Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Justiz- und Polizeidepartement EJPD

Bundesamt für Justiz BJ
Direktionsbereich Öffentliches Recht
Fachbereich Rechtsetzungsprojekte I

31.08.2024

Rechtliche Basisanalyse im Rahmen der Auslegeordnung zu den Regulierungsansätzen im Bereich künstliche Intelligenz

BJ-D-F0D63401/159

Übersicht

Die vorliegende rechtliche Analyse wird vom EJPD (BJ) im Rahmen des Auftrags des Bundesrats vom 22. November 2023, eine Auslegeordnung zu den Regulierungsansätzen im Bereich künstliche Intelligenz (KI) zu erarbeiten, erstellt. Mit dieser Auslegeordnung sollen namentlich der Regulierungsbedarf sowie Regulierungsansätze aufgezeigt werden, die dem Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit (KI-Konvention) und der Verordnung der Europäischen Union zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (KI-Verordnung) Rechnung tragen. Die rechtliche Analyse zeigt, dass das Schweizer Recht im KI-Bereich vollumfänglich Anwendung findet, aber Anpassungen erforderlich wären, wenn die Schweiz die KI-Konvention ratifizieren sollte, und allgemein, um den Herausforderungen von KI zu begegnen. Eine Annäherung an die KI-Verordnung würde ein umfassendes Eingreifen des Gesetzgebers erfordern.

Ausgangslage

Die vorliegende rechtliche Analyse soll als Unterstützung für die Auslegeordnung zu den Regulierungsansätzen im KI-Bereich dienen, die vom Bundesrat am 22. November 2023 in Auftrag gegeben wurde. Der Auftrag des Bundesrats geht von der Prämisse aus, dass der gesetzgeberische Handlungsbedarf der Schweiz im KI-Bereich erneut geprüft werden sollte. Seit 2019 und der Herausgabe des Berichts der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz», der zum Schluss kam, dass die geltende Gesetzgebung ausreichend sei, hat sich die Situation stark verändert. So werden KI-Systeme zunehmend in immer unterschiedlicheren Bereichen des privaten und des öffentlichen Sektors eingesetzt und auch auf internationaler Ebene wurden Texte erarbeitet, namentlich die KI-Konvention des Europarats und die KI-Verordnung der EU.

Die rechtliche Analyse hat zum Ziel, in Bezug auf die Herausforderungen, die sich aufgrund von KI ergeben, allfällige Lücken im Schweizer Rechtsrahmen aufzuzeigen und den gesetzgeberischen Handlungsbedarf vor dem Hintergrund des geltenden Rechts und der möglichen politischen Optionen des Bundesrats festzulegen. Sie wurde vom BJ erarbeitet und enthält Beiträge der DV (Ziff. 3 und 5.2.11), des BAKOM (Ziff. 5.2 und 5.3.2) und des IGE (Ziff. 6.2).

Inhalt

Die rechtliche Analyse stützt sich in erster Linie auf die KI-Konvention des Europarats, die für die Schweiz bei einer Ratifizierung verbindlich würde. Als erster völkerrechtlicher Vertrag über KI zeigt diese Konvention die wichtigsten rechtlichen Herausforderungen auf, die sich aufgrund von KI im Bereich des Schutzes der Menschenrechte, der Demokratie und der Rechtsstaatlichkeit ergeben. Eine Untersuchung dieser Konvention ermöglicht daher, allfällige Lücken im Schweizer Recht unter Berücksichtigung der auf internationaler Ebene erkannten Problembereiche zu ermitteln. Die Analyse betrifft das Völker- und das Landesrecht; auf das kantonale und kommunale Recht wird nicht eingegangen. In einem zweiten Teil wird die Situation dargestellt, die sich bei einer Annäherung des Schweizer Rechts an die KI-Verordnung der EU präsentieren würde. Weiter befasst sich die rechtliche Analyse mit weiteren ausgewählten Rechtsgebieten, die weder in der KI-Konvention noch in der KI-Verordnung umfassend behandelt werden, wie das Immaterialgüterrecht sowie der Bereich der zivilrechtlichen Haftung und strafrechtlichen Verantwortlichkeit.

Die Schlussfolgerungen der rechtlichen Analyse bestätigen, dass das Schweizer Recht bereits über auf KI anwendbare Rechtsnormen verfügt. Diese würden eine teilweise Übernahme der KI-Konvention ermöglichen. Es scheinen jedoch Ergänzungen nötig, insbesondere um die Transparenz der KI-Systeme zu verbessern, ihre Folgen auf die Grundrechte abzuschätzen und Kontrollmechanismen sicherzustellen. Würde eine Annäherung an die KI-Verordnung der EU in Betracht gezogen, so müssten ausführlichere Vorschriften verabschiedet werden. Bei der KI-Verordnung handelt es sich hauptsächlich um Produktsicherheitsvorschriften, wobei den verschiedenen Akteuren im KI-Bereich spezifische Pflichten auferlegt werden. Bei den weiteren ausgewählten Rechtsgebieten hat die Analyse gezeigt, dass sich gewisse Fragen stellen, die sich jedoch mit den geltenden Vorschriften grundsätzlich beantworten lassen. Ferner hätte die Verabschiedung allgemeiner Rechtsnormen für die Umsetzung der KI-Konvention, beispielsweise zur Stärkung der Transparenz der KI-Systeme, häufig zur Folge, dass der Schutz in diesen Bereichen, so etwa bei der zivilrechtlichen Haftung und strafrechtlichen Verantwortlichkeit, zusätzlich verbessert würde.

Die rechtliche Analyse konzentriert sich auf die grossen Herausforderungen und präsentiert meistens Schlussfolgerungen, die unter Berücksichtigung der getroffenen politischen Entscheidungen noch vertieft werden müssen, namentlich im Fall einer allfälligen Ratifizierung der KI-Konvention des Europarats oder einer eventuellen Annäherung an die Gesetzgebung der EU.

Die rechtliche Analyse wurde unter Berücksichtigung der Entwicklungen bis am 31. August 2024 durchgeführt.

Inhaltsverzeichnis

Übersicht **2**
Inhaltsverzeichnis **4**
1 Einleitung **8**
2 Ziel und Methodik der rechtlichen Basisanalyse **8**
3 Instrumente auf internationaler Ebene **11**
3.1 Völkerrechtliche Verträge 11
3.2 Nicht verbindliche Instrumente im KI-Bereich 12
4 Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit **13**
4.1 Einleitende Bemerkungen 13
4.1.1 Ausgangslage 13
4.1.2 Anwendbarkeit und Justiziabilität 13
4.1.3 Umsetzung in innerstaatliches Recht und Föderalismus 15
4.2 Kapitel I: Allgemeine Bestimmungen 16
4.2.1 Artikel 1 – Ziel und Zweck 16
4.2.2 Begriffsbestimmungen 17
4.2.2.1 Artikel 2 – Begriffsbestimmung «System künstlicher Intelligenz» 17
4.2.2.2 Lebenszyklus 20
4.2.3 Artikel 3 – Anwendungsbereich 21
4.2.3.1 Öffentlicher und privater Sektor 21
4.2.3.2 Nationale Sicherheit 26
4.2.3.3 Forschung und Entwicklung 27
4.2.3.4 Nationale Verteidigung 28
4.2.3.5 Zusammenfassung 28
4.3 Pflichten und Grundsätze 29
4.3.1 Kapitel II: Allgemeine Verpflichtungen 29
4.3.1.1 Artikel 4 – Schutz der Menschenrechte 29
4.3.1.2 Artikel 5 – Integrität demokratischer Prozesse und Achtung der Rechtsstaatlichkeit 30
4.3.2 Kapitel III: Grundsätze in Bezug auf Tätigkeiten im Lebenszyklus von Systemen künstlicher Intelligenz 38
4.3.2.1 Artikel 6 – Allgemeiner Ansatz 38
4.3.2.2 Artikel 7 – Menschenwürde und individuelle Autonomie 38
4.3.2.3 Artikel 8 – Transparenz und Aufsicht 40
4.3.2.4 Artikel 9 – Rechenschaftspflicht und Verantwortung 45
4.3.2.5 Artikel 10 – Gleichstellung und Nichtdiskriminierung 46
4.3.2.6 Artikel 11 – Privatsphäre und Schutz personenbezogener Daten 56
4.3.2.7 Artikel 12 – Zuverlässigkeit 60
4.3.2.8 Artikel 13 – Sichere Innovation 62
4.3.3 Kapitel IV: Rechtsmittel 67
4.3.3.1 Allgemeines 67
4.3.3.2 Artikel 14 – Rechtsmittel 67
4.3.3.3 Artikel 15 – Verfahrensgarantien 72

4.3.4 Kapitel V: Bewertung und Minderung von Risiken und nachteiligen Auswirkungen ... 76
4.3.5 Kapitel VI: Umsetzung des Übereinkommens ... 80
4.3.5.1 Allgemeines ... 80
4.3.5.2 Artikel 17 – Nichtdiskriminierung ... 80
4.3.5.3 Artikel 18 – Rechte von Menschen mit Behinderungen und von Kindern ... 80
4.3.5.4 Artikel 19 – Öffentliche Konsultation ... 81
4.3.5.5 Artikel 20 – Digitale Kompetenzen und Fähigkeiten ... 82
4.3.5.6 Artikel 21 – Schutz der bestehenden Menschenrechte ... 82
4.3.5.7 Artikel 22 – Umfassenderer Schutz ... 83
4.4 Kapitel VII: Nachfolgemechanismus und Zusammenarbeit ... 83
4.4.1 Allgemeines ... 83
4.4.2 Artikel 23 – Konferenz der Vertragsparteien ... 83
4.4.3 Artikel 24 – Berichterstattungspflicht ... 84
4.4.4 Artikel 25 – Internationale Zusammenarbeit ... 84
4.4.5 Artikel 26 – Wirksame Aufsichtsmechanismen ... 85
4.5 Kapitel VIII: Schlussbestimmungen ... 86
4.6 Zwischenfazit ... 86
5 Verordnung der Europäischen Union zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz ... 89
5.1 Struktur des Kapitels und Methode ... 89
5.2 Inhalt der Verordnung ... 89
5.2.1 Ausgangslage ... 89
5.2.2 Ziele der Regelung ... 90
5.2.3 Risikobasierter Ansatz ... 90
5.2.4 Begriffsbestimmungen ... 92
5.2.4.1 System der künstlichen Intelligenz ... 92
5.2.4.2 Modell der künstlichen Intelligenz mit allgemeinem Verwendungszweck ... 92
5.2.5 Geltungsbereich ... 93
5.2.5.1 Umfang des Geltungsbereichs ... 93
5.2.5.2 Ausnahmen ... 94
5.2.6 Verbotene Praktiken ... 96
5.2.7 Hochrisiko-Systeme der künstlichen Intelligenz ... 98
5.2.7.1 Einstufung ... 98
5.2.7.2 Anforderungen an Hochrisiko-KI-Systeme ... 100
5.2.7.3 Pflichten der Anbieter und anderer Beteiligter ... 102
5.2.7.4 Konformitätsbewertung ... 107
5.2.8 Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme ... 109
5.2.9 Andere Systeme der künstlichen Intelligenz ... 110
5.2.10 Modelle der künstlichen Intelligenz mit allgemeinem Verwendungszweck ... 111
5.2.10.1 Allgemeine Bemerkungen ... 111
5.2.10.2 Pflichten für alle Modelle der künstlichen Intelligenz mit allgemeinem Verwendungszweck ... 111

5.2.10.3 Pflichten für Modelle der künstlichen Intelligenz mit allgemeinem Verwendungszweck mit systemischem Risiko ... 112
5.2.11 Die harmonisierten Normen und ihre Rolle in der KI-Verordnung ... 114
5.2.11.1 Die harmonisierten Normen ... 114
5.2.11.2 Das Normungssystem der EU ... 114
5.2.11.3 Die Konformitätsvermutung ... 115
5.2.11.4 Die Normungsarbeiten im KI-Bereich ... 116
5.2.12 Massnahmen zur Innovationsförderung ... 117
5.2.13 Governance ... 118
5.2.14 Überwachung und Durchsetzung ... 119
5.2.15 Individuelle Rechte ... 121
5.2.16 Sanktionen ... 122
5.2.17 Inkrafttreten und Geltungsbeginn ... 122
5.3 Würdigung ... 123
5.3.1 Rechtliche Auswirkungen auf die Schweizer Akteure ... 123
5.3.1.1 Betroffene Akteure ... 123
5.3.1.2 Auswirkungen ... 124
5.3.2 Verhältnis zum Abkommen zwischen der Schweiz und der EU über die gegenseitige Anerkennung von Konformitätsbewertungen ... 125
5.3.2.1 Funktionsweise des Abkommens ... 125
5.3.2.2 Auswirkungen der KI-Verordnung auf die vom Abkommen betroffenen Schweizer Akteure ... 126
5.3.2.3 Mögliche Erweiterung des MRA ... 127
5.3.3 Verhältnis zum Angemessenheitsbeschluss der Europäischen Kommission im Bereich Datenschutz ... 128
5.3.4 Verhältnis zum Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit ... 129
5.3.5 Weitere ausgewählte Elemente ... 129
5.4 Zwischenfazit ... 132
6 Weitere ausgewählte Rechtsgebiete ... 133
6.1 Einleitung ... 133
6.2 Geistiges Eigentum ... 133
6.2.1 Urheberrecht und künstliche Intelligenz ... 133
6.2.1.1 Allgemeines ... 133
6.2.1.2 Fragestellungen und Regulierungsbedarf im Zusammenhang mit KI ... 134
6.2.2 KI und Patentrecht ... 138
6.2.2.1 Allgemeines ... 138
6.2.2.2 Herausforderungen ... 139
6.2.2.3 Regulatorischer Bedarf ... 140
6.3 Ausservertragliche Haftung ... 141
6.3.1 Allgemeines ausservertragliches Haftpflichtrecht ... 141
6.3.1.1 Vorschlag für eine EU-Richtlinie über KI-Haftung ... 141
6.3.1.2 Schweizer Recht ... 143
6.3.1.3 Würdigung ... 145

6.3.2 Produktehaftpflichtrecht ... 146
6.3.2.1 Revidierte EU-Richtlinie über die Haftung für fehlerhafte Produkte ... 146
6.3.2.2 Schweizer Recht ... 148
6.3.2.3 Würdigung ... 148
6.3.3 Schutz der Persönlichkeit ... 149
6.3.4 Fazit ... 149
6.4 Allgemeines Vertragsrecht ... 150
6.4.1 UNCITRAL Model Law on Automated Contracting ... 150
6.4.2 Schweizer Recht ... 150
6.4.2.1 Zurechnung von Willenserklärungen und vertragliche Haftung ... 150
6.4.2.2 Smart Contracts ... 151
6.4.3 Würdigung ... 152
6.5 Arbeitsrecht ... 152
6.5.1 Einleitung ... 152
6.5.2 Auf europäischer Ebene ... 153
6.5.2.1 Richtlinie zur Verbesserung der Arbeitsbedingungen in der Plattformarbeit ... 153
6.5.2.2 EU-Verordnung zu KI und Arbeitsrecht ... 155
6.5.3 Schweizerisches Arbeitsrecht: Aktueller Stand und Diskussion ... 156
6.5.3.1 Motion 23.4492 Gysi «Künstliche Intelligenz am Arbeitsplatz. Mitwirkungsrechte der Arbeitnehmenden stärken» ... 156
6.5.3.2 Einsatz von KI in der schweizerischen Arbeitswelt ... 156
6.5.3.3 Herausforderungen durch KI im schweizerischen Arbeitsrecht ... 157
6.5.3.4 Anwendbare Bestimmungen im Schweizer Recht ... 159
6.5.3.5 Würdigung ... 162
6.6 Strafrecht ... 164
6.6.1 Grundsätzliche Anwendbarkeit ... 164
6.6.2 Strafrechtliche Verantwortlichkeit ... 165
6.6.2.1 Allgemeine Bemerkungen ... 165
6.6.2.2 Beispiel: Einsatz von KI-Systemen beim automatisierten Fahren ... 167
6.6.3 Herausforderungen im Rahmen der Rechtsdurchsetzung ... 169
6.6.4 Zusammenfassung und Ausblick ... 170
7 Schlussfolgerungen ... 173
Abkürzungsverzeichnis ... 175
Anhang 1 ... 181

1 Einleitung

Mit Entscheid vom 22. November 2023 hat der Bundesrat das UVEK (BAKOM) und das EDA (Abteilung Europa) beauftragt, bis Ende 2024 eine Auslegeordnung zu den Regulierungsansätzen im Bereich künstliche Intelligenz (KI) in der Schweiz zu erarbeiten. Zur Steuerung der Arbeiten wurde ein zentraler Knotenpunkt geschaffen, der sich aus Vertreterinnen und Vertretern des UVEK (BAKOM), des EJPD (BJ) und des EDA (Abteilung Europa, DV) zusammensetzt. Das EJPD (BJ) wurde mit der Leitung der rechtlichen Basisanalyse betraut.

Zurzeit gibt es in der Schweiz keine Querschnittsgesetzgebung zu KI. Allerdings entwickelt sich die KI, wie der Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» von 2019 aufgezeigt hat, nicht im rechtsfreien Raum.[1] Der geltende Rechtsrahmen ist auf den Bereich der KI vollumfänglich anwendbar. Dabei handelt es sich namentlich um die Bundesverfassung, das DSG, das GIG, die Vorschriften über die zivilrechtliche Haftung und strafrechtliche Verantwortlichkeit usw.

Seit dem Bericht von 2019 haben im KI-Bereich bedeutende technologische Entwicklungen stattgefunden und auch der internationale Rechtsrahmen hat sich weiterentwickelt. Diese rechtliche Analyse hat zum Ziel, eine Einschätzung der gegenwärtigen Rechtslage vorzunehmen und zu untersuchen, wo im Schweizer Recht allenfalls ein Anpassungsbedarf besteht. Die Struktur und Methodik werden weiter unten vorgestellt (vgl. Ziff. 2).

Die rechtliche Analyse dient als Grundlage – neben der vom BAKOM erstellten Analyse der Regulierung von künstlicher Intelligenz in verschiedenen Ländern und Weltregionen und jener der Regulierungstätigkeiten im Zusammenhang mit KI in den verschiedenen Sektoren – für die Erarbeitung der vom Bundesrat verlangten Auslegeordnung. Diese Auslegeordnung soll dem Bundesrat als Entscheidungshilfe dienen, um 2025 einen konkreten Auftrag für eine allfällige KI-Regulierungsvorlage zu erteilen und die Zuständigkeiten in der Verwaltung zu regeln.

2 Ziel und Methodik der rechtlichen Basisanalyse

Im Rahmen dieser rechtlichen Basisanalyse soll untersucht werden, inwiefern KI-Systeme neue Herausforderungen mit sich bringen, und beurteilt werden, ob die geltende Schweizer Rechtsordnung Lücken aufweist, die geschlossen werden sollten, um diesen Herausforderungen zu begegnen. Angesichts der kurzen Frist für die Durchführung dieser Analyse konzentriert sich diese auf die wichtigsten rechtlichen Herausforderungen. Zudem konnte der gesetzgeberische Handlungsbedarf nicht immer eindeutig festgestellt werden und muss gegebenenfalls noch eingehender untersucht werden. Ferner betrifft die Analyse das Völker- und das Landesrecht; auf das kantonale und kommunale Recht wird nicht eingegangen (vgl. zu diesem Punkt Ziff. 4.1.3). Sie wurde vom BJ in Zusammenarbeit mit der DV (Ziff. 3 und 5.2.11), dem BAKOM (Ziff. 5.2 und 5.3.2) und dem IGE (Ziff. 6.2) erarbeitet.

---

[1] Herausforderungen der künstlichen Intelligenz, Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, 2019, 91, abrufbar unter: www.sbfi.admin.ch > BFI-Politik > Bildungs-, Forschungs- und Innovationspolitik des Bundes 2025–2028 > Transversale Themen im BFI-Bereich > Digitalisierung im BFI-Bereich > Künstliche Intelligenz.

Als Ausgangspunkt für diese Arbeit wurde das Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit (im Folgenden: KI-Konvention)[2] ausgewählt. Der Auftrag des Bundesrats vom November 2023 sieht in der Tat vor, dass allfällige Regulierungsansätze dieses Übereinkommen berücksichtigen müssen. Die KI-Konvention wurde vom Ausschuss für künstliche Intelligenz («Committee on Artificial Intelligence», CAI) des Europarats ausgehandelt und am 17. Mai 2024 vom Ministerkomitee des Europarats offiziell verabschiedet (vgl. Ziff. 4). Die Schweiz ist Mitglied des Europarats und die Arbeiten des CAI fanden unter ihrem Vorsitz statt. Zudem beteiligte sich eine Schweizer Delegation ebenfalls aktiv an den Arbeiten.

Würde auf politischer Ebene entschieden, dieses Übereinkommen zu ratifizieren, so wäre dieser Text für die Schweiz verbindlich. Da er sehr allgemein formuliert ist, müsste er übernommen werden (vgl. Ziff. 4.1.2). Mit dieser Analyse soll ein allfälliger Anpassungsbedarf im Schweizer Recht im Fall einer Ratifizierung skizziert werden. Dazu wird zunächst der Inhalt der Verpflichtungen präsentiert, die sich aus dem Übereinkommen ergeben, und anschliessend wird die derzeitige Schweizer Rechtslage untersucht. Gegebenenfalls wird angegeben, ob das durch das Schweizer Recht gebotene Schutzniveau ausreichend ist. Aufgrund der verfügbaren Zeit für die Erarbeitung dieser Analyse handelt es sich nicht um eine abschliessende Untersuchung; in vielen Fällen wird eine Vertiefung nötig sein, um alle bestehenden rechtlichen Fragen zu klären.

Als erster verbindlicher völkerrechtlicher Vertrag über KI schafft die KI-Konvention einen allgemeinen Rechtsrahmen und benennt die wichtigsten rechtlichen Herausforderungen in diesem Bereich in Bezug auf den Schutz der Menschenrechte, die Demokratie und die Rechtsstaatlichkeit. Thematisiert werden insbesondere Fragen des Schutzes der Grundrechte, der Transparenz, des Datenschutzes, der Gleichstellung und Nichtdiskriminierung sowie Haftungs- und Verfahrensaspekte. Die Untersuchung dieses Textes ermöglicht somit auch, allfällige Lücken im Schweizer Recht unter Berücksichtigung dieser Problembereiche ganz allgemein zu erkennen.

In einem zweiten Teil befasst sich die Analyse mit der Verordnung der EU, die harmonisierte Vorschriften für künstliche Intelligenz (im Folgenden: KI-Verordnung)[3] festlegt (vgl. Ziff. 5). Dieser Text ist für die Schweiz nicht verbindlich. Es besteht somit keine rechtliche Pflicht, diesen in das innerstaatliche Recht zu überführen. Angesichts der geografischen Lage der

---

[2] Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit, SEV 225, verfügbar in französischer Sprache unter www.coe.int > Droits humains > Intelligence artificielle et droits humains > Convention-cadre, und in englischer Sprache unter www.coe.int > Human Rights > Artificial Intelligence and Human Rights > Framework Convention (abgerufen am 26. August 2024).

[3] Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828; ABl. L 2024/1689, 12. Juli 2024, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=OJ%3AL_202401689 (abgerufen am 26. August 2024).

Schweiz und der wirtschaftlichen Beziehungen mit der EU ist eine Untersuchung dieser Verordnung jedoch unumgänglich. Zudem gilt diese für Schweizer Anbieter, die in der EU KI-Systeme in Verkehr bringen oder in Betrieb nehmen oder KI-Modelle mit allgemeinem Verwendungszweck in Verkehr bringen. Sie gilt auch für Schweizer Anbieter und Betreiber von KI-Systemen, wenn die erzeugten Ergebnisse in der EU verwendet werden. Ferner ergibt sich die Prüfung dieser Verordnung aus dem Auftrag des Bundesrats vom 22. November 2023.

Anschliessend wird eine Würdigung der KI-Verordnung vorgenommen, in der auf Fragen eingegangen wird, die aus Sicht des Schweizer Rechts relevant sind. Die Analyse beleuchtet auch das Verhältnis und die Synergien zwischen der KI-Verordnung und der KI-Konvention. Insbesondere wird untersucht, welchen Ansatz die KI-Verordnung verfolgt und inwiefern diese über die KI-Konvention hinausgeht oder sich von dieser unterscheidet.

Schliesslich geht die Analyse auf die Schweizer Rechtslage in Bezug auf Fragen der KI in Rechtsgebieten ein, die in der KI-Konvention oder der KI-Verordnung nicht umfassend behandelt werden, aber für welche die KI eine Herausforderung darstellt (vgl. Ziff. 6). Dabei handelt es sich insbesondere um relevante Aspekte des Immaterialgüterrechts, der zivilrechtlichen Haftung und der strafrechtlichen Verantwortlichkeit.

Sektorspezifische Fragen, wie die Regulierungsvorlagen in den Bereichen Energie und Gesundheit, werden in einer separaten Analyse unter der Leitung des BAKOM behandelt, wobei bestimmte transversale Aspekte, namentlich der Schutz vor Diskriminierungen, der Schutz der Privatsphäre oder die Haftungsfragen, vorbehalten sind.

Die rechtliche Analyse wurde unter Berücksichtigung der Entwicklungen bis am 31. August 2024 durchgeführt.

3 Instrumente auf internationaler Ebene[4]

Auf internationaler Ebene sind neben der KI-Konvention und der KI-Verordnung, die bereits erwähnt wurden und die gesondert untersucht werden, mehrere andere verbindliche (vgl. Ziff. 3.1) oder nicht verbindliche (vgl. Ziff. 3.2) Instrumente auf den KI-Bereich anwendbar.

3.1 Völkerrechtliche Verträge

Die KI entwickelt sich auf internationaler Ebene nicht im rechtsfreien Raum. Vielmehr ist der bestehende Rechtsrahmen auch auf den KI-Bereich anwendbar. So sind zahlreiche Verträge, Bräuche und allgemeine Grundsätze des Völkerrechts in diesem Bereich für die Schweiz bindend.

Als Beispiel seien hier die EMRK und die Konvention 108+[5] des Europarats erwähnt. Wie viele andere völkerrechtliche Verträge sind diese Konventionen auf die KI anwendbar, auch wenn sie sich nicht ausdrücklich darauf beziehen. Dank der Offenheit und der Flexibilität dieser Rechtsnormen können sich diese an die technologischen Entwicklungen anpassen.

In seinem Urteil Glukhin gegen Russland vom 4. Juli 2023[6] gelangte der EGMR beispielsweise zum Schluss, dass der Einsatz von Gesichtserkennungstechnologie eine Verletzung von Art. 8 (Recht auf Achtung des Privatlebens) und 10 EMRK (Freiheit der Meinungsäusserung) darstellen kann. Dieses Beispiel zeigt, dass gewisse Situationen, in denen ein KI-System rechtswidrig eingesetzt wird, mit dem bestehenden Recht bereits erfasst werden können (für weitere Erwägungen zu diesem Urteil vgl. Ziff. 4.3.1.2).

Der EGMR anerkennt zudem in bestimmten Fällen, dass die Staaten positive Verpflichtungen haben, Voraussetzungen zu schaffen, die den Schutz und die Ausübung der durch die EMRK garantierten Rechte ermöglichen. Somit könnte die Auslegung des bestehenden Völkerrechts unter bestimmten Voraussetzungen bereits eine positive Verpflichtung ergeben, im KI-Bereich Schutznormen zu erlassen.[7] Die internationalen Menschenrechtsnormen enthalten nicht nur die Verpflichtung, die Rechte des Einzelnen nicht einzuschränken; sie verlangen vom Staat auch, den Schutz des Einzelnen gegen Handlungen Dritter zu garantieren und positive Massnahmen zum Schutz der Menschenrechte zu ergreifen. Vor diesem Hintergrund könnte aus dem geltenden Völkerrecht eine positive Verpflichtung abgeleitet werden, im KI-Bereich Rechtsvorschriften zu erlassen.

---

[4] Dieses Kapitel wurde auf der Grundlage von Beiträgen der DV verfasst.

[5] Die Konvention 108 wurde mit dem Änderungsprotokoll zum Übereinkommen zum Schutz des Menschen bei der automatischen Verarbeitung personenbezogener Daten, das vom Ministerkomitee an seiner 128. Sitzung in Helsingör am 18. Mai 2018 verabschiedet wurde, modernisiert. Die Schweiz hat das Protokoll am 7. September 2023 ratifiziert, aber es ist noch nicht Kraft getreten.

[6] EGMR, Glukhin gegen Russland, 11519/20 (4. Juli 2023).

[7] WOLFGANG HOFFMANN-RIEM, Artificial Intelligence as a Challenge for Law and Regulation, in: Thomas Wischmeyer/Timo Rademacher (Hrsg.), Regulating Artificial Intelligence, Cham 2020, 1 ff., 8.

3.2 Nicht verbindliche Instrumente im KI-Bereich

Neben dem oben erwähnten Rechtsrahmen bestehen mehrere nicht verbindliche internationale Instrumente (*soft law*), die für die Schweiz im KI-Bereich relevant sind.

Dabei handelt es sich namentlich um folgende Texte: Empfehlung zu künstlicher Intelligenz der OECD vom 22. Mai 2019, überarbeitet am 8. November 2023, mit einer aktualisierten Definition des Begriffs «KI-System»; Erklärung des Ministerkomitees des Europarats zu den manipulativen Fähigkeiten algorithmischer Prozesse, Decl(13/02/2019)1; Empfehlung des Ministerkomitees des Europarats über die Auswirkungen algorithmischer Systeme auf die Menschenrechte, CM/Rec(2020)1; UNESCO-Empfehlung zur Ethik der Künstlichen Intelligenz, verabschiedet am 23. November 2021; Erklärung der Staats- und Regierungschefinnen und -chefs, verabschiedet am 4. Gipfeltreffen des Europarats in Reykjavik vom 16. und 17. Mai 2023; Erklärung von Bletchley der Länder, die am Gipfeltreffen zur Sicherheit künstlicher Intelligenz am 1. und 2. November 2023 teilnahmen; Resolution der UN-Generalversammlung vom 21. März 2024 zu künstlicher Intelligenz.

Zudem werden zurzeit weitere nicht verbindliche Instrumente erarbeitet. Insbesondere verfasst ein Ausschuss des Europarats bis Ende 2025 eine Empfehlung zu den Auswirkungen von KI-Systemen, ihrem Potenzial zur Förderung der Gleichstellung, einschliesslich der Geschlechtergleichstellung, und den Risiken, die sie in Bezug auf Diskriminierung mit sich bringen können. Ebenfalls bis 2025 werden von einem anderen Ausschuss des Europarats Leitlinien zu den Auswirkungen generativer KI auf die Freiheit der Meinungsäusserung fertiggestellt. Die Schweiz beteiligt sich aktiv in diesen beiden Ausschüssen.

Auch wenn nicht rechtlich verbindlich, sind diese Texte dennoch von Bedeutung, da sie einen gewissen Konsens auf internationaler Ebene im KI-Bereich widerspiegeln. Die Schweiz engagiert sich seit mehreren Jahren aktiv in diesen verschiedenen internationalen Prozessen. Zudem dienen diese Instrumente der Schweiz als Referenz. So werden sie im Bericht «Herausforderungen der künstlichen Intelligenz» der interdepartementalen Arbeitsgruppe von 2019 mehrmals erwähnt.[8]

Zum Schluss seien noch die internationalen Leitprinzipien und der internationale Verhaltenskodex des Prozesses von Hiroshima für Organisationen, die fortgeschrittene KI-Systeme entwickeln, vom 30. Oktober 2023 sowie die Erklärung der Staats- und Regierungschefinnen und -chefs der G7 zum KI-Prozess von Hiroshima vom 30. Oktober 2023 erwähnt. Die Schweiz hat sich nicht an diesen Initiativen beteiligt.

[8] Bericht «Herausforderungen der künstlichen Intelligenz» (Fn. 1).

4 Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit

4.1 Einleitende Bemerkungen

4.1.1 Ausgangslage

Der Ausschuss für künstliche Intelligenz («Committee on Artificial Intelligence», CAI) wurde im Juni 2022 vom Ministerkomitee des Europarats beauftragt, ein verbindliches, übergreifendes Rechtsinstrument zu KI zu entwickeln, das auf den Standards des Europarats für Menschenrechte, Demokratie und Rechtsstaatlichkeit aufbaut.

Der CAI setzt sich aus Vertretern aller Mitgliedstaaten des Europarats und der Europäischen Union zusammen. Darüber hinaus sind mehrere Staaten mit Beobachterstatus in den Prozess eingebunden: Argentinien, Australien, Kanada, Costa Rica, der Heilige Stuhl, Israel, Japan, Mexiko, Peru, die Vereinigten Staaten von Amerika und Uruguay. Daneben sind auch Vertreter der Zivilgesellschaft beteiligt.

Die Verhandlungen des CAI fanden unter der Leitung eines Schweizers als Vorsitzenden statt. Darüber hinaus war die Schweiz im CAI durch eine Delegation aus Repräsentanten des BAKOM, der DV und des BJ vertreten.

Die Verhandlungen zur KI-Konvention wurden im März 2024 abgeschlossen. Das Ministerkomitee hat den Text im Mai 2024 verabschiedet. Im September 2024 wurde die Konvention zur Unterzeichnung aufgelegt. Die Hauptaufgabe des CAI ist abgeschlossen, der Ausschuss arbeitet jedoch weiterhin an verwandten Themen, insbesondere an der Entwicklung einer fakultativen Methodik, die als Instrument für das Risiko- und Folgenmanagement von KI dienen soll.

4.1.2 Anwendbarkeit und Justiziabilität

Zunächst stellt sich die Frage, ob die KI-Konvention unmittelbar anwendbar (oder justiziabel bzw. *self-executing*) ist, d. h. unmittelbar Rechte und Pflichten für natürliche und juristische Personen begründet, oder ob sie sich nur an Staaten richtet, die sie noch umsetzen müssen, um ihr Wirkung zu verleihen.

Unmittelbar anwendbar sind Normen, die so konkret und bestimmt sind, dass natürliche oder juristische Personen aus ihnen Rechte und Pflichten ableiten können, und sich in Gerichts- oder Verwaltungsverfahren darauf berufen können. Entsprechend können rechtsanwendende Behörden und Gerichte solche internationalen Normen unmittelbar anwenden.[9]

[9] Das Verhältnis von Völkerrecht und Landesrecht, Bericht des Bundesrates vom 5. März 2010 in Erfüllung des Postulats 07.3764 der Kommission für Rechtsfragen des Ständerates vom 16. Oktober 2007 und des Postulats 08.3765 der Staatspolitischen Kommission des Nationalrates vom 20. November 2008, BBl **2010** 2263 2286.

Das BGer ist der Ansicht, dass die unmittelbare Anwendbarkeit einer völkerrechtlichen Bestimmung voraussetzt, dass diese nach übereinstimmender Auffassung im Gesamtzusammenhang sowie im Lichte von Ziel und Zweck des Vertrags betrachtet unbedingt und eindeutig genug formuliert sein muss, um eine direkte Wirkung erzeugen und in einem konkreten Fall angewendet werden bzw. die Grundlage für eine Entscheidung darstellen zu können.[10] Die unmittelbare Anwendbarkeit einer Norm ist letztlich eine Frage der Auslegung der betreffenden Bestimmung durch die staatlichen Behörden.

Zur Bestimmung der unmittelbaren Anwendbarkeit hat das BGer eine umfangreiche Rechtsprechung entwickelt. Für die direkte Anwendbarkeit spricht ein hohes Schutzbedürfnis des Einzelnen. Anders verhält es sich hingegen bei erheblichen Auswirkungen auf den Staat als Ganzes, bei einer komplexen Materie, wenn sich die Beurteilung des Einzelfalls bei der gerichtlichen Überprüfung als schwierig erweist, oder bei erheblichen finanziellen Auswirkungen. Zurückhaltung übt das BGer auch, wenn die unmittelbare Anwendung eine politische Beurteilung oder die Prüfung von Grundsatzfragen erfordert.[11]

Als nicht direkt anwendbar (d. h. nicht justiziabel oder *non-self-executing*) gelten Normen programmatischer Natur, bzw. Bestimmungen, die sich an den Staat als Ganzes richten. Sie bedürfen der Konkretisierung durch die zuständigen staatlichen Stellen, bevor sie für Private Rechte und Pflichten begründen.[12]

Die Prüfung der Normen der KI-Konvention nach den oben genannten Kriterien führt zu dem Ergebnis, dass sich die Konvention allgemein an Staaten richtet und nicht unmittelbar anwendbar ist. Dies ergibt sich auch aus dem Text der Konvention selbst, wo es in Artikel 1 Absatz 2 heisst, dass jede Vertragspartei geeignete Gesetzgebungs-, Verwaltungs- oder sonstigen Massnahmen trifft oder aufrechterhält, um diesem Übereinkommen Wirksamkeit zu verleihen. Es ist jedoch nicht auszuschliessen, dass ein Gericht im Einzelfall zu einem anderen Ergebnis kommt. Auch die Entwicklung der Rechtsprechung könnte eine Rolle spielen.

Ein Gericht könnte die Konvention im Rahmen einer Klage einer Einzelperson prüfen, die sich darüber beschwert, dass der Staat seinen Verpflichtungen nicht nachkommt. Ein Gericht könnte dann zum Schluss kommen, dass der Gesetzgeber aufgrund der Bestimmungen der Konvention tätig werden muss. Wenn die fragliche Norm jedoch nur eine Verpflichtung zum Tätigwerden auferlegt, aber nicht darlegt, wie dies zu geschehen hat, wird sich das Gericht darauf beschränken, den Gesetzgeber durch einen sogenannten Appellentscheid zum Handeln aufzufordern.[13]

---

[10] Das Verhältnis von Völkerrecht und Landesrecht (Fn. 9), 2303 f.

[11] Das Verhältnis von Völkerrecht und Landesrecht (Fn. 9), 2303 f.

[12] Das Verhältnis von Völkerrecht und Landesrecht (Fn. 9), 2286.

[13] Vgl. BGE **137** I 305, E. 6.6 f.

Eine Berufung auf die KI-Konvention vor dem EGMR wäre allenfalls im Zusammenhang mit der Verletzung eines durch die EMRK geschützten Rechts denkbar. Die Nichtumsetzung einer Bestimmung der KI-Konvention, die die Vertragsstaaten auffordert, in einem bestimmten Bereich gesetzgeberische oder andere Massnahmen zu ergreifen, könnte z. B. dazu führen, dass der EGMR eine Verletzung eines durch die EMRK garantierten Rechts feststellt, sofern die Untätigkeit als Verletzung positiver Verpflichtungen des Staates, z. B. aus Art. 8 EMRK, ausgelegt wird. Dies gilt selbstverständlich nur für diejenigen Vertragsstaaten, die Mitglied des Europarats sind. Dies könnte zu einer Zweiklassenstruktur führen, auf der einen Seite mit den Vertragsstaaten, die Mitglied des Europarats sind, und auf der anderen Seite den Vertragsstaaten, die nicht Mitglied des Europarats sind. Diese Situation ist jedoch allen Übereinkommen des Europarats gemein, die auch von Nichtmitgliedstaaten ratifiziert werden können, und ist daher nicht neu.

4.1.3 Umsetzung in innerstaatliches Recht und Föderalismus

Gemäss Art. 5 Abs. 4 BV haben Bund und Kantone das Völkerrecht zu beachten.

Wie bereits erwähnt (vgl. Ziff. 4.1.2) richtet sich die KI-Konvention an die Staaten und enthält allgemein formulierte Regeln und Grundsätze. Die darin enthaltenen Verpflichtungen müssen daher in innerstaatliches Recht umgesetzt werden.

In einem föderalistischen Staat wie der Schweiz kann diese Pflicht zur Umsetzung je nach Rechtsgebiet dem Bund, den Kantonen oder diesen beiden Staatsebenen gleichzeitig obliegen.[14] Bei einer allfälligen Ratifizierung der KI-Konvention wären für jede Verpflichtung die Regeln der Kompetenzverteilung zwischen Bund und Kantonen im betreffenden Bereich zu prüfen. Je nach Ergebnis läge die Pflicht zur Umsetzung beim Bund oder bei den Kantonen.

Im Bereich des Datenschutzes kann sich der Bund beispielsweise auf Artikel 95 Absatz 1, Artikel 97 Absatz 1 und Artikel 122 Absatz 1 BV sowie auf die ihm eigene Organisationskompetenz für seine Behörden stützen, um Vorschriften für den privaten Sektor und die Bundesorgane zu erlassen. Demgegenüber untersteht die Datenbearbeitung durch kantonale oder kommunale Organe unter Vorbehalt der Zuständigkeiten, die sich aus den materiellen Datenschutzbestimmungen in den Spezialgesetzen des Bundes ergeben,[15] dem kantonalen Recht.[16]

So liegt beispielsweise die Zuständigkeit für das Bildungswesen im Wesentlichen bei den Kantonen (vgl. Art. 62 BV).

---

[14] JUDITH WYTTENBACH, Umsetzung von Menschenrechtsübereinkommen in Bundesstaaten, Zürich/St. Gallen 2017, 299.

[15] So gelten beispielsweise die Artikel 89*a* ff. SVG, oder die Art. 49*a* ff. AHVG für alle rechtsanwendenden Organe oder Privatpersonen, einschliesslich kantonaler Organe.

[16] CR LPD-SYLVAIN MÉTILLE/LIVIO DI TRIA, Art. 2 N 14; PHILIPPE MEIER, Protection des données – Fondements, principes généraux et droit privé, Bern 2010, N 356 ff.

Die vorliegende Analyse konzentriert sich auf eine allfällige Umsetzung der KI-Konvention auf Bundesebene. Die Kantone wären jedoch, wie bereits erwähnt, für die Umsetzung in ihren jeweiligen Kompetenzbereichen zuständig.

4.2 Kapitel I: Allgemeine Bestimmungen

4.2.1 Artikel 1 – Ziel und Zweck

Artikel 1 regelt das Ziel und den Zweck der KI-Konvention.

Gemäss Absatz 1 soll die Konvention sicherstellen, dass die Aktivitäten im Lebenszyklus von KI-Systemen uneingeschränkt mit den Menschenrechten, der Demokratie und der Rechtsstaatlichkeit vereinbar sind. Dies entspricht der Aufgabe des Europarats. Der Begriff «KI-System» wird in Artikel 2 definiert (vgl. Ziff. 4.2.2).

Die Konvention ergänzt die in jedem Vertragsstaat bestehenden Mechanismen zum Schutz der Menschenrechte, einschliesslich der anwendbaren nationalen und völkerrechtlichen Verpflichtungen. Die Konvention schafft jedoch keine neuen Grundrechte. Sie regelt auch nicht die KI als solche, sondern nur Aspekte von KI, die einen Zusammenhang zum Schutz der Menschenrechte, der Demokratie und der Rechtsstaatlichkeit aufweisen.

Absatz 2 betrifft die Umsetzung der Konvention. Er sieht vor, dass jede Vertragspartei die erforderlichen gesetzgeberischen, verwaltungstechnischen oder sonstigen Massnahmen einführt oder beibehält, um den Bestimmungen der Konvention Wirksamkeit zu verleihen. In dieser Formulierung kommt der völkerrechtliche Grundsatz zum Ausdruck, dass die Staaten bei der Art und Weise der Umsetzung ihrer völkerrechtlichen Verpflichtungen frei sind. Bezüglich der Massnahmen zur Umsetzung der Konvention steht den Staaten ein grosser Spielraum zur Verfügung. Es handelt sich um eine Ergebnis- und nicht um eine Mittelverpflichtung. Theoretisch darf ein Staat somit frei entscheiden, ob eine Bestimmung in seiner internen Rechtsordnung durch die Legislative, die Exekutive oder die Judikative umgesetzt wird. Artikel 1 Absatz 2 sieht zudem vor, dass diese Massnahmen abgestuft und differenziert sein müssen, soweit erforderlich, je nach Schwere und Wahrscheinlichkeit des Auftretens negativer Auswirkungen auf die Menschenrechte, die Demokratie und der Rechtsstaatlichkeit während des gesamten Lebenszyklus von KI-Systemen. Dies bedeutet, dass die Massnahmen dem Risikograd eines KI-Systems anzupassen sind

Im schweizerischen Recht ist die jeweilige Normstufe für die Umsetzung der Konvention in innerstaatliches Recht anhand von Artikel 5 Absatz 1 und Artikel 164 BV zu prüfen. Letzterer Artikel bestimmt, dass alle wichtigen rechtsetzenden Bestimmungen in der Form eines Bundesgesetzes zu erlassen sind. Dazu gehören insbesondere die grundlegenden Bestimmungen über: a) die Ausübung der politischen Rechte; b) die Einschränkungen verfassungsmässiger Rechte; c) die Rechte und Pflichten von Personen; d) den Kreis der Abgabepflichtigen sowie den Gegenstand und die Bemessung von Abgaben; e) die Aufgaben und die Leistungen des Bundes; f) die Verpflichtungen der Kantone bei der Umsetzung und beim Vollzug des Bundesrechts; g) die Organisation und das Verfahren der Bundesbehörden.

Art. 164 BV schreibt vor, dass in bestimmten Bereichen gesetzgeberische Massnahmen zu treffen sind, wenn die Bedeutung der zu regelnden Materie keine andere Form der Regelung

(Verordnung, Beschluss usw.) zulässt. Es handelt sich dabei um den Grundsatz des allgemeinen Vorbehalts des Gesetzes im formellen Sinn[17], wonach wichtiges staatliches Handeln einer formell gesetzlichen Grundlage bedarf. Dieser Grundsatz stellt sicher, dass die grossen Werte- und Interessenabwägungen und die grossen politischen und rechtstechnischen Entscheidungen[18] von der Legislative getroffen werden, wodurch politische Repräsentativität und Partizipation gewährleistet werden.

Somit sind für die zur Umsetzung der Konvention im innerstaatlichen Recht zu erlassenden Rechtsakte die verfassungsrechtlichen Vorgaben zu berücksichtigen. Die wichtigsten Grundsätze und Grundregeln müssen in einem formellen Gesetz konkretisiert werden.

4.2.2 Begriffsbestimmungen

4.2.2.1 Artikel 2 – Begriffsbestimmung «System künstlicher Intelligenz»

Im Rahmen der Konvention wird ein KI-System als ein maschinengestütztes System definiert, das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können. Die verschiedenen Systeme künstlicher Intelligenz unterscheiden sich in ihrem Grad der Autonomie und der Anpassungsfähigkeit nach ihrer Betriebsaufnahme.

Diese Definition lehnt sich an die von der OECD am 8. November 2023 verabschiedete Definition[19] an und trägt der Notwendigkeit Rechnung, die internationale Zusammenarbeit im Bereich der KI zu stärken und die Bemühungen um eine weltweite Harmonisierung von KI-Governance, einschliesslich der Vereinheitlichung der Terminologie, zu unterstützen.

Diese Definition umfasst ein breites Verständnis von KI-Systemen, insbesondere in Abgrenzung zu anderen, einfacheren Arten von Softwaresystemen, die auf ausschliesslich von Menschen erstellten Regeln zur Automatisierung von Arbeitsschritten beruhen. Sie ist hinreichend abstrakt und flexibel, um zukünftige technologische Entwicklungen zu berücksichtigen. Die Definition ist auch insofern funktional, als sie speziell für die Zwecke der KI-Konvention entwickelt wurde und nicht darauf abzielt, eine universelle Bedeutung des Begriffs zu vermitteln.

---

[17] BSK BV-JUDITH WYTTENBACH/KARL-MARC WYSS, Art. 164 N 4; SGK BV-PIERRE TSCHANNEN, Art. 164 N 4; CR Cst.-JACQUES DUBEY, Art. 164 N 10.

[18] CR Cst.-JACQUES DUBEY, Art. 164 N 11.

[19] OECD, Empfehlung des Rates zu künstlicher Intelligenz, 2024, OECD/LEGAL/0449, Kapitel I, abrufbar unter: https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449 (abgerufen am 26. August 2024).

Laut dem erläuternden Bericht zur KI-Konvention können die Vertragsparteien diese Definition im Interesse der Rechtssicherheit in ihrer nationalen Rechtsordnung präzisieren, ohne dabei deren Anwendungsbereich einzuschränken.[20]

Die Definition entspricht weitgehend derjenigen in der KI-Verordnung der EU (vgl. Ziff. 5.2.4.1).

Der erläuternde Bericht zur KI-Konvention enthält nur wenige Hinweise zur Auslegung der Definition. Er verweist auf die Begründung zur aktualisierten Definition von KI durch die OECD.[21] Für Details sei daher auf dieses Dokument verwiesen. Die wichtigsten Elemente sind folgende:

- Explizite oder implizite Ziele eines KI-Systems

  Die Ziele eines KI-Systems können explizit oder implizit sein. Explizite Ziele werden von einem Menschen direkt ins System eingebettet. Implizite Ziele ergeben sich aus einem von Menschen definierten Regelwerk (z. B. ist ein System für automatisiertes Fahren so programmiert, dass es das Strassenverkehrsgesetz einhält, aber implizit schützt es Leben); weiter gibt es Zwecke, die sich aus dem System, das selbst neue Zwecke lernt, abgeleitet werden (z. B. generative KI-Systeme wie ChatGPT).

- Ein KI-System leitet von erhaltenen Inputs ab, wie Ergebnisse zu erzeugen sind

  Das KI-System wird mit Inputs versorgt, d. h. Daten und Programmierregeln, die von Menschen oder Maschinen stammen können. Es verarbeitet diese Eingaben anhand eines algorithmischen Modells, das in der Lage ist, ein Ergebnis zu berechnen.

  Beispielsweise leitet ein System zur visuellen Objekterkennung ab, wie ein Ergebnis (d. h. die Identifizierung des Objekts im Bild) zu erzeugen ist, aus einer Eingabe, die aus den Pixeln des Bildes besteht.

- Ein KI-System produziert Ergebnisse, die physische oder virtuelle Umgebungen beeinflussen können

  Die verschiedenen von KI-Systemen ausgeführten Funktionen führen zu Ergebnissen. Diese lassen sich Kategorien wie Empfehlungen, Vorhersagen und Entscheidungen

---

[20] Erläuternder Bericht zum Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit, STCE 225, N 27 (verfügbar in französischer Sprache unter www.coe.int > Droits humains > Intelligence artificielle et droits humains > Convention-cadre, und in englischer Sprache unter www.coe.int > Human Rights > Artificial Intelligence and Human Rights > Framework Convention).

[21] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 24; OCDE, Explanatory memorandum on the update OECD definition of an AI system. OECD Artificial intelligence papers, Nr. 8, OECD Publishing, März 2024, abrufbar unter: https://doi.org/10.1787/623da898-en (abgerufen am 26. August 2024).

zuordnen. Wenn beispielsweise ein System für automatisiertes Fahren vorhersagt, dass es sich bei den über seine Kamera eingegebenen Pixeln um einen Fussgänger handelt, kann es entweder empfehlen zu bremsen, oder entscheiden zu bremsen.

Die Definition nennt explizit «Inhalte» als Ergebniskategorie. Der Begriff bezieht sich insbesondere auf sogenannte generative KI-Systeme wie ChatGPT, die Inhalte wie Texte, Bilder, Ton oder Videos produzieren.

Die von KI-Systemen beeinflussten Umgebungen können physisch oder virtuell sein. Die Ergebnisse werden somit entweder von Menschen oder von Maschinen- oder Gerätekomponenten wahrgenommen.

- Unterschiedliche Grade an Autonomie und Anpassungsfähigkeit nach Inbetriebnahme

  Die *Autonomie eines KI-Systems* bezieht sich auf die Fähigkeit dieses Systems, ohne menschliches Eingreifen zu lernen oder zu handeln. Bestimmte KI-Systeme sind in der Lage, Ergebnisse zu generieren, ohne dass dies explizit in den programmierten Zwecken des Systems vorgesehen ist und ohne dass ein Mensch spezifische Anweisungen gibt.

  Die Erwähnung der *Anpassungsfähigkeit* berücksichtigt die Tatsache, dass sich einige KI-Systeme auch nach Konzeption und Inbetriebnahme weiterentwickeln können. Beispiele hierfür sind Spracherkennungssysteme, die sich an die Stimme einer Person anpassen, oder personalisierte Musikempfehlungssysteme. KI-Systeme können einmalig, periodisch oder kontinuierlich trainiert werden und arbeiten, indem sie Muster und Verbindungen in den Daten ableiten. Aufgrund dieses Trainings können einige KI-Systeme die Fähigkeit erwerben, neue Arten von Schlussfolgerungen zu ziehen, die von den Programmierern ursprünglich nicht vorgesehen waren.

Aktuell findet sich im Schweizer Recht keine Definition von KI-Systemen. Aufgrund des technologieneutralen Ansatzes des schweizerischen Gesetzgebers deckt der geltende Rechtsrahmen jedoch auch ohne eine solche Definition Sachverhalte mit KI-Systemen ab.

Es sei darauf hingewiesen, dass das beim BFS angesiedelte CNAI eine harmonisierte Terminologie für den KI-Bereich bereitstellt, um eine gemeinsame Sprache und ein entsprechendes gemeinsames Begriffsverständnis auf Ebene der Bundesverwaltung zu erleichtern.[22] Diese Terminologie enthält auch eine Definition von KI-Systemen, die derjenigen in Artikel 2 der KI-Konvention und der OECD sehr nahe kommt.

[22] Vgl. www.cnai.swiss > Dienstleistungen > Terminologie (Stand 21. Dezember 2023), 7, abgerufen am 26. August 2024: «Ein KI-System ist ein maschinenbasiertes System, das für explizite oder implizite Ziele aus den empfangenen Inputs schlussfolgert, wie es Outputs wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erzeugen kann, welche die physische oder virtuelle Umgebung beeinflussen können. KI-Systeme können mit unterschiedlichem Ausmass an Autonomie ausgestattet werden.»

Einige der aus der KI-Konvention abgeleiteten Verpflichtungen beziehen sich spezifisch auf KI-Systeme, wie z. B. die Verpflichtung zur Schaffung eines Rahmens für das Risiko- und Folgenmanagement (vgl. Ziff. 4.3.4). Im Falle einer Ratifizierung der KI-Konvention wäre es daher wünschenswert, auch im schweizerischen Recht eine Definition von KI-Systemen einzuführen, um den Anwendungsbereich einzelner Bestimmungen der Konvention genauer abzugrenzen.

4.2.2.2 Lebenszyklus

Die KI-Konvention nimmt jeweils Bezug auf den Lebenszyklus der KI-Systeme. Damit bringt sie zum Ausdruck, dass die Bestimmungen der KI-Konvention in allen Tätigkeitsphasen von KI-Systemen umzusetzen sind, von der Konzeption bis zur Ausserbetriebnahme, und zwar unabhängig davon, wer an diesen Tätigkeiten beteiligt ist.[23]

Obwohl die Nutzungsphase von KI-Systemen häufig als die risikoreichste Phase angesehen wird, kann es auch in anderen Phasen zu Verletzungen der Menschenrechte, der Demokratie und der Rechtsstaatlichkeit kommen. So soll beispielsweise Artikel 10 der KI-Konvention sicherstellen, dass Gleichstellung und Nichtdiskriminierung in allen Phasen eines KI-Systems gewahrt werden, also nicht nur beim Einsatz eines KI-Systems, sondern auch in der Entwicklungsphase, z. B. bei den Anforderungen an den Datensatz für das Training des Systems.

Im Text der Konvention fehlt eine Definition des Begriffs «Lebenszyklus», der erläuternde Bericht enthält jedoch einige Klarstellungen. Ohne die Phasen im Lebenszyklus eines KI-Systems abschliessend aufzuzählen, wird unter Bezugnahme auf die Arbeiten der OECD ein Überblick über mögliche Phasen gegeben: 1) Planung und Entwurf, 2) Datensammlung und -bearbeitung, 3) Entwicklung von KI-Systemen, einschliesslich der Entwicklung von Modellen und/oder der Anpassung vorhandener Modelle an spezifische Aufgaben, 4) Testen, Verifizierung und Validierung, 5) Lieferung/Zugänglichmachung der Systeme, 6) Einsatz, 7) Betrieb und Überwachung, und 8) Ausserbetriebnahme. Diese Aktivitäten sind häufig iterativ, aber nicht notwendigerweise sequenziell.[24]

[23] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 15.

[24] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 15.

4.2.3 Artikel 3 – Anwendungsbereich

4.2.3.1 Öffentlicher und privater Sektor

4.2.3.1.1 Allgemeines

Die KI-Konvention hat einen weiten Anwendungsbereich, um Tätigkeiten im Lebenszyklus von KI-Systemen zu erfassen, die mit den Menschenrechten, der Demokratie und der Rechtsstaatlichkeit in Konflikt geraten können.[25]

Die Konvention gilt sowohl für den öffentlichen als auch für den privaten Sektor, wenn auch mit gewissen Unterschieden. Adressat der Verpflichtungen aus der Konvention ist in beiden Fällen der Staat, der auch in beiden Fällen über einen gewissen Ermessensspielraum bei der Umsetzung seiner Verpflichtungen verfügt. Die Verhandlungsführer wollten damit der Vielfalt der verschiedenen Rechtssysteme, Traditionen und Gepflogenheiten sowie den zahlreichen Einsatzmöglichkeiten von KI-Systemen sowohl im öffentlichen als auch im privaten Sektor gerecht werden.[26]

Im öffentlichen Sektor sind die Staaten verpflichtet, die Konvention auf die Tätigkeiten während des ganzen Lebenszyklus von KI-Systemen anzuwenden. Darunter fallen auch Tätigkeiten privater Akteure, die im Auftrag der öffentlichen Hand handeln (Art. 3 Abs. 1 Bst. a).[27] KI-Systeme werden im öffentlichen Sektor beispielsweise bei der Gewährung von Sozialleistungen, im Bereich der Steuern oder im Rahmen von Asylverfahren eingesetzt.[28] In der Praxis müssen die Vertragsstaaten nach Prüfung der Risiken und Auswirkungen der betreffenden Tätigkeiten entscheiden, wie sie die Bestimmungen der Konvention umsetzen wollen. Sie können ihre Gesetze beibehalten, anpassen oder neue Gesetze erlassen. Je nach Abstufung in Artikel 1 Absatz 2 können diese Massnahmen aus administrativen Massnahmen oder Kreisschreiben, der Rechtsprechung der Gerichte oder nicht zwingenden Massnahmen bestehen.[29] In der Schweiz sind, wie bereits erwähnt (vgl. Ziff. 4.2.1), in jedem Fall die verfassungsrechtlichen Vorgaben für die Form der zu erlassenden Rechtsakte zu beachten.

Während der Verhandlungen musste ein Kompromiss zwischen den Staaten gefunden werden, die sich für bzw. gegen den Einbezug des Privatsektors in den Geltungsbereich der KI-

---

[25] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 26.

[26] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 16.

[27] Dazu gehören die Tätigkeiten privater Akteure, die im Rahmen eines Vertrags mit einer Behörde oder anderen privaten Stelle zur Erbringung öffentlicher Dienstleistungen tätig sind, sowie die öffentliche Auftragsvergabe und das öffentliche Beschaffungswesen, vgl. Erläuternder Bericht zur KI-Konvention (Fn. 20), N 28.

[28] Vgl. NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung: Entwicklungen und Herausforderungen, Jusletter IT 4. Juli 2024, 5 N 9.

[29] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 18 ff.

Konvention aussprachen. Nun sieht der Text vor, dass die Vertragsparteien in Übereinstimmung mit dem Ziel und dem Zweck der Konvention die Risiken und Auswirkungen angehen müssen, die sich aus den Tätigkeiten innerhalb des ganzen Lebenszyklus der im privaten Sektor verwendeten KI-Systeme ergeben. Die Bezugnahme auf das Ziel und den Zweck der Konvention bedeutet, dass alle in Artikel 1 enthaltenen Konzepte übernommen werden. Somit hat der Staat auch im Privatsektor die Risiken und Auswirkungen von KI-Tätigkeiten auf Menschenrechte, Demokratie und Rechtsstaatlichkeit zu prüfen und dann zu entscheiden, wie er diese angehen will.[30] Er kann entscheiden, die in den Kapiteln II bis VI der KI-Konvention genannten Grundsätze und Verpflichtungen anzuwenden, kann aber auch andere Massnahmen ergreifen.

Der Unterschied zwischen den beiden Sektoren besteht nicht so sehr im angestrebten Ziel – in beiden Fällen geht es darum, die mit KI-Systemen verbundenen Risiken und deren Auswirkungen auf Menschenrechte, Demokratie und Rechtsstaatlichkeit anzugehen – sondern in der Art und Weise, wie dieses Ziel erreicht werden soll. Im Privatsektor besteht diesbezüglich ein grösserer Spielraum.

Gemäss dem erläuternden Bericht können die in Artikel 3 Absatz 1 Buchstabe b genannten «anderen Massnahmen» für den Privatsektor Verwaltungsmassnahmen sein.[31] Gemeint sind Weisungen oder Kreisschreiben. Ob für die Schweiz der Rückgriff auf solche Rechtsakte ausreicht, ist jedoch fraglich. Die Auferlegung von Rechten und Pflichten im Bereich der KI im Privatsektor bedarf grundsätzlich einer formell gesetzlichen Grundlage, insoweit diese Grundrechte der Betroffenen (insbesondere die Wirtschaftsfreiheit) tangieren. Gemäss dem erläuternden Bericht können die Massnahmen auch freiwillig sein.[32] Dabei wird es sich aller Voraussicht nach um Selbstregulierungsmassnahmen handeln. Es ist jedoch zweifelhaft, dass die Ziele der Konvention in der Schweiz ohne staatliche Massnahmen erreicht werden können, die auch darin bestehen könnten, private Akteure zur Schaffung von Selbstregulierungsmechanismen zu verpflichten oder diese durch den Staat zu unterstützen.

In jedem Fall muss der Staat sicherstellen, dass der gewählte Ansatz die Risiken und Auswirkungen von KI im Privatsektor in einer Weise erfasst, die dem Ziel und dem Zweck der der KI-Konvention entspricht. Darüber hinaus dürfen die Vertragsparteien ihre völkerrechtlichen Verpflichtungen nicht unterschreiten (Art. 3 Abs. 1 Bst. b *in fine*).

Die Staaten müssen ihre Wahl, die jederzeit geändert werden kann, der Generalsekretärin oder dem Generalsekretär bei der Ratifikation, der Annahme oder der Genehmigung der Konvention, bzw. beim Beitritt zur Konvention mitteilen (Art. 3 Abs. 1 Bst. b). Jede Vertragspartei legt der Konferenz der Vertragsparteien innerhalb von zwei Jahren nach ihrem Beitritt und da-

---

30 Erläuternder Bericht zur KI-Konvention (Fn. 20), N 29.

31 Erläuternder Bericht zur KI-Konvention (Fn. 20), N 29.

32 Erläuternder Bericht zur KI-Konvention (Fn. 20), N 29.

nach in regelmässigen Abständen einen Bericht über die Tätigkeiten vor, die sie unternommen hat, um Artikel 3 Absatz 1 Buchstaben a und b Wirksamkeit zu verleihen (vgl. Art. 24 der Konvention, Ziff. 4.4.2).

4.2.3.1.2 Bedeutung für die Schweiz

Privatsektor

Wie oben erwähnt (vgl. Ziff. 4.2.3.1.1) ist es fraglich, ob rein administrative oder freiwillige Massnahmen im schweizerischen Recht ausreichen, um den Risiken und Auswirkungen, die sich aus den Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen im Privatsektor ergeben, in einer Weise anzugehen, die mit dem Ziel und dem Zweck der KI-Konvention vereinbar ist. Mangels anderer in Frage kommender Massnahmen geht die Analyse daher davon aus, dass der schweizerische Gesetzgeber Artikel 3 Absatz 1 Buchstabe b der Konvention umsetzen wird, indem er die Bestimmungen der Konvention auch auf den privaten Sektor anwendet.

Für die Regulierung des privaten Sektors besteht jedoch ein grösserer Spielraum. In Betracht zu ziehen sind hier Unterscheidungen zwischen öffentlichem und privatem Sektor, die die Besonderheit des schweizerischen Rechts berücksichtigen, insbesondere angesichts der Tatsache, dass im öffentlichen Sektor die rechtlichen Anforderungen grundsätzlich höher sind (siehe unten) und dass die Grundrechte grundsätzlich nur vertikal eine Wirkung haben.

Konkret bedeutet das für das schweizerische Recht, dass die Bestimmungen und Grundsätze der Konvention im Privatsektor dort zur Anwendung gelangen sollten, wo eine direkte oder indirekte horizontale Wirkung der Grundrechte besteht oder in Zukunft erkannt wird. Die KI-Konvention hat keine Erweiterung des Wirkungsbereichs der Grundrechte zwischen Privaten zur Folge, die über die Bestimmungen von Artikel 35 Absätze 1 und 3 BV hinaus geht (vgl. Ziff. 4.3.1.1). Die Anerkennung einer direkten oder indirekten horizontalen Wirkung der Grundrechte zwischen Privaten ist somit Voraussetzung, dass die Konvention für privatrechtliche Beziehungen in der Schweiz anwendbar ist.

Die Auswirkung der Konvention auf die Beziehungen zwischen Privaten könnte sich jedoch weiterentwickeln, insbesondere aufgrund von völkerrechtlichen Verpflichtungen, die die Schweiz eingegangen ist (z. B. das Internationale Übereinkommen zur Beseitigung jeder Form von Rassendiskriminierung[33], das Internationale Übereinkommen zur Beseitigung jeder Form von Diskriminierung der Frau[34] oder das Übereinkommen über die Rechte von Menschen mit Behinderungen[35]). Auch die Entwicklung der Rechtsprechung könnte eine Rolle spielen, einschliesslich der Rechtsprechung des EGMR zur Auslegung der EMRK und der sich daraus ergebenden positiven Verpflichtungen.

---

[33] SR **0.104**.

[34] SR **0.108**.

[35] SR **0.109**.

An dieser Stelle ist darauf hinzuweisen, dass bestimmte Beziehungen zwischen Privatpersonen von der Umsetzung der KI-Konvention überhaupt nicht betroffen sein könnten. Dies ist dann der Fall, wenn keine Anknüpfung an ein Grundrecht besteht. Zu denken ist hier an die klassische Gewährleistung für die Funktionsfähigkeit eines Geräts (z. B. ein Mähroboter, der nur den halben Garten mäht oder ein intelligenter Kühlschrank, der die Temperatur falsch einstellt) oder die mietrechtlichen Ansprüche des Mieters bei Mängeln. Mögliche wirtschaftliche Schäden aufgrund von Mängeln solcher KI-Systeme werden a priori keine Grundrechtsverletzung darstellen. Diese Frage lässt sich jedoch nicht eindeutig beantworten: Die Wirksamkeit eines Grundrechts im Verhältnis zwischen Privaten hängt letztlich von der Auslegung des jeweiligen Rechts, dem geschützten Rechtsgut, den Funktionen, die es erfüllt und den Umständen seiner Anwendung ab.[36]

Es sei darauf hingewiesen, dass die vorliegende Analyse sich bei der Frage nach einem allfälligen gesetzgeberischen Handlungsbedarf im Privatsektor nicht zum Regelungsort äussert. Es müssen nicht zwingend privatrechtliche Vorschriften (ZGB, OR usw.) sein, sondern es können auch öffentlich-rechtliche Vorschriften sein, die auf Privatpersonen anwendbar sind, oder gemischte Regelungen wie das DSG oder das GlG, die beide Arten von Normen enthalten.

Öffentlicher Sektor

Beim öffentlichen Sektor ist zu beachten, dass der Staat aufgrund seiner Funktion bestimmten Verpflichtungen unterliegt, die für den privaten Sektor nicht gelten und unterschiedliche Regelungen rechtfertigen können. Der wichtigste Grundsatz im schweizerischen Recht ist sicherlich das Legalitätsprinzip, wonach staatliches Handeln einer gesetzlichen Grundlage bedarf (Art. 5 und 164 BV; vgl. auch Art. 36 zur Einschränkung von Grundrechten). Dieser Grundsatz ist auch beim Einsatz von KI-Systemen zu beachten.[37]

Der Staat kann KI-Systeme in verschiedenen Bereichen und in unterschiedlichem Umfang einsetzen. Die Einsatzmöglichkeiten reichen von rein interner Unterstützung mit keiner oder nur geringer Aussenwirkung (z. B. Anwendungen zur automatisierten Aufgabenverteilung unter Mitarbeitenden, Anwendungen zum Übersetzen von Dokumenten oder Erstellen von Sitzungsprotokollen) über Systeme, welche die Verwaltung bei der Entscheidungsfindung unterstützen (Teilautomatisierung), offenbar die häufigste Form, bis hin zu Systemen, die selbst Entscheidungen treffen (Vollautomatisierung).[38]

---

[36] Vgl. CR Cst.-VINCENT MARTENET, Art. 35 N 77.

[37] NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 7 N 17.

[38] BJ, Totalrevision des Datenschutzgesetzes (DSG) – Übersicht zu den wichtigsten Änderungen für die Erarbeitung der Rechtsgrundlagen betreffend Datenbearbeitungen durch Bundesorgane, Oktober 2022, 23 f., abrufbar unter: www.bj.admin.ch > Staat & Bürger > Datenschutz > Informationen für Bundesorgane > Totalrevision des Datenschutzgesetzes (DSG) – Übersicht zu den wichtigsten Änderungen für die Erarbeitung der Rechtsgrundlagen betreffend Datenbearbeitungen durch Bundesorgane (abgerufen am 27. August 2024).

Die Beantwortung der Frage, ob, auf welcher Normstufe und in welcher Normdichte eine ausdrückliche gesetzliche Grundlage erforderlich ist, hängt im Wesentlichen von der Bedeutung der behandelten Materie (Art. 164 BV) bzw. von allfälligen Grundrechtseinschränkungen (Art. 36 BV) ab.[39] Interne Assistenzsysteme ohne oder mit geringer Aussenwirkung sind grundsätzlich der Verwaltungsunterstützung zuzuordnen und haben ihre Legitimation in der gesetzlichen Grundlage, welche die Aufgaben der Behörde regelt (z. B. Anwendungen, welche die Aufgaben automatisch unter den Mitarbeitenden verteilen, Dokumente übersetzen oder Sitzungsprotokolle erstellen).[40] Dies dürfte auch für *Chatbots* gelten, sofern sie sich auf die Erteilung von Informationen und Auskünften beschränken und für die Bearbeitung von Personendaten eine gesetzliche Grundlage besteht. In diesem Fall kann davon ausgegangen werden, dass es sich um die Wahl eines Kommunikationskanals handelt, bei der die Behörden grundsätzlich frei sind.[41] Hingegen bedürfen ganz oder teilweise automatisierte Entscheide einer formell gesetzlichen Grundlage, wenn sie schwerwiegend in die Grundrechte der betroffenen Personen eingreifen können (Art. 36 Abs. 1 BV und Art. 164 Abs. 1 Bst. b BV)[42] und/oder wenn es sich um grundlegende Bestimmungen handelt, die die Organisation und das Verfahren der Bundesbehörden betreffen (Art. 164 Abs. 1 Bst. g BV). Eine Rechtsgrundlage in einem Gesetz im materiellen Sinn dürfte z. B. zumindest dann erforderlich sein, wenn die Unterstützung so weit geht, dass der Algorithmus, insbesondere im Rahmen eines Verwaltungsverfahrens, einen Text vorbereitet, den der Mensch nur noch zu überprüfen hat.[43]

Im Datenschutzbereich hat der Gesetzgeber das Erfordernis der gesetzlichen Grundlage in Artikel 34 DSG konkretisiert. Gemäss dieser Bestimmung, die eine Konkretisierung von Artikel 36 BV darstellt, dürfen Bundesorgane, von Ausnahmen abgesehen (vgl. Art. 34 Abs. 4 und Art. 36 Abs. 2 DSG), Personendaten nur bei Vorliegen einer gesetzlichen Grundlage bearbeiten. Absatz 2 sieht vor, dass eine Grundlage in einem Gesetz im formellen Sinn in folgenden Fällen erforderlich ist: bei der Bearbeitung von besonders schützenswerten Personendaten (Bst. a), beim Profiling (Bst. b) und wenn der Bearbeitungszweck oder die Art der Datenbearbeitung zu einem schwerwiegenden Eingriff in die Grundrechte der betroffenen Person führen können (Bst. c). KI-Systeme, die Personendaten bearbeiten, unterliegen selbstverständlich diesen Anforderungen. KI-Systeme, die eine vollständig oder teilweise automatisierte Einzelentscheidung generieren, können unter Buchstabe c fallen, sofern sie nicht bereits von Artikel 34 Buchstaben a und b erfasst sind. Dies ist abhängig von der Schwere des allfälligen Grundrechtseingriffs, insbesondere im Bereich des Schutzes der Privatsphäre und des Datenschutzes (Art. 13 BV). Generell gilt: Je intransparenter die Art der Bearbeitung, je grösser der Ermessensspielraum und je mehr Daten aus verschiedenen Quellen zusammengetragen werden, desto höher ist das Risiko einzuschätzen und desto eher bedarf es einer Grundlage in einem Gesetz im formellen Sinn.[44]

---

[39] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen, Schlussbericht vom 28. Februar 2021 zum Vorprojekt IP6.4, 34.

[40] CATHERINE REITER, Künstliche Intelligenz im Verwaltungsverfahren – Ermessen als Stolperstein?, AJP 2022, 984 ff., 988; PHILIP GLASS, Datenschutzrecht für künstliche Intelligenz in der öffentlichen Verwaltung: eine Auslegeordnung am Beispiel des Kantons Zürich, in: Michael Widmer (Hrsg.), Datenschutz: Rechtliche Schnittstellen, Zürich 2023, 177 ff., 208.

[41] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al. (Fn. 39), 58 f. Dies bedeutet jedoch nicht, dass sich nicht auch andere rechtliche Fragen stellen, insbesondere im Zusammenhang mit der Einhaltung von Treu und Glauben (Art. 9 BV), vgl. für weitere Entwicklungen zu diesem Thema NADJA BRAUN BINDER/LILIANE OBRECHT/GRACE WITTMER, Vertrauensschutz bei fehlerhaften Behördenauskünften durch Chatbots, IusNet DigR 2024; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 5 N 10 ff., 18 N 45 ff.

[42] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (Fn. 39), 59.

[43] CATHERINE REITER, Künstliche Intelligenz im Verwaltungsverfahren (Fn. 40), 984 ff., 988.

[44] MONIQUE COSSALI SAUVAIN, in: Yaniv Benhamou/Bertil Cottier (Hrsg.), Petit commentaire LPD, Art. 34 N 35–37; BJ, Gesetzgebungsleitfaden Datenschutz – Auswirkungen des neuen Datenschutzgesetzes auf die Erarbeitung von Rechtsgrundlagen, Bern 2024, Kap. 3.2.3, 21, abrufbar

4.2.3.2 Nationale Sicherheit

Artikel 3 Absatz 2 ist das Ergebnis eines Konsenses zwischen den Staaten, die den Bereich der nationalen Sicherheit vom Anwendungsbereich der KI-Konvention ausnehmen wollten, und den Staaten, die eine vollständige Einbeziehung befürworteten.

Der Text sieht nun vor, dass die Vertragsparteien die KI-Konvention auf Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen für den Schutz ihrer nationalen Sicherheitsinteressen anwenden können, aber nicht müssen, unabhängig davon, welche Stelle diese Tätigkeiten ausführt. In jedem Fall müssen die betreffenden Tätigkeiten im Einklang mit dem geltenden Völkerrecht stehen, da die nationale Sicherheit in den Anwendungsbereich vieler Übereinkommen fällt. Der erläuternde Bericht nennt als Beispiele die EMRK und die beiden Pakte der Vereinten Nationen[45] und stellt darüber hinaus klar, dass diese Tätigkeiten im Einklang mit den demokratischen Institutionen und Verfahren der Vertragsparteien erfolgen müssen.[46]

Nationale Sicherheit gilt als ein «im Wesentlichen umstrittener» Begriff, d. h. als ein Konzept, über dessen Inhalt auf internationaler Ebene kein Konsens erzielt werden kann.[47] Diese Situation eröffnet den Staaten einen grossen Handlungsspielraum und führt in der Rechtswissenschaft zu einer Vielfalt von Definitionsvorschlägen.[48]

Intern hat der Bundesrat im Rahmen eines breiten Begriffsverständnisses der nationalen Sicherheit folgende Klassifikation entwickelt: 1) Polizeiliche Gefahrenabwehr, Staatsschutz und Strafverfolgung; 2) Vorbeugung, Vorsorge und Bewältigung von natur- und zivilisationsbedingten Katastrophen und Notlagen; 3) Abhalten und Abwehr eines militärischen Angriffs; 4) Wahrung der Interessen der Schweiz im Ausland und Beiträge zum internationalen Krisenmanagement.[49]

---

unter: www.bj.admin.ch > Datenschutz > Informationen für Bundesorgane > Gesetzgebungsleitfaden Datenschutz (abgerufen am 27. August 2024).

45 Internationaler Pakt über wirtschaftliche, soziale und kulturelle Rechte (UNO-Pakt I, SR **0.103.1**) und Internationaler Pakt über bürgerliche und politische Rechte (UNO-Pakt II, SR **0.103.2**).

46 Erläuternder Bericht zur KI-Konvention (Fn. 20), N 32.

47 BSK BK-OLIVER DIGGELMANN/TILMANN ALTWICKER, Art. 57, N 9 und angegebene Referenz.

48 FRANCK ELONG MBOULÉ, Le régime juridique des biens des organisations internationales, Zürich 2022, 294 ff., 296.

49 Bericht des Bundesrates an die Bundesversammlung über die Sicherheitspolitik der Schweiz vom 23. Juni 2010, BBl **2010** 5133 5158 f.

Für die Parlamentarische Versammlung des Europarats besteht die nationale Sicherheit in der Abwehr jeder sichtbaren und tatsächlichen Bedrohung der demokratischen Ordnung des Staates und der Gesellschaft.[50] Diese Definition wird jedoch von den einzelnen europäischen Staaten nicht geteilt, da sie die Bedürfnisse der Nachrichtendienste bei der Bekämpfung von verdeckten oder spekulativen Bedrohungen nicht berücksichtigt.[51]

Die KI-Konvention soll diese Frage nicht klären, aber der erläuternde Bericht stellt klar, dass routinemässige Tätigkeiten zur Aufrechterhaltung der öffentlichen Ordnung im Rahmen von Prävention, Aufdeckung, Untersuchung und Verfolgung von Straftaten, einschliesslich von Bedrohungen der öffentlichen Sicherheit, nicht betroffen sind, sofern die nationale Sicherheit der Vertragsparteien nicht auf dem Spiel steht.[52] Nach Auslegung des BJ unterscheiden sich diese regulären Tätigkeiten von den geheimen Tätigkeiten der Nachrichtendienste, die grundsätzlich unter die Ausnahmebestimmung der nationalen Sicherheit fallen.

Zu beachten ist, dass die Ausnahme nicht für Aktivitäten von KI-Systemen mit doppeltem Verwendungszweck gilt, sofern auch ein anderer Zweck als die Wahrung der nationalen Sicherheit verfolgt wird.[53]

In den Verhandlungen zur KI-Konvention hat sich die Schweizer Delegation dafür ausgesprochen, den Bereich der nationalen Sicherheit vollumfänglich in den Geltungsbereich des Konventionstextes einzubeziehen. Es gibt keinen Grund, im Rahmen der vorliegenden Analyse von dieser Position abzuweichen. Zudem gelten im schweizerischen Recht auch in diesem Bereich die Grundrechte, wobei Einschränkungen unter den üblichen Voraussetzungen von Art. 36 BV, insbesondere bei Vorliegen eines überwiegenden öffentlichen Interesses, möglich sind. In der vorliegenden Analyse wird daher von der Hypothese ausgegangen, dass die Tätigkeiten von KI-Systemen im Zusammenhang mit dem Schutz der nationalen Sicherheit, insbesondere den Tätigkeiten des NDB, in den Anwendungsbereich der Konvention fallen.

4.2.3.3 Forschung und Entwicklung

Forschungs- und Entwicklungstätigkeiten sind unter bestimmten Voraussetzungen vom Anwendungsbereich der Konvention ausgenommen. Voraussetzung ist, dass die betreffenden KI-Systeme noch nicht zur Nutzung freigegeben sind und dass die Erprobung oder vergleichbare Aktivitäten so ausgelegt sind, dass sie Menschenrechte, Demokratie und Rechtsstaatlichkeit nicht beeinträchtigen (Art. 3 Abs. 3).

---

[50] Empfehlung 1402 (1999) der parlamentarischen Versammlung des Europarats, Kontrolle der internen Sicherheitsdienste in den Mitgliedstaaten des Europarats, Punkt A. 2.

[51] CR Cst.-OLIVIER BLEICKER, Art. 57 N 34 und Referenz in Fussnote 86.

[52] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 32 *in fine*.

[53] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 32.

Gemäss dem erläuternden Bericht sollten diese Tätigkeiten in jedem Fall den geltenden Menschenrechten und dem innerstaatlichen Recht sowie den anerkannten ethischen und professionellen Standards der wissenschaftlichen Forschung entsprechen. KI-Systeme, die nach der Forschungs- und Entwicklungsphase zur Nutzung freigegeben werden, sollten grundsätzlich mit der Konvention vereinbar sein, auch in Bezug auf ihre Konzeption und Entwicklung.[54]

Zu beachten ist, dass diese Ausnahme vom Anwendungsbereich nicht die Verpflichtungen aus Artikel 13 (Sichere Innovation) und Artikel 25 Absatz 2 (Internationale Zusammenarbeit) betrifft.

4.2.3.4 Nationale Verteidigung

Der Bereich der nationalen Verteidigung ist vom Anwendungsbereich der Konvention ausgenommen (Art. 3 Abs. 4). Dies entspricht dem in Artikel 1 Buchstabe d der Satzung des Europarats verankerten Grundsatz, dass Fragen der nationalen Verteidigung nicht in die Zuständigkeit der Organisation fallen.

Dies bedeutet aber keinesfalls, dass Tätigkeiten, die innerhalb des Lebenszyklus von KI-Systemen im Zusammenhang mit der nationalen Verteidigung durchgeführt werden, nicht unter das Völkerrecht fallen.[55] Wie bei der nationalen Sicherheit gelten in der Schweiz auch in diesem Bereich die Grundrechte. Daher sollten keine Ausschlüsse vorgesehen werden, sondern allenfalls Ausnahmen oder Anpassungen, wo dies im Interesse der nationalen Verteidigung der Schweiz gerechtfertigt sein könnte.

4.2.3.5 Zusammenfassung

Die Hauptelemente des Anwendungsbereichs der Konvention gemäss Artikel 3 lassen sich wie folgt zusammenfassen:

- Öffentlicher / privater Sektor (Art. 3 Abs. 1):

  Die Konvention gilt für beide Sektoren. Im Privatsektor haben die Vertragsparteien jedoch einen grösseren Handlungsspielraum, sofern sie die Risiken und Auswirkungen, die während des Lebenszyklus von KI-Systemen auftreten, in einer Weise angehen, die mit dem Ziel und dem Zweck des Übereinkommens vereinbar ist.

  Bei einer Ratifizierung würde die Schweiz die Grundsätze und Verpflichtungen der Kapitel II bis VI der Konvention auch auf den Privatsektor anwenden. Rein administrative oder freiwillige Massnahmen scheinen angesichts unserer Rechtsordnung nicht ausreichend. Zu berücksichtigen wären Unterscheidungen zwischen den beiden Sektoren, die den Besonderheiten des schweizerischen Rechts Rechnung tragen, insbesondere im Hinblick auf die grundsätzlich nur vertikale Wirkung der

[54] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 33 f.

[55] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 36.

Grundrechte und unter Berücksichtigung der Tatsache, dass die rechtlichen Anforderungen im öffentlichen Sektor grundsätzlich höher sind.

- Nationale Sicherheit (Art. 3 Abs. 2):

  Die Staaten sind nicht zur Anwendung der Konvention im Bereich der nationalen Sicherheit verpflichtet. Über den Geltungsbereich dieses Begriffs besteht kein internationaler Konsens. Er bezieht sich vor allem auf geheime nachrichtendienstliche Tätigkeiten. Für die Zwecke der vorliegenden Analyse geht das BJ davon aus, dass die Bestimmungen der Konvention auch in diesem Bereich Anwendung finden.

- Forschung und Entwicklung (Art. 3 Abs. 3):

  Die Konvention gilt nicht für Forschungs- und Entwicklungstätigkeiten, es sei denn, die Versuche werden in einer Weise durchgeführt, die die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit beeinträchtigen könnte. Artikel 13 (Sichere Innovation) und Artikel 25 Absatz 2 (Internationale Zusammenarbeit) der KI-Konvention bleiben von dieser Ausnahme unberührt.

- Nationale Verteidigung (Art. 3 Abs. 4):

  Fragen der nationalen Verteidigung fallen nicht in die Zuständigkeit des Europarats und sind daher nicht Gegenstand der Konvention.

4.3 Pflichten und Grundsätze

4.3.1 Kapitel II: Allgemeine Verpflichtungen

4.3.1.1 Artikel 4 – Schutz der Menschenrechte

Nach Artikel 4 stellt jede Vertragspartei durch die Einführung oder Beibehaltung von Massnahmen sicher, dass die Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen im Einklang mit den Verpflichtungen zum Schutz der Menschenrechte gemäss anwendbarem Völkerrecht und innerstaatlichem Recht stehen.

Die Vertragsparteien müssen sicherstellen, dass der bestehende rechtliche Rahmen einen ebenso wirksamen wie effizienten Schutz der Menschenrechte gewährleistet, sowohl im spezifischen Kontext von KI-Systemen als auch ausserhalb solcher Systeme.

Das eigentliche Ziel der KI-Konvention besteht darin, zu gewährleisten, dass die Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen vollumfänglich mit der Achtung der Menschenrechte vereinbar sind. Zu diesem Zweck sind spezifische Verpflichtungen vorgesehen, insbesondere die Verpflichtung, einen Rahmen für das Risiko- und Folgenmanagement von KI-Systemen hinsichtlich Menschenrechte, Demokratie und Rechtsstaatlichkeit zu schaffen (vgl. Ziff. 4.3.4). Mit der Umsetzung dieser Verpflichtungen kommen die Vertragsparteien indirekt der allgemeinen Verpflichtung zum Schutz der Menschenrechte nach Artikel 4 der KI-Konvention nach.

Im schweizerischen Recht sind, wie oben erwähnt (vgl. Ziff. 3), beim Einsatz von KI-Systemen bereits verschiedene völkerrechtliche Instrumente zum Schutz der Menschenrechte an-

wendbar, insbesondere die EMRK. Im innerstaatlichen Recht gelten die Grundrechtsgarantien auch im spezifischen Kontext von KI-Systemen. Für den öffentlichen Sektor gilt insbesondere die in Artikel 36 BV verankerte Regelung, wonach Grundrechtseinschränkungen grundsätzlich vier kumulative Voraussetzungen (gesetzliche Grundlage, öffentliches Interesse, Verhältnismässigkeit und Wahrung des Kerngehalts des Grundrechts) erfüllen müssen. Dabei spielt der Ursprung des staatlichen Handelns (Mensch oder KI-System) für den Grundrechtsschutz keine Rolle.

Für den Privatsektor ist dabei zu beachten, dass nach Artikel 35 Absatz 1 BV die Grundrechte in der gesamten Rechtsordnung zur Geltung kommen. Nach Artikel 35 Absatz 3 BV sorgen die Behörden dafür, dass die Grundrechte, soweit sie sich dazu eignen, auch unter Privaten wirksam werden. Diese Bestimmung bestätigt, dass die Grundrechte ihre Wirkung zwischen Privaten grundsätzlich nicht unmittelbar entfalten können, sondern dass es zu ihrer Geltung eines besonderen Eingreifens der Behörden bedarf.[56] Dies gilt sowohl für den Erlass von Rechtsnormen als auch für deren Auslegung und Anwendung. Der einzige Fall, in dem die BV eine direkte horizontale Wirkung anerkennt, ist Artikel 8 Absatz 3 Satz 3 BV, der den Grundsatz der Lohngleichheit zwischen Mann und Frau garantiert.

Somit ist für die Umsetzung der in Artikel 4 der Konvention enthaltenen Verpflichtung zum Schutz der Menschenrechte in innerstaatliches Recht zu berücksichtigen, dass die Grundrechte im Schweizer Recht zwischen Privaten in der Regel nur eine indirekte horizontale Wirkung entfalten.

Aus dem Gesagten ergibt sich, dass der schweizerische Rechtsrahmen für den Schutz der Grundrechte beim Einsatz von KI-Systemen, insbesondere im öffentlichen Sektor, bereits zur Anwendung kommt. Die Konvention verlangt, dass dieser Rahmen ausreichend und wirksam ist, um den besonderen Herausforderungen von KI-Systemen gerecht zu werden. Das eigentliche Ziel aller in der KI-Konvention enthaltenen Regeln und allgemeinen Grundsätze ist die Stärkung des Grundrechtsschutzes. Bei der Prüfung der Wirksamkeit des Schutzes im schweizerischen Recht und des allfälligen Handlungsbedarfs wird deshalb auf die nachfolgenden Ausführungen Bezug genommen.

4.3.1.2 Artikel 5 – Integrität demokratischer Prozesse und Achtung der Rechtsstaatlichkeit

Nach Artikel 5 Absatz 1 stellt jede Vertragspartei durch die Einführung oder Beibehaltung von Massnahmen sicher, dass KI-Systeme nicht eingesetzt werden, um die Integrität, Unabhängigkeit und Effektivität demokratischer Institutionen und Prozesse zu untergraben. Dies schliesst die Gewaltenteilung, die Achtung der Unabhängigkeit der Justiz und den Zugang zur Justiz ein. Nach Artikel 5 Absatz 2 ergreift oder beibehält jede Vertragspartei Massnahmen zum Schutz und zur Aufrechterhaltung ihrer demokratischen Prozesse während der Tätigkei-

[56] GIORGIO MALINVERNI/MICHEL HOTTELIER/MAYA HERTIG RANDALL/ALEXANDRE FLÜCKIGER, Droit constitutionnel suisse – Volume II: Les droits fondamentaux, Bern 2021, N 135.

ten innerhalb des Lebenszyklus von KI-Systemen, einschliesslich des gleichberechtigten Zugangs und der gleichberechtigten Teilhabe einer Person am öffentlichen Diskurs und der Möglichkeit zur freien Meinungsbildung.

Allerdings werden die Begriffe «demokratische Institutionen und Prozesse» und «Rechtsstaatlichkeit» im Konventionstext nicht definiert und bleiben daher vage. In der Analyse wird diese Bestimmung definiert als Verpflichtung, Massnahmen zu ergreifen oder beizubehalten, um Demokratie und Rechtsstaatlichkeit als strukturelle Prinzipien demokratischer Länder zu schützen.

KI-Systeme können sehr unterschiedliche Auswirkungen auf demokratische Institutionen, Prozesse und die Rechtsstaatlichkeit haben. Wie im erläuternden Bericht ausgeführt, können sich die Vertragsparteien bei der Erfüllung ihrer Verpflichtungen nach Artikel 5 auf bestimmte von KI ausgehende Risiken konzentrieren, z. B. auf das Risiko einer Beeinträchtigung des politischen Pluralismus.[57] Die Konvention verweist weder auf ein bestimmtes zu berücksichtigendes Risiko, noch gibt sie eine zu ergreifende Massnahme vor. Die Bestimmung ist zwar weit gefasst, hat aber den Vorteil, dass sie zukunftssicher ist, da sie auch bis anhin noch nicht bekannte Risiken umfassen kann.

Nachfolgend werden Beispiele für negative Auswirkungen von KI-Systemen auf die Integrität demokratischer Prozesse und die Wahrung der Rechtsstaatlichkeit anhand ausgewählter Bestimmungen des schweizerischen Rechts aufgezeigt.

- Desinformation und Falschinformationen

  KI-Systeme können das Phänomen der Desinformation und der Verbreitung von Falschinformationen, deren Ziel es ist, die Öffentlichkeit zu falschen Annahmen zu verleiten, verstärken. Desinformation kann die öffentliche Meinung lenken und Wahlen und Abstimmungen beeinflussen. Falschinformationen können verschiedene Formen annehmen, z. B. Behauptungen, unbegründete Meinungsäusserungen oder Hassreden gegen gesellschaftliche Gruppen oder Minderheiten. Automatisierte Software («Bots»), die menschliches Verhalten in sozialen Medien nachahmen, indem sie Informationen posten, Beiträge «liken» oder sich an reale Personen wenden, können beispielsweise durch die Verbreitung von Hassreden die öffentliche Meinung polarisieren. KI-Systeme können auch dazu eingesetzt werden, Stimmen und Bilder zu imitieren, um alternative Realitäten bzw. «Deepfakes» zu schaffen und einzusetzen.[58]

---

[57] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 46.

[58] YVES-MARIE DOUBLET, Désinformation et campagnes électorales, Bericht zuhanden des Europarats, Juni 2019, 5, 11 und 16; siehe auch MURAT KARABOGA/NULA FREI et al., Deepfakes und manipulierte Realitäten. Technologiefolgenabschätzung und Handlungsempfehlungen für die Schweiz, Ta-Swiss, TA 81/2024, Zollikon:vdf.

Dabei verschwimmt die Unterscheidung zwischen Fiktion und Realität zunehmend. Darüber hinaus können erfundene und irreführende Informationen über Chatbots verbreitet werden, was zur Desinformation beiträgt und je nach Kontext zu einem Problem für die Demokratie werden kann.[59]

Die schweizerische Gesetzgebung enthält bereits einschlägige Normen, die unter bestimmten Umständen bei Desinformation und Fehlinformation zur Anwendung kommen:

- Bei Wahlen und Abstimmungen schützt Artikel 34 Absatz 2 BV die freie Willensbildung und die unverfälschte Stimmabgabe der Stimmbürgerinnen und Stimmbürger. Das BPR enthält spezifische Bestimmungen für Wahlen und Abstimmungen auf Bundesebene; so kann gemäss Artikel 77 Absatz 1 Buchstabe b BPR gegen Unregelmässigkeiten bei Abstimmungen Beschwerde erhoben werden.

  Nach Rechtsprechung des BGer darf ein Abstimmungsergebnis, das den freien Willen der Stimmberechtigten nicht zuverlässig und unverfälscht zum Ausdruck bringt, nicht anerkannt werden. Konnte der freie Wille nicht zuverlässig und unverfälscht zum Ausdruck gebracht werden, ist die Abstimmung zu wiederholen. Diese Bestimmungen sollen in erster Linie eine unzulässige Einflussnahme des Staates verhindern. Aber auch Private können Fehlinformationen verbreiten und die freie Willensbildung der Stimmberechtigten beeinflussen. Dies ist beispielsweise dann der Fall, wenn in einem so späten Zeitpunkt mit offensichtlich unwahren und irreführenden Angaben in den Abstimmungskampf eingegriffen wird, dass es den Stimmberechtigten nach den Umständen nicht mehr möglich ist, sich aus anderen Quellen ein zuverlässiges Bild von den tatsächlichen Verhältnissen zu machen. In Anbetracht der Meinungsäusserungsfreiheit wird eine derartige Beeinträchtigung der Stimmberechtigten nicht leichthin angenommen werden. Die Aufhebung einer Abstimmung fällt daher nur unter grösster Zurückhaltung und bei ganz schwerwiegenden Verstössen in Betracht.[60]

  Weiter anerkennt das BGer die Richtigstellung offensichtlich falscher oder irreführender Informationen als triftigen Grund für die Rechtfertigung von behördlichem Eingreifen in den Abstimmungskampf. Unter bestimmten Umständen besteht sogar eine Pflicht der Behörden, zur Sicherstellung des

---

[59] Vgl. die von Algorithmwatch und Forensics durchgeführte Studie, in der die Antworten eines Chatbots im Zusammenhang mit den Wahlen in der Schweiz, Bayern und Hessen im Oktober 2023 getestet wurden, abrufbar unter: https://algorithmwatch.ch > Publikationen > KI-Chatbot liefert falsche Antworten auf Fragen zu demokratischen Wahlen (abgerufen am 27. August 2024). Siehe auch die kürzliche Studie vom August 2024 im Zusammenhang mit den Regionalwahlen in Deutschland, abrufbar unter: https://algorithmwatch.org/de > Blog > Chatbots bringen noch immer viele Falschinformationen in Umlauf (abgerufen am 28. August 2024).

[60] BGE **135** I 292, E. 4.1.

bundesrechtlichen Anspruchs auf eine freie Willensbildung und unverfälschte Stimmabgabe zu intervenieren. Dabei steht den Behörden jedoch ein grosser Ermessensspielraum zu. Eine Interventionspflicht, deren Verletzung zur Aufhebung der Abstimmung führen kann, wird daher grundsätzlich nur anzunehmen sein, wenn die Einflussnahme privater Akteure die Willensbildung der Stimmberechtigten in ganz schwerwiegender Art beeinträchtigt oder geradezu verunmöglicht.[61]

Die obigen Überlegungen gelten auch für den Einsatz von KI-Systemen zur Desinformation. Die Verbreitung eines verzerrten Bildes kann in bestimmten Kontexten beispielsweise zur politischen Manipulation eingesetzt werden.[62]

Eine schwerwiegende Täuschung über den Gegenstand einer Abstimmung in den sozialen Medien durch Privatpersonen wirft erstens die Frage auf, ob die Behörden zum Eingreifen und zur Korrektur bzw. Berichtung der betreffenden Informationen verpflichtet sind.[63] Zweitens ist im Hinblick auf eine mögliche Annullierung der Abstimmung daran zu erinnern, dass die Rechtsprechung nicht den Nachweis eines Einflusses der Unregelmässigkeiten auf das Ergebnis der Abstimmung verlangt; es genügt vielmehr, dass ein solcher Einfluss möglich erscheint. In Ermangelung einer quantifizierten Feststellung der Auswirkungen einer Unregelmässigkeit im Verfahren wird ihr Einfluss auf das Ergebnis anhand der Gesamtumstände und grundsätzlich mit voller Kognition beurteilt. [64] Diese Erleichterung könnte sich als besonders nützlich erweisen, wenn Desinformation online über KI-Systeme verbreitet wird.

Die genannten Bestimmungen zu den politischen Rechten greifen jedoch erst nach einer Verletzung, und erlauben keine präventive Regulierung der Strukturen der Informationsproduktion und -verbreitung an sich.

- In der öffentlichen Kommunikation und Meinungsbildung nehmen Online-Plattformen heutzutage eine immer bedeutendere Stellung ein. Plattformen tragen damit einerseits zu einer besseren Verwirklichung der Meinungsäusserungs- und Informationsfreiheit bei, andererseits beeinflussen sie auch die öffentliche

---

[61] Urteil des BGer, 1C_472/2010, E. 4.3; NADJA BRAUN BINDER/MANUELA KÄLIN, Rechtliche Aspekte der politischen Meinungsbildung, in: Urs Bieri et al. (Hrsg.), Digitalisierung der Schweizer Demokratie. Technologische Revolution trifft auf traditionelles Meinungsbildungssystem, Ta-Swiss, TA 75/2021, Zürich:vdf, 125 ff., 129.

[62] Für eine Darstellung der verschiedenen möglichen Gefahren vgl. TOM LEBRUN, La liberté d'expression aux Etats-Unis et au Canada face au risque de propagande générée par l'intelligence artificielle, in: Céline Castets-Renard/Jessica Eynard (Hrsg.), Un droit de l'intelligence artificielle. Entre règles sectorielles et régime général. Perspectives comparées, Bruxelles, 2023, 409 ff.

[63] NADJA BRAUN BINDER/MANUELA KÄLIN, Rechtliche Aspekte der politischen Meinungsbildung (Fn. 61), 145.

[64] BGE **135** I 292, E. 4.4.

Debatte. Zudem hat die niederschwellige Möglichkeit zur Produktion und Diffusion von Inhalten verschiedene negative Auswirkungen zur Folge.

Neben den allgemeinen Rechtsregeln bestehen bislang keine spezifischen gesetzlichen Regelungen oder eine ausführliche Rechtspraxis bezüglich der Verantwortlichkeit von Intermediären für fremde widerrechtliche Inhalte auf ihren Plattformen. Auch bestehen keine bzw. nur punktuelle Vorschriften, welche die Rechte der Nutzerinnen und Nutzer gegenüber den Intermediären stärken oder die Plattformen zu mehr Transparenz verpflichten.

Aus diesem Grund hat der Bundesrat das UVEK (BAKOM) damit beauftragt, eine Vernehmlassungsvorlage für die Regulierung von grossen Online-Plattformen auszuarbeiten. Diese soll gewisse Sorgfaltspflichten für Plattformen enthalten. So sollen diese beispielsweise Melde- und Abhilfeverfahren für rechtswidrige Hassrede einrichten sowie eine Kontaktstelle und einen Rechtsvertreter in der Schweiz benennen müssen. Ebenfalls sollen die Rechte von Nutzenden gestärkt werden, z. B. indem Plattformen Massnahmen gegen Nutzende begründen und ihnen ein Beschwerdeverfahren zur Verfügung stellen müssen.

Verschiedene dieser Aspekte werden nicht zuletzt auch algorithmische Systeme betreffen, welche von Plattformen eingesetzt werden (z. B. zur Moderation oder zum Ranking von Inhalten). Entsprechend wird auch diese Regulierung zum Schutz demokratischer Prozesse (in diesem Fall der öffentlichen Meinungsbildung) und dem Rechtsstaat beitragen.

- Die Identifizierung von KI-generierten Inhalten ist ebenfalls ein Mittel zur Bekämpfung von Online-Desinformation. Eine solche Identifizierung schränkt die Verbreitung von Desinformation zwar nicht per se ein, ermöglicht den Empfängern aber eine bessere Beurteilung und Einordnung der erhaltenen Informationen. In diesem Sinne ist die Umsetzung der in Artikel 8 der KI-Konvention verankerten Transparenzpflicht im Zusammenhang mit dem Schutz der Integrität demokratischer Prozesse und der Rechtsstaatlichkeit von besonderer Relevanz. In der Analyse wird nachfolgend darauf eingegangen (vgl. Ziff. 4.3.2.3). Mit einer Kennzeichnung könnten die Nutzerinnen und Nutzer darüber aufgeklärt werden, dass ein Account oder eine Publikation nicht von einem Menschen stammt.

- Auch das Strafrecht bietet einen rechtlichen Rahmen, der unter bestimmten Voraussetzungen einschlägig sein und bestimmten strafrechtlich relevanten Verhaltensweisen Grenzen setzen kann. Die Bestimmungen des StGB sind grundsätzlich technologieneutral und gelten unabhängig vom Tatmittel des Täters. Das Strafrecht bietet daher einen allgemeinen Rahmen, der bestimmte Verbrechen und Vergehen gegen den öffentlichen Frieden (Art. 258 ff. StGB) und Vergehen gegen den Volkswillen (Art. 279 ff. StGB) erfasst. So ist beispielsweise Artikel 261bis StGB über Diskriminierung und Aufruf zu Hass unter bestimmten Voraussetzungen auf Hassreden in sozialen Netzwerken anwendbar. Gleiches gilt für Artikel 173 ff. StGB über die Ehrverletzungsdelikte oder Artikel 179^{decies} StGB über Identitätsmissbrauch (vgl. zu diesen Bestimmungen Ziff. 6.6). Zivilrechtlich kann auch der Persönlichkeitsschutz (Art. 28 ff. ZGB)

zur Anwendung kommen (vgl. zu den Überlegungen im Zusammenhang mit Deepfakes Ziff. 6.3.3).

Bei der Anwendung dieser Normen ist die in Artikel 16 Absatz 2 BV verankerte Meinungsfreiheit zu beachten. Diese gilt auch für falsche und irreführende Behauptungen.[65] Einschränkungen bleiben möglich, müssen aber den Anforderungen von Artikel 36 BV genügen.[66]

- Targeting zu politischen Zwecken

Im Zusammenhang mit den Risiken für die freie Meinungsbildung der Bürgerinnen und Bürger ist auch die Frage des Targetings zu politischen Zwecken durch den Einsatz von KI-Systemen und die Aspekte des Datenschutzes zu nennen. Mithilfe von Algorithmen lässt sich das Verhalten der Bürgerinnen und Bürger beeinflussen, z. B. durch das gezielte Versenden von Informationen an Gruppen, von denen ein Profil erstellt wurde, in der Annahme, dass deren Mitglieder besonders sensibel auf eine bestimmte Art von Nachrichten reagieren.

a) Rechtlicher Rahmen gemäss DSG

Im schweizerischen Recht ist in diesem Zusammenhang insbesondere der durch das neue, am 1. September 2023 in Kraft getretene DSG geschaffene rechtliche Rahmen zu beachten.[67]

Die Bearbeitung von Daten über die politische Einstellung der Bürgerinnen und Bürger stellt eine Bearbeitung von besonders schützenswerten Personendaten dar (Art. 5 Abs. 1 Bst. c DSG). Diese Daten geniessen somit einen qualifizierten Schutz. Wenn Daten aus verschiedenen Quellen mittels automatisierter Analysemethoden miteinander verknüpft werden, um politische Meinungen und Tendenzen der Stimmbürgerinnen und Stimmbürger zu ermitteln, handelt es sich um eine Profiling-Tätigkeit (Art. 5 Abs. 1 Bst. f DSG). Je nach Umständen kann es sich dabei sogar um ein Profiling mit hohem Risiko handeln (Art. 5 Abs. 1 Bst. g DSG).

Auch hier sind die allgemeinen Grundsätze von Treu und Glauben, Transparenz, Verhältnismässigkeit, Zweckbindung, Richtigkeit und Datensicherheit zu beachten. Wenn

---

65 FLORENT THOUVENIN/STEPHANIE VOLZ/MARK EISENEGGER/DANIEL VOGLER/MARIELA JAFFÉ, Governance von Desinformation in digitalisierten Öffentlichkeiten, Jusletter 5. Februar 2024, 9 N 24, und die angegebenen Ref.

66 RAPHAELA CUENI, Falsche und irreführende Informationen im Verfassungsrecht der Schweiz, ex ante 1/2019, 3 ff., insbesondere 9 ff.

67 Vgl. zu diesem Thema den Leitfaden vom 15. Dezember 2022 der Datenschutzbehörden des Bundes und der Kantone zur digitalen Bearbeitung von Personendaten im Rahmen von Wahlen und Abstimmungen in der Schweiz, abrufbar unter: https://www.edoeb.admin.ch > Datenschutz > Internet & Technologie > Leitfaden zu Wahlen und Abstimmungen (abgerufen am 27. August 2024); siehe auch MICHAEL MONTAVON, Big Data, un outil d'influence en période électorale, 2023, www.swissprivacy.law/258.

die geplante Bearbeitung ein hohes Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person mit sich bringen kann, muss der für die Bearbeitung Verantwortliche vorgängig eine Datenschutz-Folgenabschätzung durchführen (Art. 22 Abs. 1 DSG). Ob ein hohes Risiko besteht, insbesondere bei der Verwendung neuer Technologien wie KI, hängt von der Art, dem Umfang, den Umständen und dem Zweck der Bearbeitung ab. Ein solches Risiko besteht insbesondere bei der umfangreichen Bearbeitung von besonders schützenswerten Personendaten (Art. 22 Abs. 2 Bst. b DSG), namentlich beim umfassenden Profiling zur Ermittlung von politischen Meinungen und Tendenzen der Stimmberechtigten.

b) Politische Werbung

In der Schweiz ist politische Werbung in Radio und Fernsehen gemäss Artikel 10 RTVG und Artikel 17 Absatz 3 RTVV verboten. Der Anwendungsbereich dieser Bestimmungen ist jedoch beschränkt und andere Medien, auch die sozialen Medien, sind nicht erfasst.[68]

In der EU bestehen seit Anfang April 2024 spezifische Regeln für politische Werbung.[69] Diese umfassen auch spezifische Pflichten für politische Werbung online. Das BAKOM prüft im Rahmen der Vernehmlassungsvorlage für die Regulierung von Kommunikationsplattformen, ob ähnliche Pflichten wie in der EU auch in der Schweiz sinnvoll wären.

- Risiken für die Ausübung der Grundrechte, insbesondere die Kommunikationsfreiheiten

Der Schutz von Demokratie und Rechtsstaatlichkeit geht mit dem Schutz der Grundrechte einher, insbesondere der Kommunikationsfreiheiten (u. a. Meinungs- und Informationsfreiheit [Art. 16 BV], Medienfreiheit [Art. 17 BV], Versammlungsfreiheit [Art. 22 BV]), die für diese konstitutiv sind. KI-Systeme können insbesondere zu Überwachungszwecken eingesetzt werden, z. B. zur Gesichtserkennung im öffentlichen Raum; damit können sie potenziell insbesondere die Meinungsfreiheit, die Versammlungsfreiheit und die Privatsphäre einer grossen Anzahl von Personen beeinträchtigen.

Im bereits erwähnten Urteil vom 4. Juli 2023 Glukhin gegen Russland[70] verurteilte der EGMR Russland wegen Verletzung von Artikel 8 EMRK (Recht auf Achtung des Privatlebens) und Artikel 10 EMRK (Recht auf Freiheit der Meinungsäusserung) auf-

---

[68] NADJA BRAUN BINDER/MANUELA KÄLIN, Rechtliche Aspekte der politischen Meinungsbildung (n. 61), 134 ff.

[69] Verordnung (EU) 2024/900 des Europäischen Parlamentes und des Rates vom 13. März 2024 über die Transparenz und das Targeting politischer Werbung, PE/90/2023/REV/1; JO L, 2024/900, 20. März 2024.

[70] EGMR, Glukhin gegen Russland (Fn. 6).

grund des Einsatzes von Gesichtserkennungstechnologie zur Identifizierung und Lokalisierung von Herrn Glukhin, nachdem dieser allein in der Moskauer Metro demonstriert hatte. Der Gerichtshof kam zum Schluss, dass die Bearbeitung der Personendaten des Beschwerdeführers im Zusammenhang mit seiner friedlichen Demonstration, die weder die öffentliche Ordnung noch die öffentliche Sicherheit bedroht hatte, einen Eingriff von besonderer Schwere darstellte. Der Einsatz von Gesichtserkennungstechnologie in seinem Fall war mit den Idealen und Werten einer demokratischen und rechtsstaatlichen Gesellschaft unvereinbar. Im schweizerischen Recht finden sich die Artikel 8 und 10 EMRK in den Artikeln 13 und 16 Absatz 2 BV wieder. Die von der EMRK abgeleiteten Grundsätze sind für die Schweiz verbindlich.

Der in der Schweiz geltende gesetzliche Rahmen zum Schutz der Grundrechte erweist sich somit auch im Hinblick auf die Umsetzung von Artikel 5 der KI-Konvention als relevant.

Es bleibt somit festzuhalten, dass das Schutzobjekt von Artikel 5, nämlich die Integrität des demokratischen Prozesses und die Achtung der Rechtsstaatlichkeit, in der Schweiz von einem breiten Spektrum von Normen erfasst wird. Diese reichen vom Grundrechtsschutz über den Persönlichkeitsschutz und den Datenschutz bis hin zu den Rechten im Bereich der Ausübung der politischen Rechte. Die geplante Regulierung der grossen Kommunikationsplattformen soll zudem die Rechte der Nutzerinnen und Nutzer in der Schweiz und die Transparenz seitens dieser Plattformen stärken und damit auch zu einer gut funktionierenden Demokratie beitragen.

Aufgrund der ersten oben aufgeführten Analysen drängen sich keine weiteren Gesetzesänderungen auf. Da sich die Gefahren für die Demokratie besonders schnell entwickeln und potenziell wichtige Folgen zeitigen, ist die Situation wie auch die Entwicklung der Rechtsprechung und ihre Tragweite im digitalen Kontext weiterhin genau zu beobachten, um bei Bedarf rechtzeitig reagieren zu können. Der Bericht des Bundesrates in Erfüllung des Postulats 22.3006 SiK-N «Auslegeordnung zur Bedrohung der Schweiz durch Desinformationskampagnen»[71] ist diesbezüglich hilfreich.

Weiter ist hervorzuheben, dass die KI-Konvention auch andere Bestimmungen enthält, die für den Schutz von Demokratie und Rechtsstaatlichkeit relevant sein können, insbesondere das Transparenzprinzip (Art. 8) und der rechtliche Rahmen für das Risiko- und Folgenmanagement von KI-Systemen (Art. 16). Die Umsetzung dieser allgemein ausgerichteten Bestimmungen würde ebenfalls eine Vervollständigung und Verstärkung des geltenden rechtlichen Rahmens im Bereich des Schutzes der Demokratie und des Rechtsstaates ermöglichen.

[71] Beeinflussungsaktivität und Desinformation. Bericht des Bundesrates in Erfüllung des Postulats 22.3006 SiK-N, 19. Juni 2024, abrufbar unter: www.admin.ch > Dokumentation > Medienmitteilungen > Beeinflussungsaktivität und Desinformation: Bundesrat betont Resilienz und stärkt Analyse und Koordination (abgerufen am 27. August 2024).

4.3.2 Kapitel III: Grundsätze in Bezug auf Tätigkeiten im Lebenszyklus von Systemen künstlicher Intelligenz

4.3.2.1 Artikel 6 – Allgemeiner Ansatz

Artikel 6 sieht vor, dass die Vertragsparteien die in Kapitel III der KI-Konvention enthaltenen Grundsätze bei der Umsetzung an ihre innerstaatliche Rechtsordnung und ihre anderen Verpflichtungen aus der Konvention anpassen.

Laut dem erläuternden Bericht zur Konvention ist dieser Artikel insofern von besonderer Bedeutung, da jede Vertragspartei über ein detailliertes Rechtssystem zum Schutz der Menschenrechte mit eigenen Regeln, Grundsätzen und Gepflogenheiten verfügt, die den Geltungsbereich, den Inhalt der wesentlichen Rechte und Einschränkungen, mögliche Abweichungen oder Ausnahmen von diesen Rechten sowie die Funktionsweise der anwendbaren Überwachungs- und Durchsetzungsmechanismen betreffen. [72]

Zusammenfassend lässt sich sagen, dass diese Bestimmung den bereits oben (vgl. Ziff. 4.2.1) dargelegten völkerrechtlichen Grundsatz widerspiegelt, wonach die Staaten in der Umsetzung ihrer völkerrechtlichen Verpflichtungen frei sind. Im schweizerischen Recht sieht Artikel 164 BV vor, dass dort, wo Bundeskompetenz besteht, alle wichtigen rechtsetzenden Bestimmungen in der Form eines Bundesgesetzes zu erlassen sind.

4.3.2.2 Artikel 7 – Menschenwürde und individuelle Autonomie

Gemäss Artikel 7 beschliessen die Vertragsparteien die Einführung oder Beibehaltung von Massnahmen zur Wahrung der Menschenwürde und der individuellen Autonomie in Bezug auf Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen.

Diese Bestimmung unterstreicht die Bedeutung der Menschenwürde und der individuellen Autonomie im Rahmen einer menschenzentrierten Regulierung. So dürfen Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen nicht zur Entmenschlichung des Einzelnen führen.[73]

Individuelle Autonomie ist ein wichtiger Aspekt der Menschenwürde; sie betrifft die Selbstbestimmungsfähigkeit des Individuums, d. h. seine Fähigkeit, Wahlmöglichkeiten zu nutzen und Entscheidungen ohne unzulässige Beeinflussung oder Zwang zu treffen. In Bezug auf KI erfordert die individuelle Autonomie, dass der Mensch die Kontrolle über den Einsatz und die Auswirkungen von KI-Technologien auf sein Leben hat.[74] Dieses Prinzip ist daher eng mit dem Grundsatz von Transparenz und Aufsicht verbunden (Art. 8 der Konvention, vgl. Ziff. 4.3.2.3).

---

[72] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 51.

[73] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 53.

[74] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 55.

Im schweizerischen Recht wird die Menschenwürde auf Verfassungsstufe durch Artikel 7 BV garantiert. Die Rechtsprechung erkennt dieser Bestimmung die Bedeutung eines eigenständigen Grundrechts zu, das per Definition einen absoluten Schutz all jener Aspekte der Menschenwürde gewährleistet, die von den anderen Grundrechten nicht erfasst werden.[75] In der Praxis wird die Menschenwürde jedoch vor allem zur Unterstützung eines anderen Grundrechts herangezogen und war offenbar noch nie allein ausschlaggebend für ein Urteil einer Justizbehörde des Bundes.[76] Sie hat daher eher den Charakter eines Auffanggrundrechts oder eines letzten Rechtsmittels. Dies liegt an der Unbestimmtheit ihres Schutzbereichs, denn es geht um nichts anderes als um das Menschsein selbst.[77] Der Schutzbereich der Menschenwürde deckt sich mit ihrem unantastbaren Kern (Art. 36 Abs. 4 BV). Jeder Eingriff des Staates stellt somit eine Verletzung dieses Grundrechtes dar.[78] Selbstbestimmung und Schutz der individuellen Autonomie sind sowohl durch Artikel 10 Absatz 2 BV (Recht auf persönliche Freiheit) als auch durch Artikel 13 BV (Schutz der Privatsphäre) abgedeckt.

Angesichts der mit KI verbundenen technologischen Herausforderungen, insbesondere der Gefahr der Entmenschlichung des Einzelnen[79], dürfte die Rolle des Garanten der Menschenwürde noch an Bedeutung gewinnen. Bestimmte Anwendungen, wie etwa der Einsatz von KI-Tools durch Behörden im Rahmen einer permanenten Überwachung des Einzelnen zum Zwecke des «Social Scoring» oder der Einsatz von Emotionserkennung in bestimmten Situationen, können per se einen nicht zu rechtfertigenden Eingriff in die Menschenwürde darstellen. Solche Anwendungen von KI sind daher bereits verfassungsrechtlich verboten.

In diesem Zusammenhang sind auch die vom Bund im November 2020 verabschiedeten KI-Leitlinien relevant. Diese unverbindlichen Leitlinien sollen als allgemeine Orientierungshilfe für die Entwicklung von KI in der Bundesverwaltung dienen und eine einheitliche Praxis sicherstellen. Die erste Leitlinie stellt den Menschen in den Mittelpunkt. Betont wird darin die besondere Bedeutung des Grundrechtsschutzes beim Einsatz von KI.[80]

---

[75] CR Cst.-JACQUES DUBEY, Art. 7, N 25, und die angegebenen Referenzen.

[76] CR Cst.-JACQUES DUBEY, Art. 7, N 26; BSK BV-EVA MARIA BELSER/EVA MOLINARI, Art. 7 N 39 f.; *contra* SGK BV-RAINER J. SCHWEIZER/CHRISTOPH SPENLÉ, Art. 7 N 22.

[77] CR Cst.-JACQUES DUBEY, Art. 7, N 27.

[78] CR Cst.-JACQUES DUBEY Art. 7 N 56 ff., insbesondere N 59; BSK BV-EVA MARIA BELSER/EVA MOLINARI, Art. 7 N 63.

[79] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 53.

[80] Leitlinien «Künstliche Intelligenz» für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, 2020, abrufbar unter: www.sbfi.admin.ch > BFI-Politik > Bildungs-, Forschungs- und Innovationspolitik 2021-2024 > Transversale Themen > Digitalisierung im BFI-Bereich > Künstliche Intelligenz (abgerufen am 26. August 2024). Die Leitlinien sollen alle zwei Jahre evaluiert werden. Die nächste Evaluierung ist für dieses Jahr im Rahmen der Bestandsaufnahme der Regulierungsansätze im Bereich der KI vorgesehen. Die Evaluation im Jahr 2022 hat gezeigt, dass die KI-Leitlinien in der Bundesverwaltung gut bekannt sind und von den Mitarbeitenden, die mit KI arbeiten (sei es beim Einsatz von KI, auf regulatorischer Ebene usw.), berücksichtigt und angewendet werden. Es wird zwar darauf hingewiesen, dass die Leitlinien und ihre konkrete Anwendung innerhalb der Bundesverwaltung weiter diskutiert werden müssen, die Evaluation hat

Es stellt sich die Frage nach der Tragweite von Artikel 7 der Konvention bei der Umsetzung dieser Bestimmung im privaten Sektor (vgl. Art. 3 Abs. 1 Bst. b der Konvention, Ziff. 4.2.3.1). In der schweizerischen Rechtsordnung bestehen bereits Regeln, welche die Garantie der Menschenwürde und der individuellen Autonomie zwischen Privaten verwirklichen. Dies gilt insbesondere für Artikel 28 ZGB zum Schutz der Persönlichkeit (vgl. zur Entwicklung in diesem Zusammenhang Ziff. 6.3.3).

Angesichts der obigen Ausführungen ergibt sich, dass die Verfassung das Recht auf Menschenwürde und individuelle Autonomie gegenüber dem Staat bereits schützt. Entsprechende Regelungen bestehen auch im privaten Sektor. Grundsätzlich eignet sich das schweizerische Zivilrecht mit seinen offenen Generalklauseln gut zur Bewältigung der anstehenden Probleme. Dies gilt beispielsweise für Artikel 28 ZGB zum Schutz der Persönlichkeit. Die Frage, ob diese Regeln im Einzelfall präzisiert werden müssen, kann im Falle einer Ratifikation der Konvention vertieft werden.

4.3.2.3 Artikel 8 – Transparenz und Aufsicht

Artikel 8 verpflichtet die Vertragsparteien zur Einführung oder Beibehaltung von Massnahmen, um die auf den jeweiligen Kontext und die spezifischen Risiken zugeschnittenen Transparenz- und Kontrollanforderungen für Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen, einschliesslich der Identifizierung von KI-generierten Inhalten, sicherzustellen.

Diese Bestimmung deckt zwei Aspekte ab:

- Der erste Aspekt betrifft die Transparenz von KI-Systemen. Transparenz soll durch die Verfügbarkeit angemessener Informationen über das KI-System gewährleistet werden, wie z. B. Zweck(e), bekannte Beschränkungen, Annahmen, technische Entscheidungen bei der Konzeption, Merkmale, Einzelheiten der zugrunde liegenden Modelle oder Algorithmen, Lernmethoden und Qualitätssicherungsverfahren. Darüber hinaus kann Transparenz bedeuten, dass die betroffenen Personen oder die Öffentlichkeit gegebenenfalls über die Einzelheiten der zum Training des Systems verwendeten Daten und den Schutz von Personendaten sowie über den Zweck des Systems und die Art und Weise, in der es konzipiert und eingesetzt wird, informiert werden müssen.[81]

  Das Prinzip der Transparenz zielt auch darauf ab, dass KI-Systeme für die relevanten KI-Akteure verständlich und zugänglich sein müssen. Dieser Grundsatz umfasst die Erklärbarkeit und die Interpretierbarkeit von KI-Systemen. *Erklärbarkeit* bezieht sich

jedoch keinen Bedarf für eine Anpassung oder Aktualisierung der Leitlinien ergeben, siehe BAKOM-Bericht, Monitoring der Leitlinien «Künstliche Intelligenz für den Bund», Evaluation der Anwendung und Aktualität der Leitlinien, 9. Dezember 2022, abrufbar unter: www.cnai.swiss > Ressourcen > KI-Leitlinien (abgerufen am 26. August 2024).

[81] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 58.

auf die Fähigkeit, vorbehaltlich der technischen Machbarkeit, hinreichend verständliche Erklärungen darüber zu liefern, warum ein KI-System Informationen, Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen generiert. Dies ist besonders in sensiblen Bereichen wie dem Gesundheitswesen, dem Finanzwesen und der Strafverfolgung wichtig, wo es unerlässlich ist, die Logik zu verstehen, auf der die mithilfe eines KI-Systems getroffenen Entscheidungen beruhen.[82]

Interpretierbarkeit bezieht sich auf die Fähigkeit zu verstehen, wie ein KI-System Vorhersagen macht oder Entscheidungen trifft. Es geht darum, die interne Funktionsweise, die Logik und die Entscheidungsprozesse von KI-Systemen für menschliche Nutzerinnen und Nutzer, einschliesslich Entwicklern, Interessenvertretern und Endnutzern, sowie für Betroffene verständlich und zugänglich zu machen.[83]

Unter dem Gesichtspunkt der Erklärbarkeit und Interpretierbarkeit stellen sich bei KI-Systemen besondere Herausforderungen, die je nach Art des KI-Systems mehr oder weniger stark zum Tragen kommen:[84] Ein regelbasierter Algorithmus bedeutet, dass Menschen Arbeitsschritte programmieren, mit denen die Maschine ein Ergebnis generiert. Dabei führt ein *Input* immer zum gleichen *Output*. Das Ergebnis lässt sich aus den einzelnen Schritten erklären. Diese Algorithmen können jedoch auch eine gewisse Komplexität aufweisen, was die Begründung der auf ihnen basierenden Entscheidungen erschweren kann.

Ganz anders verhält es sich mit datenbasierten Algorithmen des maschinellen Lernens («machine learning»). Diese Algorithmen analysieren riesige Datensätze, um Korrelationen zwischen den Daten zu finden und so selbst eine verallgemeinerbare Regel zu definieren. Solche Algorithmen arbeiten mit Wahrscheinlichkeiten. Das bedeutet, dass die gleiche Eingabe nicht immer zum gleichen Ergebnis führt. Ausserdem ist das Zustandekommen der Ergebnisse nicht immer nachvollziehbar. Da die Ergebnisse in diesem Fall nicht auf kausalen Zusammenhängen, sondern auf Korrelationen zwischen verschiedenen Daten beruhen, kann es im Einzelfall schwieriger sein, die für das Ergebnis des Systems massgeblichen Sachdaten und Rechtsgrundlagen zu erklären. Die Erklärbarkeit und Interpretierbarkeit solcher Systeme ist daher nur sehr schwer zu gewährleisten (Black-Box-Effekt).

- Der zweite in Artikel 8 der Konvention angesprochene Aspekt betrifft die Notwendigkeit von Kontrollmechanismen zur Überwachung, Bewertung und Steuerung der Tätig-

---

[82] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 60.

[83] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 61.

[84] Zur Begründungspflicht der Behörden vgl. NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (Fn. 39), 37; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 10 N 23; NADJA BRAUN BINDER/LILIANE OBRECHT, Die Begründung von Verfügungen beim Einsatz algorithmischer Systeme, SJZ 2024, 707 ff.

> keiten innerhalb des Lebenszyklus von KI-Systemen. Gemäss dem erläuternden Bericht kann die Kontrolle aus einer Vielzahl von Rahmenwerken und Prozessen bestehen. Genannt werden beispielsweise Rahmen für das Risiko- und Folgenmanagement, ethische Leitlinien, Zertifizierungsverfahren, Instrumente zur Erkennung und Abschwächung von Verzerrungen, die Rolle der zuständigen Behörde wie z. B. die sektoriellen Aufsichtsbehörden, Datenschutzbehörden usw.[85] In diesem Zusammenhang soll die Verpflichtung zur Transparenz auch sicherstellen, dass KI-Systeme so konzipiert, entwickelt und eingesetzt werden, dass menschliche Kontrollmechanismen während des gesamten Lebenszyklus von KI-Systemen wirksam eingesetzt werden können.[86]

Die Bestimmung bezieht sich auch auf die Transparenz und Aufsicht bei der Identifizierung von KI-generierten Inhalten. Ziel ist es, das Risiko der Täuschung zu verringern und die Unterscheidung zwischen authentischen, von Menschen erzeugten Inhalten und KI-generierten Inhalten zu ermöglichen. Solche Massnahmen könnten Techniken wie Kennzeichnungen oder Wasserzeichen umfassen, bei denen in der Regel eine erkennbare Signatur im KI-generierten Ergebnis eingebettet wird, unter Vorbehalt der Verfügbarkeit dieser Technologien und ihrer nachgewiesenen Wirksamkeit, des allgemein anerkannten Stands der Technik und der Besonderheiten der verschiedenen Inhaltstypen.[87]

Im schweizerischen Recht kann Transparenz im öffentlichen Sektor und im staatlichen Handeln als Element der Rechtsstaatlichkeit verstanden werden (Art. 5 BV);[88] eine weitere relevante verfassungsrechtliche Grundlage ist der Informationsauftrag des Bundesrates (Art. 180 Abs. 2 BV).[89]

Auf Gesetzesstufe soll das BGÖ die Kontrolle der Verwaltung durch die Bevölkerung verbessern, indem allen interessierten Personen Einsicht in die Dokumente der Bundesbehörden gewährt wird. Voraussetzung ist, dass die Dokumente vorhanden sind und der Zugang nicht aufgrund der im BGÖ vorgesehenen Ausnahmen ausgeschlossen ist. Die aktive Informationspflicht des Bundesrats ist hingegen insbesondere in Artikel 10 RVOG geregelt.

Im Übrigen ist zu beachten, dass die Transparenz staatlichen Handelns auch durch das Legalitätsprinzip gewährleistet werden kann (vgl. die Ausführungen zu diesem Prinzip in Ziff. 4.2.3.1.2).

---

[85] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 63.

[86] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 65.

[87] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 59.

[88] ROLF H. WEBER/SIMON HENSELER, Regulierung von Algorithmen in der EU und in der Schweiz, Zeitschrift für Europarecht 2020, 28 ff., 36.

[89] DOMINIQUE HÄNNI, Vers un principe d'intégrité de l'administration publique – La prévention de la corruption en droit administratif, Genf/Zürich/Basel 2019, 305 N 794.

Das Öffentlichkeitsprinzip wird in Bezug auf staatliches Handeln in einzelnen Bereichen teilweise konkretisiert. So müssen die Gerichte den betroffenen Parteien Akteneinsicht gewähren, damit diese ihr rechtliches Gehör wahrnehmen können (Art. 29 Abs. 2 BV). Zudem müssen die Behörden ihre Entscheide begründen. Für die Ausführungen zu diesen Aspekten und zu den Anforderungen der KI-Konvention in Bezug auf Rechtsmittel wird auf die nachfolgenden Ausführungen zu den Artikeln 14 und 15 der KI-Konvention verwiesen (vgl. Ziff. 4.3.3).

Auch das DSG konkretisiert das Öffentlichkeitsprinzips und betont die Transparenz bei der Bearbeitung von Personendaten. Nach Artikel 6 Absatz 3 DSG dürfen Personendaten nur für einen bestimmten und für die betroffene Person erkennbaren Zweck beschafft werden; sie dürfen nur so bearbeitet werden, dass es mit diesem Zweck vereinbar ist. Nach Artikel 12 DSG sind sowohl die Verantwortlichen als auch die Auftragsbearbeiter zum Führen eines Verzeichnisses über ihre Bearbeitungstätigkeiten verpflichtet (für weitere Einzelheiten siehe Ziff. 4.3.2.5).

Zudem sieht die Informationspflicht nach Artikel 19 DSG vor, dass die betroffene Person grundsätzlich angemessen über die Beschaffung von Personendaten informiert wird. Für automatisierte, von KI-Systemen getroffene Einzelentscheidungen sieht Artikel 21 Absatz 1 DSG vor, dass der Verantwortliche, d. h. der Betreiber des Systems, die betroffene Person über jede Entscheidung informieren muss, die ausschliesslich auf einer automatisierten Bearbeitung beruht und die für sie mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt. Artikel 21 Absatz 4 DSG sieht insbesondere vor, dass eine automatisierte Einzelentscheidung eines Bundesorgans als solche zu kennzeichnen ist. Die betroffene Person hat zudem grundsätzlich das Recht, ihren Standpunkt darzulegen und zu verlangen, dass die automatisierte Einzelentscheidung durch eine natürliche Person überprüft wird. Artikel 25 Absatz 2 Buchstabe f DSG sieht ein entsprechendes Auskunftsrecht vor. Danach ist der betroffenen Person zumindest gegebenenfalls Auskunft darüber zu erteilen, ob eine automatisierte Einzelentscheidung vorliegt und auf welcher Logik diese Entscheidung beruht. Dies gilt allerdings nur für ausschliesslich automatisierte Entscheidungen und nicht, wenn KI-Systeme als – wenn auch sehr wesentliche – Entscheidungshilfen eingesetzt werden.[90] Schliesslich sieht Artikel 56 DSG vor, dass der EDÖB ein Register der Bearbeitungstätigkeiten der Bundesorgane führt und dass dieses Register veröffentlicht wird.

[90] Der EuGH scheint den Begriff der automatisierten Einzelentscheidung weiter auszulegen. So hat er kürzlich entschieden, dass bereits das «Scoring» eines Unternehmens, das einen Kreditwürdigkeitsprüfung durchführt, eine solche Entscheidung darstellt, wenn sich der Empfänger bei seiner Entscheidung über einen Kreditantrag der betroffenen Person massgeblich auf dieses Ergebnis stützt (EuGH, Rechtssache C-634/21 vom 7. Dezember 2023, SCHUFA Holding [Scoring]). Die Auswirkungen dieser Rechtsprechung in der Schweiz sollten weiterverfolgt werden. In jedem Fall werden die Besonderheiten des Schweizer Rechts zu berücksichtigen sein.

Überdies sehen einige technische Regelwerke punktuell Transparenzanforderungen für die Verwendung von Algorithmen vor.[91]

Im privaten Sektor sollten die Grundsätze der Transparenz und der Aufsicht auch im Verhältnis zwischen Privaten gelten, wo eine direkte oder indirekte horizontale Wirkung der Grundrechte besteht oder in Zukunft erkannt wird (vgl. Ziff. 4.2.3.1). Das DSG gilt diesbezüglich auch für Private und bietet Schutz in Bezug auf die Transparenz der Bearbeitung (vgl. insbesondere Art. 6 Abs. 3, Art. 12, 19, 21 und Art. 25 Abs. 2 Bst. f DSG). Zur allfälligen Pflicht, im Streitfall vor Gericht die Funktionsweise eines in einer Rechtsbeziehung eingesetzten KI-Systems zu erklären, siehe Ziff. 6.3.1.

Zusammenfassend lässt sich feststellen, dass das Prinzip der Transparenz und der Aufsicht in der schweizerischen Rechtsordnung bereits punktuell verwirklicht ist. Die Analyse kommt jedoch zum Schluss, dass der geltende Rechtsrahmen zumindest für den öffentlichen Sektor nicht ausreicht, um die Anforderungen von Artikel 8 der Konvention zu erfüllen. Auf Ebene der Verwaltung besteht insbesondere die Möglichkeit, ein öffentliches Register aller von den öffentlichen Stellen eingesetzten KI-Systeme mit Meldepflicht einzurichten. Um dem Black-Box-Effekt entgegenzuwirken, könnten Anforderungen wie eine detaillierte Dokumentation sowie regelmässige Audits und Tests der KI-Systeme vorgesehen werden. Ziel ist es, die Kontrolle über die Systeme zu behalten, indem sichergestellt wird, dass die Funktionsweise der Systeme nachvollziehbar bleibt oder zumindest, dass die bei der generativen KI inhärenten Halluzinationen[92] erkannt und korrigiert werden können.

Sowohl im öffentlichen als auch im privaten Sektor stellt sich die Frage nach einer allfälligen Ausdehnung der Informationspflichten bzw. der zu erteilenden Auskünfte bei der Ausübung des Auskunftsrechts nach DSG auf teilautomatisierte Entscheide.

Hinsichtlich der Kennzeichnung von KI-generierten Inhalten enthält das schweizerische Recht grundsätzlich keine Kennzeichnungspflicht. Die Einführung einer solchen Massnahme könnte angezeigt sein, um KI-generierte Inhalte erkennen zu können.

Dabei ist zu beachten, dass die getroffenen Massnahmen den entgegenstehenden Interessen Rechnung tragen müssen, die mit dem Interesse an der Offenlegung bestimmter Informationen kollidieren können, wie z. B. der Schutz von Daten Dritter, Geschäftsgeheimnisse oder Interessen der Strafverfolgung.

---

[91] Vgl. Artikel 24 Absatz 1 der Verordnung des EDI über den Strahlenschutz bei medizinischen Teilchenbeschleunigeranlagen (SR **814.501.513**): «Die Lieferantin oder der Lieferant des Bestrahlungsplanungssystems hat in der technischen Beschreibung genaue Angaben über die zur Berechnung der Dosisverteilungen verwendeten Algorithmen zu machen […]».

[92] Halluzinationen äussern sich in Antworten, die in keiner Weise der Realität entsprechen, vgl. Merkblätter vom 26. April 2024 zur Sensibilisierung betreffend grossen KI-Sprachmodellen (large language models, LLM) in der Bundesverwaltung, 2, abrufbar unter: www.cnai.swiss > Dienstleistungen > Weitere Dienstleistungen > Merkblätter zu KI in der Bundesverwaltung (abgerufen am 26. August 2024).

4.3.2.4 Artikel 9 – Rechenschaftspflicht und Verantwortung

Artikel 9 verpflichtet die Staaten zur Einführung oder Beibehaltung von Massnahmen zur Sicherstellung der Rechenschaftspflicht und der Verantwortung für nachteilige Auswirkungen, die sich aus Tätigkeiten im Lebenszyklus von KI-Systemen ergeben.

Gemäss dem erläuternden Bericht zur Konvention bedeutet diese Verpflichtung, dass Personen, Organisationen oder Stellen, die für Tätigkeiten während des Lebenszyklus von KI-Systemen verantwortlich sind, für nachteilige Auswirkungen dieser Tätigkeiten auf die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit zur Rechenschaft gezogen werden müssen. Die Bestimmung fordert die Schaffung von Rahmenbedingungen und Mechanismen zur Durchsetzung dieser Verpflichtung.[93]

Artikel 9 fordert darüber hinaus die Schaffung klarer Verantwortungslinien, um Handlungen und Entscheidungen bis zu bestimmten Personen oder Institutionen zurückverfolgen zu können. Es muss sichergestellt werden, dass nachteilige Auswirkungen ermittelt und die Verantwortlichkeiten angemessen zugewiesen werden können.[94] Dies kann neue Rahmenbedingungen und Mechanismen umfassen, aber auch gerichtliche und administrative Massnahmen, zivil-, straf- und andere Haftungsregelungen sowie im öffentlichen Sektor Verwaltungs- und andere Verfahren, die eine Anfechtung von Entscheidungen ermöglichen, oder spezifische Verantwortlichkeiten und Pflichten für Betreiber, die an den Tätigkeiten im Lebenszyklus von KI-Systemen beteiligt sind.[95]

Diese Verpflichtung ist untrennbar mit dem Grundsatz der Transparenz und Aufsicht (Art. 8) verbunden, da über die Transparenz indirekt erklärt werden kann, wie KI-Systeme funktionieren und wie sie Entscheidungen generieren. Sie steht auch im Zusammenhang mit der Verpflichtung, einen Rahmen für das Risiko- und Folgenmanagement (Art. 16) zur proaktiven Vermeidung und Minderung von Risiken und Folgen zu schaffen.[96]

Für den privaten Sektor wird das in Artikel 9 umschriebene Verantwortlichkeitsprinzip zwischen Privaten dort umzusetzen sein, wo eine direkte oder indirekte horizontale Wirkung der Grundrechte besteht oder in Zukunft erkannt wird (vgl. Ziff. 4.2.3.1). Für Überlegungen zum Zivil- und Strafrecht, insbesondere zur Haftung, bzw. Verantwortlichkeit, wird auf die nachfolgenden Ausführungen verwiesen (vgl. Ziff. 6.3 und 6.6).

Im schweizerischen Recht sind die Haftungsregeln so formuliert, dass sie unabhängig davon gelten, ob eine Verletzung durch ein KI-System verursacht wurde oder nicht. Das DSG

[93] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 66.

[94] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 68.

[95] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 66.

[96] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 69 f.

sieht zudem, unabhängig von KI, Massnahmen zur Rechenschaftspflicht («Accountability») vor (vgl. Ziff. 4.3.2.6).

Im öffentlichen Sektor gelten die Staatshaftungsnormen und die Vorschriften, die den Rahmen für die Tätigkeit seiner Organe bilden, unabhängig davon, ob bei dieser Tätigkeit ein KI-System eingesetzt wird oder nicht. Es muss jedoch sichergestellt werden, dass der aus Regeln, Rechtsnormen und anderen Mechanismen bestehende Rahmen eine wirkungsvolle Zuweisung der Verantwortungen erlaubt. Die Analyse zeigt, dass die geplanten Massnahmen zur Umsetzung von Artikel 8 (Transparenz und Aufsicht), Artikel 14 (Rechtsmittel), Artikel 15 (Verfahrensgarantien) und 16 (Rahmen für das Risiko- und Folgenmanagement), für welche Handlungsbedarf festgestellt wurde, dazu beitragen werden, die Wirksamkeit der bestehenden Normen zu erhöhen.

Bezüglich der Überlegungen zur zivilrechtlichen Haftung und strafrechtlichen Verantwortlichkeit wird auf die nachfolgenden Ausführungen verwiesen (vgl. Ziff. 6.3 und 6.6).

4.3.2.5 Artikel 10 – Gleichstellung und Nichtdiskriminierung

Artikel 10 Absatz 1 sieht vor, dass jede Vertragspartei Massnahmen ergreift oder beibehält, um die Achtung der Gleichheit, einschliesslich der Gleichstellung der Geschlechter, und der Nichtdiskriminierung bei Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen in Übereinstimmung mit dem anwendbaren Völkerrecht und dem innerstaatlichen Recht zu gewährleisten. Diese Formulierung bezieht sich insbesondere auf alle für die Vertragsparteien relevanten völkerrechtlichen und innerstaatlichen Rechtsinstrumente. Diese bieten zusammen eine solide Rechtsgrundlage und Orientierungshilfen, die es jedem Staat ermöglichen, Massnahmen zur Gewährleistung der Gleichstellung aller und zur Nichtdiskriminierung bei Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen zu ergreifen.[97]

Nach Absatz 2 verpflichtet sich jede Vertragspartei dazu, Massnahmen zur Beseitigung von Ungleichheiten zu ergreifen oder beizubehalten, um im Einklang mit ihren innerstaatlichen und völkerrechtlichen Verpflichtungen auf dem Gebiet der Menschenrechte unparteiische, faire und gleichberechtigte Ergebnisse der Tätigkeit innerhalb des Lebenszyklus von KI-Systemen zu erzielen. Gemäss dem erläuternden Bericht soll sich der hier geforderte Ansatz nicht auf die Forderung beschränken, dass eine Person «ohne objektive und angemessene Rechtfertigung» aufgrund eines oder mehrerer ihrer geschützten Merkmale nicht weniger günstig behandelt werden darf. Vielmehr verpflichten sich die Vertragsparteien zur Einführung neuer oder zur Beibehaltung bestehender Massnahmen zur Überwindung struktureller und historisch bedingter Ungleichheiten.[98]

[97] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 71.

[98] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 77.

Der Einsatz von KI stellt eine grosse Herausforderung für die Wahrung des Grundsatzes der Gleichstellung und Nichtdiskriminierung dar.[99] Da Algorithmen lediglich das Ergebnis von Berechnungen darstellen, die von Menschen festgelegt werden und auf Daten basieren, die von Menschen, Maschinen oder einer Kombination aus beiden gesammelt wurden, können sie menschliche Verzerrungen widerspiegeln und verarbeiten, die in ihre Programmierung und Datenverarbeitung eingeflossen sind.[100] Diese meist unbeabsichtigten Verzerrungen[101] können sich aus den verwendeten Daten (mangelnde Repräsentativität, veraltete Daten, unzureichende Datenverarbeitung), aus dem Algorithmus selbst (welche Parameter werden im Modell berücksichtigt, welche nicht[102]) oder aus der Art und Weise seiner Anwendung (welche Bedürfnisse soll das System erfüllen und wie wird es in der Praxis eingesetzt) ergeben.[103] Diese Risiken können durch Rückkopplungsschleifen und mangelnde Transparenz der Prozesse noch verstärkt werden[104] (zur Frage der Transparenz vgl. Ziff. 4.3.2.3).

Das Arbeitsrecht, insbesondere die Rekrutierungsphase, wird diesbezüglich häufig als sensibler Bereich benannt. Die in diesem Bereich verwendeten KI-Systeme sind in der Tat sehr vielfältig und reichen von der «einfachen» Auswahl von Lebensläufen bis hin zu Programmen, welche die Mimik eines Bewerbers während eines Videointerviews analysieren oder die berufliche Leistung auf der Grundlage eines Lebenslaufs vorhersagen.[105] Ein bekanntes Beispiel in diesem Zusammenhang ist Amazon. Im Jahr 2018 berichtete die Nachrichtenagentur Reuters, dass das Unternehmen eine auf maschinellem Lernen beruhende Bewerbungssoftware entwickelt habe, um die besten Lebensläufe ausfindig zu machen. Das Programm benachteiligte jedoch systematisch die Lebensläufe von Frauen, da es die geschlechtsspezifischen Unterschiede bei den Einstellungen in den letzten zehn Jahren widerspiegelte. Auch der Verzicht auf das Ankreuzen des Geschlechts änderte nichts daran, da das Programm aus anderen Abgleichen auf das Geschlecht der Person schliessen konnte.[106]

Der Einsatz von KI kann in allen Bereichen zu Ungleichheiten und Diskriminierungen führen. Beispiele finden sich insbesondere in folgenden Bereichen: Migration; Zugang zu Gütern und

---

[99] Bericht «Herausforderungen der künstlichen Intelligenz» (Fn. 1), 38 f.

[100] Europarat, Commission pour l'égalité de genre (GEC) und Comité directeur sur l'anti-discrimination, la diversité et l'inclusion (CDADI), Etude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination, erstellt durch IVANA BARTOLETTI/RAPHAËLE XENIDIS, Strassburg 2023, 17, abrufbar unter: https://rm.coe.int/prems-107623-fra-2530-etude-sur-l-impact-de-ai-a5-web/1680ac99e2 (abgerufen am 15. August 2024).

[101] GRÉGOIRE LOISEAU, Intelligence artificielle et droit des personnes, in: Alexandra Bensamoun/Grégoire Loiseau (Hrsg.), Droit et intelligence artificielle, Paris 2022, 35 ff., 49.

[102] Die Proxy-Diskriminierung ist im Zusammenhang mit KI besonders relevant. Proxys (oder Stellvertreter) sind scheinbar harmlose Kriterien, die jedoch stark mit sensiblen Kriterien korrelieren können. Beispielsweise bedeutet die Angabe «30 Jahre Berufserfahrung», dass die betreffende Person mindestens 45 Jahre alt sein muss, und die Postleitzahl lässt Annahmen über den sozioökonomischen Status oder die Herkunft einer Person zu, vgl. NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 12 N 30; AlgorithmWatch/CH, Positionspapier: Schutz vor algorithmischer Diskriminierung, September 2023, 6, abrufbar unter: https://algorithmwatch.ch > Positionen > Diskriminierende Algorithmen: So gelingt der Schutz (abgerufen am 15. August 2024).

[103] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 75 und 76.

[104] Bericht «Herausforderungen der künstlichen Intelligenz» (Fn. 1), 38 f. Vgl. für eine umfassende Übersicht der Problematik: FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung beim Einsatz von Künstlicher Intelligenz (KI), Jusletter IT 4. Juli 2024.

[105] AVI ASHER-SCHAPIRO, AI is taking over job hiring, but can it be racist?, Reuters 2021, abrufbar unter: https://www.reuters.com/article/business/healthcare-pharmaceuticals/ai-is-taking-over-job-hiring-but-can-it-be-racist-idUSL5N2NF5ZC/ (abgerufen am 15. August 2024).

[106] JEFFREY DASTIN, Insight – Amazon scraps secret AI recruiting tool that showed bias against women, Reuters 2018, in englischer Sprache abrufbar unter: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ (abgerufen am 15 August 2024).

Dienstleistungen, Banken und Versicherungen; Zugang zu Sozialleistungen[107]; Risikobewertung in den Bereichen Sicherheit, Verbrechensprävention, Aufrechterhaltung der öffentlichen Ordnung und Justizwesen; Zugang zu öffentlichen und administrativen Dienstleistungen; Bildung; Medien und Suchmaschinen; Gesundheitswesen.[108] Insbesondere im letztgenannten Bereich zeigen Beispiele, dass Systeme manchmal auf der Grundlage von nicht repräsentativen Daten trainiert werden. Dies kann zu Fehldiagnosen und zu Diskriminierung von Frauen[109], Minderheiten und Menschen *of Color*[110] oder älteren Menschen[111] führen.

Es ist wichtig anzumerken, dass ein gut konzipiertes KI-System auch dazu beitragen kann, diskriminierten oder benachteiligten Menschen zu helfen. KI-Systeme können beispielsweise eingesetzt werden, um den Zugang zu Informationen oder zu vorhandenen Gütern und Dienstleistungen zu verbessern. Der Einsatz maschineller Übersetzungssysteme für Regional- oder Minderheitensprachen könnte den Zugang zur Grundversorgung verbessern. KI könnte auch zur Förderung der Gleichstellung im Polizeisektor eingesetzt werden, z. B. zur Prävention geschlechtsspezifischer Gewalt (siehe die Software VioGen in Spanien). Im Gesundheitswesen könnte KI eingesetzt werden, um den Zugang zur Gesundheitsversorgung in benachteiligten Gebieten und die Diagnostik für traditionell unterrepräsentierte Gruppen zu verbessern.[112] KI-Systeme können auch menschliche Fehler (Rückkopplungsfehler, unvollständige Informationen) korrigieren. In solchen Fällen kann das KI-System zur Aufdeckung von Ungleichbehandlungen beitragen, um diese zu korrigieren und ihnen entgegenzuwirken.[113]

Im schweizerische Recht gibt es keine KI-spezifischen Bestimmungen zur Rechtsgleichheit und zur Nichtdiskriminierung. Es ist daher zu prüfen, ob der bestehende Rechtsrahmen, der auch für den Einsatz von KI gilt, die Herausforderungen dieser Technologie ausreichend abdeckt.

Die internationalen Verpflichtungen der Schweiz in den Bereichen Gleichstellung und Nichtdiskriminierung ergeben sich insbesondere aus den beiden UNO-Pakten (Art. 2 Abs. 2 UNO-Pakt I und Art. 2 Abs. 1 UNO-Pakt II, Art. 3 beider UNO-Pakte und Art. 26 UNO-Pakt II) sowie aus der EMRK (Art. 14). Diese Bestimmungen haben jedoch aus verschiedenen Gründen

---

[107] Hier ist der Skandal bezüglich der Familienzulagen in der Niederlanden zu nennen. Bei der Entwicklung des algorithmischen Systems zur Feststellung, ob Anträge für Familienzulagen als fehlerhaft und potenziell betrügerisch einzustufen sind, wurden Kriterien aus dem Bereich des *Racial Profiling* einbezogen. Dies führte dazu, dass Zehntausende von Eltern und Erziehungsberechtigten, zumeist aus einkommensschwachen Familien, von den niederländischen Steuerbehörden fälschlicherweise des Betrugs bezichtigt wurden, vgl. https://www.amnesty.org/fr/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/ (abgerufen am 15. August 2024).

[108] Conseil de l'Europe, Etude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination (Fn. 100), 23 ff. (abgerufen am 15. August 2024); Algorithm-Watch/CH, Positionspapier: Schutz vor algorithmischer Diskriminierung, September 2023 (Fn. 102) (abgerufen am 15. August 2024).

[109] ANNE-SOPHIE MORAND, A WEIRD AI system?, Jusletter 20. September 2021, 9 N 21 ff.

[110] ANNE-SOPHIE MORAND, A WEIRD AI system? (Fn. 109), 7 N 16 ff.

[111] Vgl. JUSTYNA STYPINSKA, AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies, AI & SOCIETY 2023, 665 ff.

[112] Europarat, Etude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination (Fn. 100), 89.

[113] Universität Zürich, Digital Society Initiative, Positionspapier, Ein Rechtsrahmen für Künstliche Intelligenz, November 2021, 4, abrufbar unter: https://www.dsi.uzh.ch > Forschung > Strategy Lab > 1. DSI Strategy Lab (abgerufen am 19. August 2024); FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung (Fn. 104), 17.

keine eigenständige Bedeutung und können in der Schweiz nur zusammen mit anderen Bestimmungen der Texte geltend gemacht werden.[114] Die Schweiz hat zudem mehrere UNO-Übereinkommen ratifiziert, welche die Rechtsgleichheit in bestimmten Bereichen garantieren, wie das Übereinkommen zur Beseitigung jeder Form von Rassendiskriminierung[115] und das Übereinkommen zur Beseitigung jeder Form von Diskriminierung der Frau[116]. Die meisten Bestimmungen dieser Texte sind vor dem Bundesgericht einklagbar.[117] Zu erwähnen ist auch das Übereinkommen über die Rechte von Menschen mit Behinderungen[118], dessen Artikel 5 direkt einklagbar ist.[119]

Auf Verfassungsstufe ist der Grundsatz der Rechtsgleichheit in Artikel 8 Absatz 1 und das Diskriminierungsverbot in Artikel 8 Absatz 2 verankert. Absatz 3 dieser Bestimmung garantiert die rechtliche und tatsächliche Gleichstellung von Mann und Frau, und Absatz 4 enthält eine Verpflichtung zur Beseitigung von Benachteiligungen von Menschen mit Behinderungen. Der Grundsatz der Rechtsgleichheit gilt für alle Beziehungen, die der Staat eingehen kann. Er steht in engem Zusammenhang mit dem Gebot der Gerechtigkeit und Billigkeit. Mit Ausnahme der Lohngleichheit von Mann und Frau in Absatz 3 haben das Gebot der Rechtsgleichheit und das Diskriminierungsverbot keine unmittelbare Wirkung gegenüber Privaten. Allerdings verpflichtet Artikel 35 Absatz 1 BV den Staat, dafür zu sorgen, dass die Grundrechte in der gesamten Rechtsordnung zur Geltung kommen und, soweit sie sich dazu eignen, auch unter Privaten wirksam werden (Art. 35 Abs. 3 BV, vgl. Ziff. 4.3.1.1). Andere Verfassungsbestimmungen wiederholen das Gebot der Rechtsgleichheit in spezifischen Bereichen. So garantiert etwa Artikel 27 BV den freien Wettbewerb, oder Artikel 15 BV verbietet es dem Staat, für oder gegen eine Religion oder Weltanschauung Partei zu ergreifen (Grundsatz der religiösen Neutralität des Staates).[120]

Der schweizerische Gesetzgeber hat den Grundsatz der Rechtsgleichheit in mehreren Gesetzestexten konkretisiert. Diese gelten, wie bereits erwähnt, auch für KI-Systeme:

- GIG: Das GIG bezweckt die rechtliche und tatsächliche Gleichstellung von Frau und Mann in Arbeitsverhältnissen im öffentlichen und privaten Sektor. Konkret schreibt es

---

[114] GIORGIO MALINVERNI et. al., Droit constitutionnel suisse – Volume II (Fn. 56), N 1103 ff.; für Art. 14 EMRK vgl. BGE **123** II 472, E. 4c.

[115] SR **0.104**.

[116] SR **0.108**.

[117] Zur ersten dieser beiden Übereinkommen siehe BGE **123** IV 202, E. 2; BGE **123** II 472, 479, E. 4d; für das zweite, BGE **125** I 21, E. 4a; BGE **145** I 308, E. 3.1.

[118] SR **0.109**.

[119] Urteil des BGer 8C_633/2021, E. 4.2.

[120] GIORGIO MALINVERNI et al., Droit constitutionnel suisse – Volume II (Fn. 56), N 1100, 1113 f.

vor, dass Arbeitnehmerinnen und Arbeitnehmer aufgrund ihres Geschlechts weder direkt noch indirekt benachteiligt werden dürfen, namentlich nicht unter Berufung auf den Zivilstand, auf die familiäre Situation oder, bei Arbeitnehmerinnen, auf eine Schwangerschaft (Art. 3 Abs. 1 und 2). Dieses Diskriminierungsverbot gilt insbesondere für die Einstellung, Aufgabenzuteilung, Entlöhnung und Entlassung. Weil Diskriminierungen schwierig zu beweisen sind, führt das Gesetz auch eine Beweislasterleichterung ein (Art. 6). Das Gesetz sieht auch ein besonderes Klagerecht für Organisationen vor, wenn eine erhebliche Anzahl von Arbeitsverhältnissen betroffen ist (Art. 7).

- BehiG und dazugehörende Ausführungsverordnungen: Dieses Gesetz soll die Benachteiligung von Menschen mit Behinderungen verhindern, verringern und beseitigen. Es gilt insbesondere für verschiedene Arten von öffentlich zugänglichen Bauten oder für den Zugang zu Bildung. Es richtet sich in erster Linie an den Staat, unter Vorbehalt der in Artikel 3 genannten Bauten.

- Zu erwähnen ist auch Artikel 261bis StGB, der im Rahmen der Ratifizierung des Übereinkommens zur Beseitigung jeder Form von Rassendiskriminierung[121] eingeführt wurde, um diese völkerrechtlichen Verpflichtungen in nationales Recht umzusetzen. Dieser Artikel bezweckt in erster Linie den Schutz der Menschenwürde derjenigen Personen, gegen die sich rassistische Herabsetzung richtet. In zweiter Linie schützt er den öffentlichen Frieden, der durch Handlungen bedroht ist, die geeignet sind, Menschen gegeneinander aufzubringen.[122] Mit Ausnahme von Absatz 5 (Verweigerung einer der Allgemeinheit angebotenen Leistung aus Gründen der Rasse, Ethnie oder Religion) muss die strafrechtlich relevante Handlung in der Öffentlichkeit stattfinden. Der Begriff der Öffentlichkeit wird vom BGer relativ weit gefasst: Darunter fallen alle Äusserungen oder Verhaltensweisen, die nicht dem privaten Bereich zugeordnet werden können, also z. B. im Familien- oder Freundeskreis oder in einem Umfeld persönlicher Beziehungen oder besonderen Vertrauens.[123]

- Schliesslich ist auf das Bundesgesetz über genetische Untersuchungen beim Menschen[124] hinzuweisen, das in Artikel 4 festhält, dass niemand wegen seines Erbguts

[121] SR **0.104**.

[122] CR CP II-MIRIAM MAZOU, Art. 261bis N 3 und die zitierten Referenzen.

[123] BGE **130** IV 111, E. 5.2.1

[124] SR **810.12**.

diskriminiert werden darf, oder auf Artikel 2 des Freizügigkeitsabkommens[125], der die Diskriminierung aufgrund der Staatsangehörigkeit verbietet.

In den nicht von diesen Texten erfassten Bereichen können sich diskriminierte Personen je nach Umständen direkt auf die Verfassung oder auf allgemeinere Schutzbestimmungen wie Artikel 6 des ArG berufen.[126] Im schweizerischen Recht hat sich der Gesetzgeber dafür entschieden, den Grundsatz der Rechtsgleichheit und Nichtdiskriminierung unter Privaten je nach Art der Diskriminierung oder in bestimmten Bereichen umzusetzen. Im Privatrecht bieten bestimmte Normen indirekten Schutz vor Diskriminierung. So kann man sich beispielsweise bei einer die Persönlichkeit tangierenden Diskriminierung auf Artikel 28 ZGB[127] und Artikel 328 OR[128] berufen, um die Unterlassung der Diskriminierung oder eine Entschädigung zu verlangen. Es besteht auch ein Schutz vor diskriminierenden Kündigungen, die in die Kategorie der missbräuchlichen Kündigungen fallen, wenn diese wegen einer Eigenschaft, die der anderen Partei kraft ihrer Persönlichkeit zusteht, oder wegen der Ausübung eines verfassungsmässigen Rechts ausgesprochen wurde (Art. 336 Abs. 1 Bst. a und b OR).

Neben den oben genannten Bestimmungen ist im Hinblick auf den Schutz natürlicher Personen auch das DSG zu erwähnen. Dieses Gesetz zielt zwar nicht primär auf die Bekämpfung von Diskriminierungen und Ungleichheiten ab, enthält aber dennoch zahlreiche Bestimmungen, die einen indirekten Schutz im privaten und öffentlichen Sektor bieten können. Die Bearbeitung von Personendaten innerhalb des Lebenszyklus eines KI-Systems muss unter Einhaltung der im Text festgelegten Regeln erfolgen. Im Hinblick auf die hier behandelte Thematik erscheinen folgende Punkte relevant:

- *Besonders schützenswerte Personendaten:* Das DSG sieht einen erhöhten Schutz vor, wenn besonders schützenswerte Personendaten bearbeitet werden, sowie bei Profiling und Profiling mit hohem Risiko (Art. 5 Bst. c, f und g), was einen Schutz vor Diskriminierung bieten kann. Besonders schützenswerte Daten sind insbesondere Daten über religiöse, weltanschauliche, politische oder gewerkschaftliche Ansichten oder

---

[125] Abkommen zwischen der Schweizerischen Eidgenossenschaft einerseits und der Europäischen Gemeinschaft und ihren Mitgliedstaaten andererseits über die Freizügigkeit vom 21. Juni 1999, SR **0.142.112.681**.

[126] Siehe insbesondere SECO, Kommentar, Art. 2 ArGV 3, 302-4, abrufbar unter: https://www.seco.admin.ch > Arbeit > Arbeitsgesetz und Verordnungen > Kommentar zum Arbeitsgesetz und seinen Verordnungen > Kommentar zur ArGV 3 (abgerufen am 19. August 2024).

[127] Vgl. zum Beispiel: FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper: Diskriminierung, Juni 2024, 7, abrufbar unter: https://www.itsl.uzh.ch > Wissenstransfer > Publikationen > Positionspapiere > Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von künstlicher Intelligenz (abgerufen am 19. August 2024); ELEONOR KLEBER, La discrimination multiple – Etude de droit international, suisse et européen, Zürich 2015, 210 und die zitierten Referenzen.

[128] CR CO I-LEMPEN, Art. 328 N 1; BSK OR I-PORTMANN/RUDOLPH, Art. 328 N 45 f. und Art. 320 N 2a für diese vorvertragliche Verhandlungsphase.

Tätigkeiten, Daten über die Gesundheit, die Intimsphäre oder die Zugehörigkeit zu einer Rasse oder Ethnie, genetische Daten sowie Daten über Massnahmen der sozialen Hilfe.

- *Erfordernis einer gesetzlichen Grundlage:* Bundesorgane dürfen Personendaten grundsätzlich nur bearbeiten, wenn eine gesetzliche Grundlage dies erlaubt. Der Entscheid über den Einsatz von KI läge somit bei Vorliegen eines Gesetzes im materiellen Sinn beim Bundesrat, und bei einer formell gesetzlichen Grundlage allenfalls beim Parlament, sofern nicht bereits eine genügende gesetzliche Grundlage besteht (vgl. Ziff. 4.2.3.1). Dieses Erfordernis stellt auch eine Form des indirekten Schutzes vor Ungleichbehandlung und Diskriminierung dar, da der Staat bei der Ausarbeitung von Gesetzen an den Grundsatz der Rechtsgleichheit gebunden ist. Es sei daran erinnert, dass das geltende Recht bei der Ausarbeitung von Gesetzesentwürfen auch eine spezifische Verpflichtung des Gesetzgebers vorsieht, in bestimmten Fällen zusätzlich zur Regulierungsfolgenabschätzung (RFA) eine Analyse der Auswirkungen des Gesetzes auf die Gleichstellung von Frauen und Männern durchzuführen.

- *Grundsätze der Verhältnismässigkeit, der Richtigkeit und des Datenschutzes durch Technik und datenschutzfreundliche Voreinstellungen* (Art. 6 Abs. 2 und Art. 7 DSG)*:* Die Umsetzung dieser Grundsätze kann dazu beitragen, die Beschaffung und Bearbeitung von Daten auf ein Minimum zu beschränken, ihre Richtigkeit und Sicherheit zu gewährleisten und sie sogar zu anonymisieren, wodurch das Risiko einer Diskriminierung aufgrund dieser Daten verringert wird. Dies kann auch dazu beitragen, die Datenbearbeitung transparent zu machen und die betroffenen Personen in die Lage zu versetzen, ihre Rechte geltend zu machen und so mögliche Diskriminierungen festzustellen.

- *Pflicht zur Durchführung einer Folgenabschätzung*: Wenn die geplante Bearbeitung voraussichtlich ein hohes Risiko für die Persönlichkeit oder die Grundrechte mit sich bringt, muss der Verantwortliche eine Datenschutz-Folgenabschätzung durchführen (Art. 22 DSG). Bei der Risikobeurteilung muss er die möglichen Schäden oder Folgen für die betroffene Person berücksichtigen, die mit einer möglichen Verletzung ihrer Persönlichkeit und ihres Rechts auf informationelle Selbstbestimmung verbunden sind. Zu den möglichen Schäden gehören z. B. die Verweigerung von Krediten oder die Verweigerung einer Anstellung.[129] Diese Bestimmungen können bei der Aufdeckung von Diskriminierung helfen. Die Folgenabschätzung enthält eine Beschreibung der geplanten Bearbeitung, eine Bewertung der Risiken für die Persönlichkeit oder die Grundrechte der betroffenen Person sowie die geplanten Massnahmen zum Schutz ihrer Persönlichkeit und ihrer Grundrechte. Zu beachten ist, dass es im privaten Sektor Ausnahmen von der Pflicht zur Durchführung einer Folgenabschätzung gibt (Art. 22 Abs. 4 und 5 DSG).

[129] CR LPD-PHILIPPE GILLIÉRON, Art. 22 N 24; DAVID ROSENTHAL/SAMIRA STUDER/ALEXANDRE LOMBARD (für die Übersetzung), La nouvelle loi sur la protection des données, Jusletter 16. November 2020, 58, N 149.

- *Informationspflicht im Allgemeinen und Informationspflicht bei einer automatisierten Einzelentscheidung und Auskunftsrecht (Art. 19, 21 und 25 Abs. 2 Bst. f DSG)*: Der für die Bearbeitung Verantwortliche muss die betroffene Person über die Beschaffung ihrer Daten informieren, sofern keine Ausnahme vorliegt. Darüber hinaus muss er die betroffene Person grundsätzlich über jede Entscheidung informieren, die ausschliesslich aufgrund einer automatisierten Bearbeitung von Personendaten getroffen wird und die für die betroffene Person rechtliche Folgen nach sich zieht oder sie erheblich beeinträchtigt. Die betroffene Person hat grundsätzlich das Recht, ihren Standpunkt darzulegen und zu verlangen, dass die Entscheidung von einer natürlichen Person überprüft wird. Diese Verpflichtungen gelten nicht, wenn die Entscheidung im Zusammenhang mit dem Abschluss oder der unmittelbaren Erfüllung eines Vertrags steht und dem Antrag der betroffenen Person stattgegeben wird, oder wenn diese ausdrücklich in eine automatisierte Entscheidung eingewilligt hat. Im Rahmen des Auskunftsrechts erhält die betroffene Person die erforderlichen Informationen, damit sie ihre Rechte gemäss DSG wahrnehmen kann und die Transparenz der Datenbearbeitung gewährleistet ist. Gegebenenfalls muss sie über das Vorliegen einer automatisierten Einzelentscheidung und über die Logik, auf der diese Entscheidung beruht, informiert werden (Art. 25 Abs. 2 DSG). Diese Bestimmungen können zur Aufdeckung von Diskriminierungen führen. Das Vorgesagte gilt jedoch nur für den Fall einer ausschliesslich automatisierten Entscheidung und nicht, wenn KI-Systeme als – wenn auch sehr wesentliche – Entscheidungshilfe eingesetzt werden.[130]

- *Dokumentations- und Protokollierungspflichten*: Gemäss Artikel 12 DSG und Artikel 3, 4 und 5 DSV ist der Verantwortliche verpflichtet, ein Verzeichnis der Bearbeitungstätigkeiten zu führen und unter bestimmten Voraussetzungen zu protokollieren. Zudem muss er technische und organisatorische Massnahmen treffen, um die Vertraulichkeit, die Verfügbarkeit, die Integrität und die Nachvollziehbarkeit der Daten zu gewährleisten. Diese Massnahmen können somit eine Kontrolle der durchgeführten Bearbeitung[131] und die Aufdeckung von Unregelmässigkeiten, die zu Diskriminierungen führen könnten, ermöglichen.

- *Grundsatz der Richtigkeit:* Gemäss Artikel 6 Absatz 5 DSG muss, wer Personendaten bearbeitet, sich über deren Richtigkeit vergewissern. Es sind alle angemessenen

[130] Der EuGH scheint den Begriff der automatisierten Einzelentscheidung weiter auszulegen. So hat er kürzlich entschieden, dass bereits das «Scoring» eines Unternehmens, das einen Kreditwürdigkeitsprüfung durchführt, eine solche Entscheidung darstellt, wenn sich der Empfänger bei seiner Entscheidung über einen Kreditantrag der betroffenen Person massgeblich auf dieses Ergebnis stützt (EuGH, Rechtssache C-634/21 vom 7. Dezember 2023, SCHUFA Holding [Scoring]). Die Auswirkungen dieser Rechtsprechung in der Schweiz sollten weiterverfolgt werden. In jedem Fall werden die Besonderheiten des Schweizer Rechts zu berücksichtigen sein.

[131] YVES POULLET, L'IA un défi pour nos législations vie privée, in: Astrid Epiney/Sophia Rovelli (Hrsg.): Künstliche Intelligenz und Datenschutz / L'intelligence artificielle et la protection des données, Freiburg 2021, 1 ff., 40, im Zusammenhang mit dem in der DSGVO vorgesehenen Grundsatz der Verantwortlichkeit.

Massnahmen zu treffen, damit Daten, die im Hinblick auf den Zweck ihrer Beschaffung oder Bearbeitung unrichtig oder unvollständig sind, berichtigt, gelöscht oder vernichtet werden.

- *Grundsatz von Treu und Glauben:* Einige Autoren vertreten die Ansicht, dass eine Bearbeitung von Personendaten, die zu einer Diskriminierung führt, als Verletzung des datenschutzrechtlichen Grundsatzes von Treu und Glauben betrachtet werden könnte (Art. 6 Abs. 2 DSG). Für eine solche Auslegung spricht, dass der Zweck des DSG der Schutz der Grundrechte und der Persönlichkeit der betroffenen Personen ist (Art. 1 DSG).[132]

Ungeachtet dessen vertreten verschiedene Autorinnen und Autoren die Auffassung, dass das schweizerische Diskriminierungsverbot heute – unabhängig davon, ob KI zur Anwendung gelangt oder nicht – bisweilen lückenhaft ist.[133] Auch das Schweizerische Kompetenzzentrum für Menschenrechte (SKMR) hat seinerzeit im Rahmen einer Studie, die der Bundesrat auf das Postulat Naef 12.3543 hin in Auftrag gegeben hatte, Empfehlungen zur Verstärkung der Gesetzgebung in diesem Bereich formuliert.[134] Der Bundesrat ist jedoch den meisten dieser Empfehlungen nicht gefolgt und hat auch auf die Einführung eines privatrechtlichen Diskriminierungsverbots in Ergänzung zu den Artikeln 27 ff. ZGB und auf eine Beweislasterleichterung im Diskriminierungsfall verzichtet, und auch den Anwendungsbereich von Artikel 261bis StGB nicht erweitert.[135] Der Bundesrat hat einen bereichsspezifischen Ansatz bevorzugt.

Die Überlegungen zu algorithmischer Diskriminierung kommen meist zu dem Schluss, dass der Einsatz von KI die bereits festgestellten Schwächen noch verschärft. So wird die Erarbei-

[132] In diesem Sinn: FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper: Diskriminierung (Fn. 127), 7.

[133] Insbesondere CR Cst.-VINCENT MARTENET, Art. 8, N 141 ff. und die zitierten Ref.; PREVITALI ADRIANO/MICHAEL MONTAVON, L'interdiction des discriminations, in: Oliver Diggelmann/Maya Hertig Randall/Benjamin Schindler (Hrsg.), Verfassungsrecht der Schweiz Bd. II / Droit constitutionnel suisse Vol. II, 2020, 1453 ff.; SAMANTHA BESSON, L'égalité horizontale: l'égalité de traitement entre particuliers, Freiburg 1999, 1352 ff.

[134] SCHWEIZERISCHES KOMPETENZZENTRUM FÜR MENSCHENRECHTE (SKMR), Der Zugang zur Justiz in Diskriminierungsfällen, Synthesebericht, Bern 2015, abrufbar unter: www.bj.admin.ch > Gesellschaft > Gleichstellung der Geschlechter und Schutz vor Diskriminierung > Schutz vor Diskriminierung > Frühere Berichte und Studien > Postulat Naef 12.3543 (abgerufen am 19. August 2024).

[135] Bericht des Bundesrates vom 25. Mai 2016 in Erfüllung des Postulats Naef 12.3543, Recht auf Schutz vor Diskriminierung, insbesondere Ziff. 4.2.1, 4.3.1, 4.2.3 und 4.2.6, abrufbar unter: www.bj.admin.ch > Gesellschaft > Gleichstellung der Geschlechter und Schutz vor Diskriminierung > Schutz vor Diskriminierung > Frühere Berichte und Studien > Postulat Naef 12.3543 (abgerufen am 27. August 2024).

tung eines allgemeinen Gleichbehandlungsgesetzes bzw. eines gesetzlichen Diskriminierungsverbots[136], die Einführung eines Grundsatzes im DSG, der diskriminierende Datenbearbeitungen verbietet,[137] die Einführung eines Grundsatzes der *Non-Discrimination by Design*[138] oder auch eine Umkehr bzw. Erleichterung der Beweislast gefordert.[139]

Zusammenfassend lässt sich sagen, dass die Problematik der Rechtsgleichheit und des Diskriminierungsverbots beim Einsatz von KI in einigen Aspekten durch die geltende Gesetzgebung abgedeckt wird.

Es stellt sich jedoch die Frage, ob der bestehende Rechtsrahmen gegebenenfalls ergänzt werden muss. Es ist darauf hinzuweisen, dass die im Zusammenhang mit KI geäusserte Kritik und die Forderungen hinsichtlich der Eignung des geltenden Rechts zur Lösung der anstehenden Probleme nicht neu sind. Sie entsprechen vielmehr den allgemeinen und bereits bekannten Forderungen.

Es ist jedoch zu erwarten, dass KI die bereits identifizierten Probleme verschärfen, beschleunigen und ihnen eine kollektive Dimension verleihen wird. Beim Diskriminierungsschutz sollte grundsätzlich nicht danach unterschieden werden, ob die Diskriminierung durch ein KI-System verursacht wird oder nicht. Ist der politische Wille vorhanden, wird jede allgemeine Stärkung des Diskriminierungsverbots, d. h. unabhängig von der verwendeten Technologie, auch im Bereich der KI von Vorteil sein.

Unabhängig von dieser Frage erscheint ein Eingreifen des Gesetzgebers notwendig, um die Transparenz und die Folgenabschätzung im Zusammenhang mit möglichen Diskriminierungen zu verbessern, da es nach wie vor schwierig ist, Verzerrungen im Zusammenhang mit KI-Systemen zu erkennen und nachzuweisen. Denkbar sind dabei folgende Massnahmen:

- Pflicht zur Berücksichtigung von Gleichstellungs- und Nichtdiskriminierungsaspekten bei der Risiko- und Folgenabschätzung von KI-Systemen (vgl. Ziff. 4.3.4, Art. 16 der KI- Konvention).

- Informationspflicht bei automatisierten Einzelentscheidungen: Derzeit besteht diese Pflicht nur bei Entscheidungen, die ausschliesslich auf einer automatisierten Bearbeitung von Personendaten beruhen (vgl. Art. 21 und 25 DSG). In vielen Fällen werden KI-Systeme jedoch nur als Unterstützung oder Hilfe bei der Entscheidungsfindung eingesetzt und fallen nicht unter diese Bestimmung. Eine Ausweitung der

[136] FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper: Diskriminierung (Fn. 127), 6; AlgorithmWatch/CH, Papier de position: Protection contre la discrimination algorithmique (Fn. 102), 8.

[137] FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper: Diskriminierung (Fn. 127), 6.

[138] FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA/WEINER/CHRISTOPH HEITZ, Diskriminierung (Fn. 104), 26 f.

[139] In diesem Sinn: FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper: Diskriminierung (Fn. 127), 6 f.

Informationspflicht auf teilautomatisierte Entscheidungen könnte daher von Vorteil sein.

- Meldung von KI-Systemen, die im öffentlichen Sektor eingesetzt werden: Derzeit führt das CNAI ein Register der im öffentlichen Sektor eingesetzten KI-Systeme. Eine solche Meldung erfolgt derzeit auf freiwilliger Basis. Die Meldung könnte obligatorisch werden.

- Bekämpfung des Black-Box-Effekts: Detaillierte Dokumentation, regelmässige Audits und Tests von KI-Systemen. Hier geht es darum, die Kontrolle über die Systeme zu behalten, indem sichergestellt wird, dass man weiterhin versteht, wie sie funktionieren, oder dass man sogenannte Halluzinationen erkennen kann. Die getroffenen Massnahmen müssen jedoch auch entgegenstehende Interessen berücksichtigen, die mit dem Interesse an der Offenlegung bestimmter Informationen kollidieren können, z. B. der Schutz von Daten Dritter, Geschäftsgeheimnisse oder Interessen der Strafverfolgung.

4.3.2.6 Artikel 11 – Privatsphäre und Schutz personenbezogener Daten

Artikel 11 der Konvention sieht vor, dass jede Vertragspartei Massnahmen ergreift oder beibehält, um sicherzustellen, dass das Recht des Einzelnen auf Schutz der Privatsphäre und seiner personenbezogenen Daten gewährleistet ist, unter anderem durch geltende nationale und internationale Gesetze, Normen und Rahmenvorschriften. Der Artikel sieht auch vor, dass wirksame Garantien und Schutzmassnahmen in Übereinstimmung mit den geltenden nationalen und internationalen rechtlichen Verpflichtungen eingeführt werden müssen.

Der erläuternde Bericht stellt fest, dass der Schutz der Privatsphäre und der Schutz personenbezogener Daten ein gemeinsamer Grundsatz ist, der für die Einhaltung fast aller anderen Grundsätze des Übereinkommens erforderlich ist. Die Beschaffung und Bearbeitung von Personendaten sind im KI-Bereich allgegenwärtig, und die eingesetzten Technologien und maschinengestützten Systeme wirken sich unmittelbar auf das Leben der Menschen aus. Da KI-Systeme in erster Linie datenbasiert sind, könnten die Tätigkeiten während des gesamten Lebenszyklus dieser Systeme ohne angemessene Garantien ernsthafte Risiken mit sich bringen.[140]

Herausforderungen bestehen in allen Phasen des Lebenszyklus von KI-Systemen, insbesondere bei generativer KI. Je nachdem, ob man sich in der Konzeptions-, Trainings-, Nutzungs- oder Feedbackphase des Systems befindet, ergeben sich unterschiedliche Problematiken.[141]

---

[140] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 79.

[141] SAMUEL KLAUS, KI trifft Datenschutz – Risiken und Lösungsansätze, in: Astrid Epiney/Sophia Rovelli (Hrsg.): Künstliche Intelligenz und Datenschutz (Fn. 131), 81 ff., 83 f.; DAVID ROSENTHAL, Datenschutz und KI: Worauf in der Praxis zu achten ist, Jusletter 26. April 2022, 3; vgl. auch

Es ist zu prüfen, ob der bestehende regulatorische Rahmen im nationalen Recht, der auch auf den Bereich der KI anwendbar ist, die Risiken dieser Technologie ausreichend abdeckt.

Auf Verfassungsstufe werden das Grundrecht auf Privatleben und Datenschutz durch Artikel 10 Absatz 2 BV (persönliche Freiheit) sowie spezifischer durch Artikel 13 BV (in geringerem Masse durch Art. 7) geschützt.[142] Artikel 13 sieht einerseits vor, dass jede Person Anspruch auf Achtung ihres Privat- und Familienlebens, ihrer Wohnung, sowie ihres Brief-, Post- und Fernmeldeverkehr hat (Abs. 1), und andererseits, dass jede Person Anspruch auf Schutz vor Missbrauch ihrer persönlichen Daten hat (Abs. 2). In diesem zweiten Absatz ist das Recht auf informationelle Selbstbestimmung verankert, d. h. das Recht der betroffenen Person, selbst bestimmen zu können, ob und zu welchem Zweck Informationen über sie bearbeitet werden dürfen.[143]

Auf Gesetzesstufe wird Artikel 13 BV für natürliche Personen hauptsächlich durch das DSG und seine Ausführungsverordnungen, namentlich die DSV und die VDSC, gewährleistet. Diese Bestimmungen gelten sowohl für den privaten als auch für den öffentlichen (Bundes-) Sektor. Das Datenschutzgesetz wurde kürzlich revidiert; die Änderungen traten am 1. September 2023 in Kraft. Dieser rechtliche Rahmen wird durch zahlreiche sektorspezifische Bundesgesetze ergänzt. Hervorzuheben sind hier das Fernmeldegesetz[144], welches das Fernmeldegeheimnis konkretisiert, und die in Artikel 1 des Bundesgesetzes betreffend die Überwachung des Post- und Fernmeldeverkehrs[145] aufgeführten Gesetze, welche die Voraussetzungen festlegen, unter denen eine Überwachung durchgeführt werden darf. Ebenfalls relevant sind weitere Bestimmungen wie die Artikel 28 ff. ZGB, die insbesondere das Recht auf Privatleben und das Recht am eigenen Bild schützen[146], oder die Artikel 179 ff. StGB (strafbare Handlungen gegen den Geheim- oder Privatbereich).

European Data Protection Board, Report of the work undertaken by the ChatGPT Taskforce, 23. Mai 2024, 4 N 14, abrufbar unter: https://www.edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-chatgpt-taskforce_en (abgerufen am 19. August 2024).

142 CR LPD-BERTIL COTTIER, Art. 1 N 19, BSK DSG/BGÖ - MATTHIAS R. SCHÖNBÄCHLER/URS MAURER-LAMBROU/SIMON KUNZ, Art. 1, N 5.

143 BGE **145** IV 42, E., 4.2; JULIEN FRANÇAIS, in: Petit commentaire LPD (Fn. 44), Art. 1 N 12.

144 SR **784.10**.

145 SR **780.1**.

146 BSK ZGB I-ANDREAS MEILI, Art. 28 N 17.

Das DSG und seine Verordnungen sind auf KI-Systeme anwendbar, wenn diese – was häufig der Fall ist – Personendaten verwenden.[147] Die für die Datenbearbeitung Verantwortlichen und in bestimmten Fällen auch die Auftragsbearbeiter müssen daher die zahlreichen in diesen Texten vorgesehenen Verpflichtungen einhalten. Dazu gehören insbesondere die Einhaltung der allgemeinen Grundsätze (Rechtmässigkeit, Erkennbarkeit, Zweckbindung, Richtigkeit, Verhältnismässigkeit, Treu und Glauben, Datenschutz durch Technik und durch datenschutzfreundliche Voreinstellungen, Sicherheit; vgl. Art. 6, 7 und 8 DSG), die Meldepflicht für automatisierte Bearbeitungen und das Führen von Verzeichnissen (Art. 12 DSG und Art. 31 DSV), die Dokumentations- und Protokollierungspflichten (Art. 3 bis 5 DSV), die Informationspflichten (Art. 19 DSG), insbesondere bei automatisierten Entscheidungen (Art. 21 DSG), die Informationspflichten bei Ausübung des Auskunftsrechts (Art. 25 DSG), die Pflicht zur Durchführung einer Folgenabschätzung (Art. 22). Diese verschiedenen Elemente werden in Art. 10 der Konvention (Gleichstellung und Nichtdiskriminierung) näher erläutert (vgl. Ziff. 4.3.2.5).

Für die Bearbeitung von Personendaten im öffentlichen Bereich gilt zudem der Grundsatz der gesetzlichen Grundlage (Ziff. 4.2.3.1.2). Das DSG hat diesen Grundsatz konkretisiert. So bedarf der staatliche Einsatz eines Gesichtserkennungsalgorithmus, der biometrische Personendaten (besonders schützenswerte Personendaten) bearbeitet oder Vorhersagen aufgrund von Profiling macht, eines Gesetzes im formellen Sinn (Art. 34 Abs. 2 Bst. a und b DSG). Zudem kann auch der blosse Einsatz von KI unter diese Voraussetzung fallen, wenn der Zweck oder die Art und Weise der Bearbeitung von Personendaten voraussichtlich schwerwiegende Auswirkungen auf die Grundrechte der betroffenen Person hat (Art. 34 Abs. 2 Bst. c DSG).[148] Das DSG verfolgt damit einen «risikobasierten» Ansatz.

Bei ungerechtfertigter Verletzung der vorgenannten Pflichten kann die betroffene Person von der für die Datenbearbeitung verantwortlichen Person insbesondere die Einstellung der Datenbearbeitung, die Berichtigung unrichtiger Daten oder die Löschung der Daten verlangen (Art. 30 und 41 DSG). Sie kann sich auch an den Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten (EDÖB) wenden, der Datenbearbeitungen, die gegen das DSG verstossen, verbieten kann. Der EDÖB kann auch von Amtes wegen tätig werden.

Allerdings ist an dieser Stelle anzumerken, dass die Entwicklung von KI-Systemen sehr oft im Widerspruch zu den allgemeinen Grundsätzen des DSG steht. Eine umfassende Beschaffung von Personendaten aus einer Vielzahl von Quellen für das Training von KI-Systemen ist

---

[147] FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper: Datenschutz, Juni 2024, 3, abrufbar unter: https://www.itsl.uzh.ch > Wissenstransfer > Publikationen > Positionspapiere > Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von künstlicher Intelligenz (abgerufen am 19. August 2024).

[148] BJ, Totalrevision des Bundesgesetzes über den Datenschutz (DSG) – Übersicht zu den wichtigsten Änderungen (Fn. 38), 10 ff., Ziff. 2.2.1; siehe auch BJ, Gesetzgebungsleitfaden Datenschutz (Fn. 44).

schlecht vereinbar mit den Grundsätzen der Datenminimierung, der Erkennbarkeit, der Richtigkeit oder der Zweckbindung.[149] Es stellt sich die Frage, ob der private Datenbearbeiter in diesem Fall ein überwiegendes Interesse geltend machen kann, insbesondere den Rechtfertigungsgrund der Bearbeitung zu nicht personenbezogenen Zwecken, namentlich für Forschung, Planung oder Statistik (vgl. Art. 31 Abs. 2 Bst. e DSG). Dies ist im Einzelfall und insbesondere in Abhängigkeit von den jeweiligen Interessen und der Art des KI-Systems zu beurteilen.[150] Im öffentlichen Sektor wird die Anwendbarkeit von Art. 39 DSG ebenfalls im Einzelfall beurteilt.[151]

Es sei noch darauf hingewiesen, dass im Rahmen der Totalrevision des DSG das StGB um eine Bestimmung zum Identitätsmissbrauch (Art. 179^{decies} StGB) erweitert wurde, welche die zahlreichen Bestimmungen zum Schutz der Privatsphäre und der Daten ergänzt. Diese Bestimmung wird im Zusammenhang mit KI und «Deepfakes» sicherlich an Bedeutung gewinnen (vgl. Ziff. 6.6.1).

Schliesslich sehen die Leitlinien des Bundes vor, dass die vom Bund eingesetzten KI-Technologien so zu gestalten sind, dass die Privatsphäre gemäss den Vorschriften geschützt und die Datenschutzbestimmungen jederzeit eingehalten werden.[152]

Zusammenfassend lässt sich sagen, dass KI zahlreiche Herausforderungen für den Schutz der Privatsphäre und den Datenschutz mit sich bringt, die jedoch durch den bestehenden Rechtsrahmen relativ gut abgedeckt sind.

Gleichwohl ist ein Tätigwerden des Gesetzgebers unter dem Gesichtspunkt der Transparenz nicht auszuschliessen, um eine wirksamere Umsetzung der bestehenden Garantien zu ermöglichen:

- Informationspflicht bei automatisierten Einzelentscheidungen: Derzeit besteht diese Pflicht nur für Entscheidungen, die ausschliesslich auf einer automatisierten Datenbearbeitung beruhen. In vielen Fällen werden KI-Systeme jedoch nur zur Unterstützung oder als Hilfe bei der Entscheidungsfindung eingesetzt und fallen nicht unter diese Vorschrift.

[149] In diesem Sinne im Zusammenhang mit Gesichtserkennungssoftware, vgl. NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 16 N 39.

[150] DAVID ROSENTHAL, Datenschutz beim Einsatz generativer künstlicher Intelligenz, Jusletter 6. November 2023, 12 N 34 ff.; ROLF H. WEBER, Künstliche Intelligenz und Datenschutz, Jusletter IT 4. Juli 2024, 7 N 19 ff.

[151] Die Problematik der Weiterverwendung von Daten wird im Rahmen der Arbeiten zur Erfüllung der Motion 22.3890 «Rahmengesetz für die Sekundärnutzung von Daten» vom 22. August 2022 untersucht.

[152] Leitlinien «Künstliche Intelligenz» für den Bund (Fn. 80), 3.

- Meldung der eingesetzten KI-Systeme: Das CNAI führt aktuell ein Register der im öffentlichen Sektor eingesetzten KI-Systeme. Eine Meldung ist derzeit freiwillig, könnte aber verpflichtend werden.

- Bekämpfung des Black-Box-Effekts: detaillierte Dokumentation, regelmässige Audits und Tests von KI-Systemen. Hier geht es darum, die Kontrolle über die Systeme zu behalten, indem sichergestellt wird, dass man weiterhin versteht, wie sie funktionieren, oder dass man sogenannte Halluzinationen erkennen kann. Wie bereits erwähnt, müssen die getroffenen Massnahmen den entgegenstehenden Interessen Rechnung tragen, die mit dem Interesse an der Offenlegung bestimmter Informationen kollidieren können, wie z. B. der Schutz der Daten Dritter, Geschäftsgeheimnisse oder Interessen der Strafverfolgung.

- Zudem muss der Gesetzgeber die Koordination zwischen einer allfälligen Regulierung der KI und dem DSG regeln, insbesondere im Hinblick auf die Pflicht zur Durchführung einer Folgenabschätzung.

4.3.2.7 Artikel 12 – Zuverlässigkeit

Artikel 12 sieht vor, dass jede Vertragspartei geeignete Massnahmen zur Förderung der Zuverlässigkeit von KI-Systemen und des Vertrauens in deren Ergebnisse trifft. Dies kann angemessene Qualitäts- und Sicherheitsanforderungen für den gesamten Lebenszyklus von KI-Systemen umfassen.

Gemäss dem erläuternden Bericht unterstreicht diese Bestimmung die wichtige Rolle, die insbesondere Normen, technische Spezifikationen oder Absicherungstechniken bei der Überwachung und Überprüfung der Zuverlässigkeit von KI-Systemen und bei der transparenten Dokumentation und Kommunikation von Nachweisen für diesen Prozess spielen können. Der Schwerpunkt liegt dabei auf spezifischen Schlüsselelementen in der Funktionsweise von KI-Systemen wie Zuverlässigkeit, Robustheit, Sicherheit, Genauigkeit und Leistung sowie auf funktionalen Voraussetzungen wie Qualität, Datenintegrität und -sicherheit sowie Cybersicherheit. Ziel dieser technischen Standards ist die Schaffung eines begründeten Vertrauens der Öffentlichkeit. Soweit deren Entwicklungsprozess transparent und inklusiv ist, können diese Normen dazu beitragen, eine gegenseitig verstandene und erweiterbare Sicherheit und Konformität im KI-Bereich zu schaffen.[153]

In diesem Rahmen sollen Massnahmen getroffen werden, die darauf abzielen, KI-Systeme wie jedes andere Softwaresystem technisch sicher und zuverlässig zu machen. Sicherheit und Zuverlässigkeit sollten als Grundvoraussetzungen für den gesamten Lebenszyklus von KI-Systemen und nicht als blosse Eigenschaften betrachtet werden. Je nach Einzelfall können die Massnahmen die Bereitstellung klarer Informationen darüber umfassen, wie die KI-Akteure diese Standards praktisch umgesetzt haben. Dies bedeutet, dass die Verantwortlichkeit von Anfang bis Ende mittels ausreichender Transparenz der Prozesse und Dokumentationsprotokolle sicherzustellen ist. Zwischen dieser Bestimmung und Artikel 8 (Transparenz

[153] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 84 ff.

und Aufsicht) sowie Artikel 9 (Rechenschaftspflicht und Verantwortung) besteht ein klarer Zusammenhang. Technische Normen spielen eine wichtige Rolle bei der Risikobewertung und -minderung, wenn die bestehenden Vorschriften keine ausreichenden Vorgaben enthalten. Dies gilt insbesondere für den Rahmen für das Risiko- und Folgenmanagement.[154]

Die schweizerische Datenschutzgesetzgebung sieht relativ weitgehende Verpflichtungen bezüglich der Sicherheit von Personendaten vor. Die Einhaltung der in Artikel 6 DSG festgelegten Grundsätze, die Grundsätze des Datenschutzes durch Technik und durch datenschutzfreundliche Voreinstellungen (Art. 7 DSG) sowie die Vorschriften über die Datensicherheit (Art. 8 DSG, ergänzt durch Art. 1–6 DSV), die Zertifizierung (Art. 13 DSG) und die Meldung von Verletzungen der Datensicherheit (Art. 24 DSG) gewährleisten einen kohärenten Schutz. Was insbesondere die Zertifizierung betrifft, sieht die VDSZ (Art. 6) vor, dass der EDÖB Richtlinien über die Mindestanforderungen für ein Managementsystem erlässt, wobei insbesondere folgende technische Normen zu berücksichtigen sind: SN EN ISO 9001, SN EN ISO/IEC 27001 und SN EN ISO/IEC 2770.

Das am 1. Januar 2024 in Kraft getretene Informationssicherheitsgesetz vom 18. November 2020[155] (ISG) und die dazugehörenden Verordnungen[156] enthalten ebenfalls Regeln zur Gewährleistung der Sicherheit der Bearbeitung von Informationen im Zuständigkeitsbereich des Bundes und der Sicherheit seiner Informatikmittel (Art. 1 Abs. 1 ISG). Dieses Gesetz legt den Schwerpunkt auf die Vereinheitlichung der Massnahmen. In diesem Rahmen kann es auch zum Schutz von KI-Systemen beitragen, die ohne Personendaten betrieben werden können.[157]

Die KI-Leitlinien des Bundes halten zudem fest, dass KI-Systeme sicher, robust und resilient konzipiert sein müssen, damit sie sich positiv auf Mensch und Umwelt auswirken und nicht missbraucht oder missbräuchlich eingesetzt werden können. Es sind geeignete Massnahmen zu treffen, um schwerwiegende Fehlentscheidungen zu vermeiden.[158]

In diesem Zusammenhang stellt sich die Frage nach der Bedeutung harmonisierter Normen bei der Konformitätsbewertung eines KI-Systems mit dem rechtlichen Rahmen. Im Folgenden wird erläutert, dass die KI-Verordnung der EU unter bestimmten Bedingungen, bei Einhaltung

---

[154] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 87 ff.

[155] SR **128**.

[156] Informationssicherheitsverordnung vom 8. November 2023 (SR **128.1**); Verordnung vom 8. November 2023 über das Betriebssicherheitsverfahren (SR **128.41**); Verordnung vom 19. Oktober 2016 über die Identitätsverwaltungs-Systeme und Verzeichnisdienste des Bundes (SR **172.010.59**).

[157] Universität Zürich, Digital Society Initiative, Positionspapier, Ein Rechtsrahmen für Künstliche Intelligenz (Fn. 113), 6, das zum damaligen Zeitpunkt die Schaffung eines allgemeinen Gesetzes zur Informatiksicherheit forderte.

[158] Leitlinien «Künstliche Intelligenz» für den Bund (Fn. 80), 5.

der harmonisierten Normen von der Vermutung der Konformität mit der Verordnung ausgeht (vgl. Ziff. 5.2.11).

Zusammenfassend lässt sich sagen, dass die Schweiz bereits gewisse Standards im Bereich des Datenschutzes und der Informationssicherheit anwendet, welche die Zuverlässigkeit und die funktionalen Voraussetzungen eines KI-Systems wie Qualität, Integrität, Datensicherheit und Cybersicherheit verbessern können.

Im Falle einer Ratifizierung der KI-Konvention muss die Schweiz ihre Kontakte und ihre Zusammenarbeit mit den Normierungsgremien zur Entwicklung von Normen und Standards im Bereich der KI fortsetzen.

Gegenwärtig sieht das schweizerische Recht keinen Mechanismus vor, der es erlaubt, von der Einhaltung bestimmter standardisierter KI-Normen auf die Konformität mit dem geltenden Rechtsrahmen zu schliessen. Im Falle einer schweizerischen Regulierung von KI würde sich unweigerlich die Frage nach der rechtlichen Tragweite von Normen stellen.

4.3.2.8 Artikel 13 – Sichere Innovation

Gemäss Artikel 13 wird jede Vertragspartei aufgefordert, zur Förderung von Innovation und zur Vermeidung negativer Auswirkungen auf Menschenrechte, Demokratie und Rechtsstaatlichkeit die Schaffung kontrollierter Umgebungen für die Entwicklung, Erprobung und die Testung von KI-Systemen unter Aufsicht der zuständigen Behörden zu ermöglichen.

Es sollen Anreize geschaffen werden, damit bereits in den frühen Phasen der Konzeption von KI-Systemen Fragen der System- und Datenqualität, Sicherheit und Absicherung sowie Menschenrechtsbelange berücksichtigt werden. Dies ist umso wichtiger, als bestimmte Risiken erst nach Tests erkannt und/oder nur in der Konzeptionsphase wirksam angegangen werden können.[159]

Die Konvention überlässt die Art und Weise der Umsetzung dieser Bestimmung den Vertragsparteien. Laut dem erläuternden Bericht kann es sich dabei insbesondere um «regulatorische Sandboxen» («regulatory sandboxes», bzw. Reallabore), informelle regulatorische Leitlinien oder Garantien der Verfahrensaussetzung («no-action letter»)[160] handeln, um zu

[159] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 90 ff.

[160] Gemäss dem erläuternden Bericht sollen Garantien der Verfahrensaussetzung klarstellen, wie die Regulierungsbehörden mit der Gestaltung, der Entwicklung oder dem Einsatz von KI-Systemen in neuen Kontexten umgehen werden (vgl. erläuternder Bericht KI-Konvention [Fn. 20], N 92). Sie kann auch in einer behördlichen Zusage bestehen, keine Aufsichtsmassnahmen zu ergreifen, solange sich die Unternehmen an die getroffenen Vereinbarungen oder an die von den Behörden gegebenen Auslegungshilfen zu den anwendbaren Gesetzen halten. Dies gibt den Betroffenen Rechts- und Planungssicherheit, insbesondere in Fällen, in denen unklar ist, wie ein Geschäftsmodell rechtlich zu behandeln ist. Garantien der Verfahrensaussetzung bedeutet für Unternehmen, dass während der Testphase in der Sandbox keine Rechtsanwendung erfolgt (vgl. STEPHANIE VOLZ, KI Sandboxen für die Schweiz?, SZW 2022, 51 ff., 63 f.).

klären, wie die Regulierungsbehörden die verschiedenen Lebenszyklen der KI angehen werden.[161]

Eine Sandbox beinhaltet eine sehr breite Palette an Instrumenten, deren Merkmale sich sehr oft vermischen. Die Analyse folgt hier der Kategorisierung des vom SECO im Jahr 2022 erteilten «Prüfauftrags zu Regulatory Sandboxes». Demnach umfasst eine Sandbox im weiteren Sinne zwei Hauptinstrumente: «Pilotprojekte» mit experimenteller Regulierung (im Folgenden: Pilotprojekte) und eine Sandbox im engeren Sinne.[162] In beiden Fällen wird durch die Behörden ein Rahmen festgelegt und überwacht.

Gemäss der obgenannten Studie dienen Pilotprojekte dazu, neue Technologien oder Verfahren für einen begrenzten Zeitraum unter realen Bedingungen zu testen, wobei in bestimmten Aspekten von bestehenden Vorschriften abgewichen werden kann. Ihre Ergebnisse zeigen, inwieweit die Vorschriften überarbeitet werden müssen.[163] Eine Sandbox im engeren Sinne ermöglicht es Unternehmen, Verfahren, Produkte und Dienstleistungen am Markt zu erproben, indem sie zeitlich befristet von bestimmten Vorschriften ausgenommen wird. Verstösse gegen solche Regeln werden korrigiert, aber nicht sanktioniert. Dies erlaubt Unternehmen herauszufinden, ob ihre innovativen Geschäftsmodelle innerhalb des bestehenden rechtlichen Rahmens bestehen können oder nicht.[164] Sandbox-Tests können sowohl unter realen Bedingungen als auch in einer simulierten und/oder kontrollierten Umgebung durchgeführt werden.[165]

Gemäss SECO-Studie sind die Anforderungen an die gesetzliche Grundlage bei der Entwicklung einer Sandbox weniger streng. Während Pilotprojekte oft ein Gesetz im formellen Sinn als Grundlage benötigen, genügt für eine Sandbox im engeren Sinn je nach Fall eine Verordnung oder sogar nur eine Änderung der behördlichen Praxis.[166]

---

[161] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 92.

[162] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN/ANNICK PIETZONKA/GUIDO SAURER, Prüfauftrag zu Regulatory Sandboxes, in SECO (Hrsg.), Grundlagen für die Wirtschaftspolitik, Juni 2022, 6.

[163] Vergleiche auch BJ, Gesetzgebungsleitfaden 4. Auflage 2019, mit Aktualisierung 2023, N 1044 ff., abrufbar unter: www.bj.admin.ch > Staat & Bürger > Legistik > Legistische Hauptinstrumente (abgerufen am 25. Juli 2024).

[164] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN u. a., Prüfauftrag zu Regulatory Sandboxes (Fn. 162), N 2.1.2 f.

[165] Vgl. auch Art. 3 Nr. 55 KI-Verordnung.

[166] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN u. a., Prüfauftrag zu Regulatory Sandboxes (Fn. 162), N 2.2.2 und 2.2.3.

Eine Alternative zur Sandbox im weiteren Sinne sind sogenannte «risikobasierte Regulierungen». In diesem Fall wird die Regulierung proportional zu den von den Unternehmen eingegangenen Risiken angepasst. Die eingeführten regulatorischen Vereinfachungen sind dauerhaft. Unternehmen mit relativ geringen Risiken für Kunden und Wirtschaft profitieren so von vereinfachten Vorschriften.[167] Ein weiteres wichtiges Instrument, das ebenfalls in diesen Rahmen der sicheren Innovation passt und eine Ergänzung oder sogar eine Alternative zur Sandbox im engeren Sinne darstellt, sind die «Innovation Hubs». Das sind Plattformen, auf denen sich Behörden, Unternehmen, Universitäten und Investoren über technologische Trends und branchenspezifisches Wissen austauschen können.[168]

Die schweizerische Gesetzgebung enthält Bestimmungen, die Pilotprojekte in vielen sektoriellen Bereichen zulassen. Die meisten dieser Bestimmungen sind so neutral gehalten, dass sie auch auf KI anwendbar sind.

Im Bereich des Datenschutzes sieht Art. 35 DSG vor, dass der Bundesrat die automatisierte Bearbeitung von besonders schützenswerten Personendaten oder andere Datenbearbeitungen im Sinne von Artikel 34 Absatz 2 Buchstaben b und c bereits vor Inkrafttreten eines formellen Gesetzes bewilligen kann, wenn: 1. die Aufgaben, für welche die Bearbeitung erforderlich ist, in einem bestehenden Gesetz formell geregelt sind; 2. ausreichende Massnahmen getroffen werden, um den Eingriff in die Grundrechte der betroffenen Personen auf ein Minimum zu beschränken; und 3. für die praktische Umsetzung der Datenbearbeitung, insbesondere aus technischen Gründen, eine Testphase vor dem Inkrafttreten unerlässlich ist. Diese Bedingungen müssen kumulativ erfüllt sein.[169] Das Genehmigungsverfahren durch den EDÖB und die Modalitäten sind in der DSV (Art. 32 ff.) detailliert beschrieben. Pilotversuche können dazu dienen, die Funktionsweise neuer Technologien, mit denen Personendaten bearbeitet werden, zu testen[170]; dies gilt also möglicherweise auch für KI-Systeme.

Ein weiteres Beispiel für ein Pilotprojekt, das für KI zur Anwendung gelangen könnte, ist Artikel 15 des teilweise am 1. Januar 2024 in Kraft getretenen Bundesgesetzes vom 17. März

---

[167] Dies ist beispielsweise in der Schweiz bei der Fintech-Regulierung der Fall, die es ermöglicht, Unternehmen, die Publikumseinlagen von bis zu einer Million Franken verwalten, von der Bewilligungspflicht zu befreien (vgl. YVES SCHNEIDER/PATRICK ZENHÄUSERN/PETER HETTICH/ROGER KÜTTEL, Welche Regulierungen begünstigen Innovation, Die Volkswirtschaft – Plattform für Wirtschaftspolitik, 23. Juni 2022, abrufbar unter: https://dievolkswirtschaft.ch/de/2022/06/welche-regulierungen-beguenstigen-innovationen/ [abgerufen am 25. Juli 2024]).

[168] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN U. A., Prüfauftrag zu Regulatory Sandboxes (Fn. 162), N 2.2.4. Die Schweiz verfügt über zahlreiche Innovation Hubs, die einen Informations- und Beratungsaustausch zwischen den verschiedenen KI-Akteuren ermöglichen. Auf Bundesebene wäre zum Beispiel Innosuisse zu nennen, der Schweizerische Innovationspark oder das KI-Zentrum der EPFL. Weitere wichtige Zentren gibt es auf kantonaler Ebene.

[169] BSK DSG/BGÖ-ANDREAS STÖCKLI/CHRISTOPH GRÜNINGER, Art. 35 N 8; CR DSG-ASTRID EPINEY/SAMAH POSSE, Art. 35 N 2.

[170] MONIQUE COSSALI SAUVAIN, in: Petit commentaire LPD (Fn. 44), Art. 35 N 20.

2023 über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben[171] (EMBAG). Dieses Gesetz bezweckt die Förderung der elektronischen Abwicklung von Geschäftsprozessen im Bund (Grundsatz «digital first»). Diese Prozesse umfassen die Interaktion der Behörden aller Staatsebenen untereinander sowie der Behörden zu Unternehmen und zur Bevölkerung. Im EMBAG werden im Wesentlichen die allgemeinen Rahmenbedingungen für verschiedene Tätigkeiten festgelegt, namentlich die Entwicklung des Einsatzes von E-Government auf Bundesebene, die Zusammenarbeitsformen des Bundes mit anderen Gemeinwesen und Organisationen im Bereich E-Government sowie für die elektronischen Behördenleistungen des Bundes.[172] Artikel 15 EMBAG sieht auch die Möglichkeit der Durchführung von Pilotversuchen vor. Wenn Letztere Bearbeitungen beinhalten, die in den Anwendungsbereich von Artikel 35 DSG fallen, dürfte von den dort festgelegten Voraussetzungen nicht abgewichen werden können.[173]

Weitere Beispiele finden sich im sektoriellen Recht. So ermächtigt Artikel 106 Absatz 5 SVG den Bundesrat, Pilotprojekte im Bereich des automatisierten Fahrens zu bewilligen. Mehr als zehn Projekte wurden bereits auf dieser Grundlage realisiert.[174] Auch im Bereich der Stromversorgung können unter bestimmten Voraussetzungen Pilotprojekte zur Entwicklung innovativer Technologien, Geschäftsmodelle oder Produkte im Energiesektor bewilligt werden.[175] Pilotversuche werden auch in der Krankenversicherung[176] zur Dämpfung des Kostenwachstums im Gesundheitswesen, zur Stärkung der Qualität oder zur Förderung der Digitalisierung, in der Invalidenversicherung zur Erleichterung der Eingliederung[177] oder in der Berufsbildung[178] durchgeführt.

Die obigen Ausführungen lassen den Schluss zu, dass das Bundesrecht Bestimmungen enthält, die gegebenenfalls die Durchführung von Pilotprojekten im KI-Bereich ermöglichen. Das Spektrum der Innovationszentren in der Schweiz ist ebenfalls sehr reichhaltig. Es fällt jedoch auf, dass keine oder wenige Sandboxen im engeren Sinne, die es Unternehmen ermöglichen würden, innovative Systeme unter realen oder simulierten Bedingungen zu testen, entwickelt werden. Eine solche Sandbox wäre aber gemäss einer Umfrage im Rahmen der vom SECO in Auftrag gegebenen Studie zur Prüfung regulatorischer Sandboxen gerade im Zusammen-

---

171 SR **172.019**.

172 Pressemitteilung vom 22. November 2023, Bundesrat setzt Bundesgesetz über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben in Kraft, abrufbar unter: www.admin.ch > Dokumentation > Medienmitteilungen > Bundesrat setzt Bundesgesetz über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben in Kraft (abgerufen am 27. August 2024).

173 MONIQUE COSSALI SAUVAIN, Petit commentaire LPD (Fn. 44), Art. 35 N 11 ff.

174 Vgl. die auf der Website des ASTRA veröffentlichte Liste, «Intelligente Mobilität in der Schweiz: Übersicht über abgeschlossene Pilotversuche», Stand 3. Februar 2023, abrufbar unter: www.astra.admin.ch > Themen > Intelligente Mobilität > Pilotversuche (abgerufen am 25. Juli 2024).

175 Art. 23*a* des Stromversorgungsgesetzes (SR **734.7**).

176 Art. 59*b* des Bundesgesetzes über die Krankenversicherung (SR **832.10**).

177 Art. 68^{quater} des Bundesgesetzes über die Invalidenversicherung (SR **831.20**).

178 Art. 4 des Bundesgesetzes über die Berufsbildung (SR **412.10**).

hang mit KI und Datenschutz sinnvoll. Datenschutzbestimmungen bzw. Unsicherheiten bezüglich ihrer konkreten Anwendung würden die Entwicklung von potenziell gesellschaftlich nützlichen Anwendungen behindern.[179]

Vor diesem Hintergrund sollte im Rahmen einer möglichen Umsetzung der KI-Konvention näher geprüft werden, ob die bestehenden Regelungen zu Pilotprojekten bereits eine ausreichende Grundlage für die Entwicklung einer (ggf. sektorspezifischen) Sandbox für KI bieten. Gegebenenfalls ist eine Einrichtung solcher Instrumente zu prüfen. Dabei ist zu beachten, dass einige Autoren die Sandbox als «technisch-organisatorische Massnahme» zur Begrenzung von Grundrechtseingriffen im Rahmen von Pilotprojekten unter Federführung des Bundes ansehen.[180]

Eine Inspirationsquelle in diesem Zusammenhang könnte die «*Innovation Sandbox*»[181] sein, die 2022 vom Kanton Zürich ins Leben gerufen wurde. Dieses Projekt schafft eine Testumgebung, in der Akteure KI-Projekte innerhalb eines klar definierten Rahmens durchführen können. Unterschiedliche Organisationen wie Start-ups, KMUs, Grossunternehmen oder Forschungsinstitute erhalten durch diese Sandbox Zugang zu regulatorischem Know-how und neuen Datenquellen. Im Gegenzug werden alle Erkenntnisse und Ergebnisse öffentlich zugänglich gemacht. Im Gegensatz zu vielen ausländischen Ansätzen geht die «Innovation Sandbox» für KI insofern einen Schritt weiter, als einige der eingereichten Projekte nicht nur geprüft, sondern auch in die Praxis umgesetzt werden. Ziel ist es, verantwortungsvolle Innovation unter Berücksichtigung rechtlicher und ethischer Kriterien zu fördern und die Verbreitung von KI in Verwaltung, Wirtschaft und Forschung zu unterstützen.

Der schweizerische rechtliche Rahmen sieht sektorspezifische Möglichkeiten für Pilotprojekte im KI-Bereich vor. Auch die Innovationzentren sind hier gut vertreten. Die Schaffung von kontrollierten Umgebungen, in denen Unternehmen ihre Produkte und Systeme unter realen Bedingungen testen können (Sandbox im engeren Sinne), scheint hingegen wenig oder gar nicht entwickelt zu werden. Der diesbezügliche Bedarf wird im Rahmen der Umsetzung der Konvention genauer zu ermitteln sein.

[179] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN et al., Prüfauftrag zu Regulatory Sandboxes (Fn. 162), N 3.1. Ebenso: STEPHANIE VOLZ, KI Sandboxen für die Schweiz? (Fn. 160), 67 f., die sich für die Einführung einer Experimentierklausel im DSG ausspricht.

[180] SANDRA HUSI-STÄMPFLI/ANNE-SOPHIE MORAND, Datenschutzrecht, Zürich 2024, 180 N 336.

[181] Innovation Sandbox for Artificial Intelligence (AI), abrufbar unter: https://www.zh.ch > Wirtschaft & Arbeit > Wirtschaftsstandort > Innovation-Sandbox für KI (abgerufen am 25. Juli 2024).

4.3.3 Kapitel IV: Rechtsmittel

4.3.3.1 Allgemeines

Kapitel IV ist den Rechtsmitteln gewidmet. Artikel 14 enthält Bestimmungen über Massnahmen, die zu ergreifen sind, um zugängliche und wirksame Rechtsmittel gegen durch KI-Systeme entstandene Menschenrechtsverletzungen zu gewährleisten. Artikel 15 behandelt Verfahrensgarantien im Allgemeinen.

4.3.3.2 Artikel 14 – Rechtsmittel

Artikel 14 Absatz 1 – Zugängliche und wirksame Beschwerde

Gemäss Artikel 14 Absatz 1 ergreift jede Vertragspartei im Einklang mit ihrer innerstaatlichen Rechtsordnung Massnahmen oder behält diese bei, um die Verfügbarkeit zugänglicher und wirksamer Rechtsmittel gegen Menschenrechtsverletzungen sicherzustellen, die sich aus Tätigkeiten im Lebenszyklus von KI-Systemen ergeben.

Diese Bestimmung verlangt keine Schaffung neuer Rechtsmittel. Sie fordert die Vertragsparteien jedoch auf, die bestehenden Rechtsmittel bei Menschenrechtsverletzungen durch Massnahmen zu stärken, die den von KI-Systemen ausgehenden Gefahren Rechnung tragen. KI-Systeme sind technisch sehr komplex und ihre Funktionsweise ist oft undurchsichtig. Dies erschwert es Personen, die vom Einsatz von KI-Systemen betroffen sein können, ihre Rechte geltend zu machen. Insbesondere ist es für betroffene Personen schwierig, Zugang zu relevanten Informationen zu erhalten und diese zu verstehen.[182]

Aus diesem Grund betont Artikel 14 der Konvention die Notwendigkeit sowohl zugänglicher als auch wirksamer Beschwerde. Diese Terminologie lehnt sich insbesondere an Artikel 2 UNO-Pakt II und Artikel 13 EMRK an. Um wirksam zu sein, muss die betreffende Beschwerde in der jeweiligen Situation unmittelbar Abhilfe schaffen können. Um zugänglich zu sein, muss die Beschwerde mit ausreichenden Verfahrensgarantien ausgestattet sein, damit sie für die betroffene Person von Nutzen ist.[183] Bei der Zugänglichkeit ist die Möglichkeit des praktischen Zugangs zu den notwendigen Informationen und zu sachkundiger Beratung zu berücksichtigen. Bezüglich der rechtlichen Zugänglichkeit muss die Beschwerde insbesondere hinreichende Erfolgsaussichten bieten.[184]

Es stellt sich die Frage nach der Anwendbarkeit von Artikel 14 im Verhältnis zwischen Privaten. Artikel 14 der Konvention ist auf Rechtsmittel *gegen Menschenrechtsverletzungen* anwendbar, die sich aus Tätigkeiten während des Lebenszyklus von KI-Systemen ergeben. Die Verpflichtungen aus Artikel 14 der KI-Konvention sollten dort Anwendung finden, wo eine direkte oder indirekte horizontale Wirkung der Grundrechte zwischen Privaten besteht oder in

---

[182] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 98 f.

[183] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 98.

[184] OLIVIER BIGLER, in: Luc Gonin/Olivier Bigler (Hrsg.), Commentaire de la Convention européenne des droits de l'homme (CEDH), Bern 2018, Art. 13 N 21 f., und die zitierten Referenzen.

Zukunft erkannt wird (vgl. die Überlegungen in Ziff. 4.2.3.1). In diesem beschränkten Rahmen bieten die zivilrechtlichen Rechtsmittel bereits gewisse Garantien (vgl. insbesondere die Erwägungen in Ziff. 6.3).

Artikel 14 Absatz 2 – Massnahmen zur Verstärkung der Tragweite von Artikel 14 Absatz 1

Die Konvention enthält bereits in Kapitel III, namentlich mit Artikel 8 (Transparenz und Aufsicht) und Artikel 9 (Rechenschaftspflicht und Verantwortung), einschlägige Verpflichtungen in Bezug auf die Rechtsmittel. Artikel 14 Absatz 2 geht darüber hinaus, indem er die Verpflichtung einführt, drei spezifische Massnahmen zur Verstärkung der Tragweite von Artikel 14 Absatz 1 zu ergreifen oder beizubehalten:

- Buchstabe a sieht eine Dokumentationspflicht vor. Diese beinhaltet einschlägige Informationen über KI-Systeme, die die Menschenrechte erheblich beeinträchtigen können, und über ihre entsprechende Nutzung. Diese Informationen sind den zum Zugang zu diesen Informationen befugten Stellen zu liefern sowie – sofern angemessen und anwendbar – den betroffenen Personen zur Verfügung zu stellen oder mitzuteilen.

  Diese Bestimmung soll Transparenz beim Einsatz von KI-Systemen gewährleisten. Letztlich besteht der Zweck dieses Artikels jedoch darin, betroffenen Personen die Möglichkeit zu geben, Entscheidungen, die mit Hilfe von KI-Systemen vorbereitet oder getroffen wurden, und gegebenenfalls sogar den Einsatz des Systems an sich anzufechten (vgl. Art. 14 Abs. 2 Bst. b und die Erläuterungen dazu weiter unten).

  Im Schweizer Recht ergeben sich die Dokumentations- und Informationspflichten gegenüber den betroffenen Personen aus dem Anspruch auf rechtliches Gehör gemäss Artikel 29 Absatz 2 BV. Zudem gewährt das BGÖ unter bestimmten Voraussetzungen Zugang zu amtlichen Dokumenten der Bundesverwaltung.

  Im Datenschutzrecht besteht zudem gemäss DSG die Pflicht zur Führung eines Verzeichnisses der Bearbeitungstätigkeiten (vgl. Art. 12 DSG). Im Zusammenhang mit der Information der betroffenen Personen sind auch die Artikel 19, 21 und 25 DSG über die Informationspflicht und das Auskunftsrecht zu erwähnen (vgl. Ziff. 4.3.2.5 und 4.3.2.6). Angesichts der weitergehenden Verpflichtungen nach Artikel 14 Absatz 2 Buchstabe a erscheinen die geltenden Bestimmungen jedoch nicht ausreichend.

Zusammenfassend kann festgestellt werden, dass das schweizerische Recht bereits Bestimmungen enthält, an die die Verpflichtung nach Artikel 14 Absatz 2 Buchstabe a der Konvention geknüpft werden kann. Der bestehende Rechtsrahmen scheint jedoch nicht ausreichend zu sein. Es bleibt jedoch darauf hinzuweisen, dass die Begriffe «einschlägig», «erheblich» und «sofern anwendbar» den Vertragsparteien einen Ermessensspielraum bei der Umsetzung einräumen. Insbesondere fallen nur Systeme, die «erhebliche» Auswirkungen auf die Menschenrechte haben können, unter diese Bestimmung, die somit keine Anforderung für alle in Artikel 3 der Konvention definierten KI-Systeme darstellt. Es wird Aufgabe des Gesetzgebers sein, die entsprechende Schwelle festzulegen.

Darüber hinaus könnte die Pflicht, relevante Informationen zur Verfügung zu stellen oder zugänglich zu machen, durch andere überwiegende Interessen eingeschränkt werden, beispielsweise im Zusammenhang mit dem Schutz von Daten Dritter oder von Geschäftsgeheimnissen.

Artikel 14 Absatz 2 Buchstabe a erfordert daher einer Präzisierung des anwendbaren Rechtsrahmens, so dass ein Tätigwerden des Gesetzgebers in diesem Sinne angezeigt erscheint. Zudem werden die Informationen in der Regel nicht bei den Behörden, sondern bei den Entwicklern der KI-Systeme vorhanden sein.

Das Führen von Registern relevanter Informationen zu KI-Systemen ist in der Bestimmung nicht vorgesehen, könnte aber eine mögliche Form der Umsetzung dieser Verpflichtung sein.

- Buchstabe b schreibt vor, dass die in Artikel 14 Absatz 2 Buchstabe a genannten Informationen ausreichen müssen, damit die betroffene Person die Entscheidungen, die aufgrund der Verwendung des Systems getroffen wurden oder massgeblich auf der Verwendung des Systems beruhen, oder, falls erforderlich und angemessen, den Einsatz des Systems anfechten kann. Der letzte Fall kann Situationen abdecken, in denen die Verwendung eines bestimmten Systems hypothetisch verboten wäre. Denkbar sind auch Situationen, in denen der Behörde vorgeworfen wird, ein KI-System ohne angemessene Rechtsgrundlage eingesetzt zu haben, und in denen dann dieser Einsatz des KI-Systems als solcher angefochten wird.

  KI-Systeme stellen besondere Herausforderungen bezüglich Begründungspflicht.[185] Diese sind je nach Art des eingesetzten KI-Systems unterschiedlich. Wie oben erläutert (vgl. *ad* Art. 8, Ziff. 4.3.2.3), sind regelbasierte Algorithmen hinsichtlich Erklärbarkeit und Interpretierbarkeit ihrer Ergebnisse weniger problematisch als Algorithmen des maschinellen Lernens («machine learning»). Bei Letzteren erweist sich die Erfüllung der Begründungspflicht als besonders schwierig.

  In jedem Fall sollte sich die Begründung nicht auf die allgemeine Funktionsweise eines KI-Systems beziehen, sondern auf die *Logik, auf der die Entscheidung im Einzel-*

[185] Vgl. zu den Herausforderungen im öffentlichen Sektor NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (Fn. 39), 37; NADJA BRAUN BINDER/LILIANE OBRECHT, White Paper: Transparenz durch Begründung von Verfügungen, Juni 2024, 3, abrufbar unter: https://www.itsl.uzh.ch > Wissenstransfer > Publikationen > Positionspapiere > Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von künstlicher Intelligenz (abgerufen am 27. August 2024); NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 10 N 23; NADJA BRAUN BINDER/LILIANE OBRECHT, Die Begründung von Verfügungen (Fn. 84), 707 ff.

fall beruht. Inhaltlich müssen die übermittelten Informationen dem Kontext entsprechen, hinreichend klar und vor allem für die betroffene Person nützlich sein, damit sie sich in einem Rechtsmittel wirksam darauf berufen kann.[186]

Im schweizerischen Recht ergibt sich das Recht auf Begründung von Verwaltungs- und Gerichtsentscheiden aus dem Anspruch auf rechtliches Gehör nach Artikel 29 Absatz 2 BV. Die Begründung eines Entscheides muss sich sowohl auf den zugrunde gelegten Sachverhalt als auch auf die rechtlichen Erwägungen stützen. Das Bundesgericht betont, dass es entscheidend ist, dass aus dem Entscheid klar hervorgeht, welche Tatsachen festgestellt und welche rechtlichen Folgerungen aus den massgebenden Tatsachen gezogen werden.[187] Der Anspruch auf rechtliches Gehör wird durch das anwendbare Verfahrensrecht konkretisiert. Für das Verwaltungsverfahren auf Bundesebene sind insbesondere die Bestimmungen des VwVG, namentlich Artikel 35 VwVG, massgebend. Die Begründung muss angemessen sein und variiert je nach konkretem Fall. Je komplexer der Sachverhalt, desto dichter muss die Begründung sein. Hat die Behörde einen grossen Ermessensspielraum, ist der Entscheid ebenfalls ausführlicher zu begründen. In Bereichen, in denen viele gleichartige Entscheidungen getroffen werden, darf die Begründung aus Gründen der Verfahrensökonomie weniger detailliert ausfallen. Dabei ist zu beachten, dass eine ausreichende Begründung staatlichen Handelns auch der Rechtssicherheit dient, indem sie staatliches Handeln vorhersehbar macht und eine einheitliche Praxis fördert. Setzt die Behörde KI-Systeme ein, um behördliche oder gerichtliche Entscheidungen zu treffen, sind daher die Garantien zu beachten, die sich aus dem Anspruch auf rechtliches Gehör ergeben.

In diesem Zusammenhang ist auch das DSG zu erwähnen, das entsprechende Bestimmungen für automatisierte Einzelentscheidungen kennt (vgl. Art. 21 und 25 Abs. 2 Bst. f DSG). Eine automatisierte Einzelentscheidung liegt vor, wenn eine Datenbearbeitung automatisch, ohne Dazutun einer natürlichen Person erfolgt, und für die betroffene Person eine rechtliche Folge nach sich zieht oder sie erheblich beeinträchtigt (automatisierte Einzelentscheidung). Es kann sich dabei um eine Entscheidung einer privaten Stelle oder eines öffentlichen (Bundes-)Organs handeln, die Entscheidung muss jedoch einen gewissen Komplexitätsgrad aufweisen. Einfache Entscheidungen, wie sie etwa bei einem Geldbezug am Bancomaten getroffen werden, fallen nicht darunter.[188] Im öffentlichen Recht entfaltet eine Verfügung im Sinne von Artikel 5 VwVG, die ausschliesslich auf einer automatisierten Bearbeitung von Personendaten beruht, Rechtswirkungen gegenüber der betroffenen Person und ist daher als automatisierte

---

[186] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 99.

[187] BGE **142** II 154, E. 4.2; JACQUES DUBEY, Droits fondamentaux, Volume II: Liberté, garanties de l'Etat de droit, droits sociaux et politiques, Basel 2018, 815.

[188] Botschaft vom 15. September 2017 zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz, BBl **2017** 6941 7056 s.

Einzelentscheidung im Sinne von Art. 21 DSG zu qualifizieren (vgl. z. B. Art. 38 Abs. 2 ZG).[189]

Im Rahmen des Auskunftsrechts sieht Artikel 25 Absatz 2 Buchstabe f DSG vor, dass die betroffene Person zumindest über das Vorliegen einer automatisierten Einzelentscheidung informiert werden muss, falls eine solche vorliegt, sowie über die Logik, auf der diese Entscheidung beruht. Gemäss Botschaft des Bundesrates sind nicht die verwendeten Algorithmen, die häufig unter das Geschäftsgeheimnis fallen, sondern die Grundannahmen der Algorithmus-Logik, auf der die automatisierte Einzelentscheidung beruht, offenzulegen.[190]

Zu beachten ist jedoch, dass diese Bestimmungen nur anwendbar sind, wenn die Entscheidung auf einer *ausschliesslich* automatisierten Personendatenbearbeitung beruht (vgl. Art. 21 Abs. 1 DSG). Artikel 14 Absatz 2 Buchstabe b KI-Konvention ist demgegenüber weiter gefasst, da er auch Fälle erfasst, in denen die Entscheidung «erheblich» auf dem Einsatz eines KI-Systems beruht. Erfasst werden also auch Situationen, in denen KI-Systeme bei einer Entscheidung als Unterstützung eingesetzt werden, wobei diese Entscheidung dann aber letztlich von einer natürlichen Person getroffen wird. Im Gegensatz zu Artikel 21 DSG beschränkt sich Artikel 14 Absatz 2 Buchstabe b zudem nicht auf Fälle von automatisierten Entscheidungen, bei denen personenbezogene Daten verarbeitet werden. Denn auch wenn bei den meisten KI-Systemen personenbezogene Daten bearbeitet werden, ist dies doch nicht immer der Fall.[191]

Im Übrigen ist daran zu erinnern, dass die Bestimmungen des DSG auf die Bearbeitung von Personendaten im Rahmen von Gerichtsverfahren oder bundesrechtlich geregelte Verfahren nicht anwendbar sind. Auf erstinstanzliche Verwaltungsverfahren hingegen ist das DSG anwendbar (Art. 1 Abs. 3 DSG).

Zusammenfassend lässt sich sagen, dass das schweizerische Recht bereits verschiedene einschlägige Bestimmungen zur Begründungspflicht im Zusammenhang mit automatisierten Einzelentscheidungen enthält, insbesondere im Rahmen des DSG. Diese decken jedoch nicht alle Konstellationen ab, d. h. insbesondere nicht Entscheidungen, die *mithilfe* von KI-Systemen getroffen werden. Ein Tätigwerden des Gesetzgebers erscheint notwendig, um auch diesen Fall zu erfassen und die Anforderungen an die Begründungspflicht in diesem Zusammenhang zu definieren.

Bei erstinstanzlichen Verwaltungsentscheidungen ist selbstverständlich der Anspruch des rechtlichen Gehörs zu beachten. Es sollte jedoch weiter geprüft werden,

[189] LISA JACCOUD/SÉBASTIEN FANTI/ALEXANDRE STAGER, in: Petit commentaire LPD (Fn. 44), Art. 21 N 22.

[190] Botschaft Totalrevision DSG (Fn. 188), BBl **2017** 6941 7067.

[191] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI u. a., Einsatz Künstlicher Intelligenz (Fn. 39), 37, Fussnote 246.

ob der Rechtsrahmen in Bezug auf die Art und Weise, wie die Einhaltung dieser Anforderungen im Rahmen automatisierter Verwaltungsverfahren gewährleistet wird, einer Präzisierung bedarf.

Für gerichtliche Entscheidungen scheint kein Handlungsbedarf zu bestehen, da es nach dem derzeitigen Stand der Verfahrensvorschriften keine Rechtsgrundlage für solche Entscheidungen gibt. Sollten solche Rechtsgrundlagen geschaffen werden, wäre ein allfälliger Anpassungsbedarf vertieft zu prüfen.

- Artikel 14 Absatz 2 Buchstabe c sieht vor, dass den Betroffenen eine wirksame Beschwerde bei den zuständigen Behörden zur Verfügung stehen muss. Gemäss dem erläuternden Bericht kann dies die in Artikel 26 der Konvention vorgesehenen Kontrollmechanismen einschliessen (vgl. Ziff. 4.4.5).[192]

Gemäss Analyse des BJ verlangt die Bestimmung keine Schaffung neuer Rechtsmittel. Vielmehr verlangt sie von den Vertragsparteien die Gewährleistung der Wirksamkeit der bestehenden Rechtsmittel bei Menschenrechtsverletzungen im Zusammenhang mit Tätigkeiten im Rahmen des Lebenszyklus von KI-Systemen. Diesem Zweck dienen die in Artikel 14 Absatz 2 Buchstaben a und b der Konvention vorgesehenen Massnahmen. Hinsichtlich eines allfälligen Handlungsbedarfs im schweizerischen Recht kann daher auf die obigen Ausführungen verwiesen werden.

4.3.3.3 Artikel 15 – Verfahrensgarantien

Artikel 15 Absatz 1 – Wirksame Verfahrensgarantien, Schutzmassnahmen und Verfahrensrechte

Nach Artikel 15 Absatz 1 stellt jede Vertragspartei sicher, dass in Fällen, in denen sich ein KI-System erheblich auf die Wahrnehmung der Menschenrechte auswirkt, den betroffenen Personen im Einklang mit dem geltenden Völker- und innerstaatlichen Recht wirksame Verfahrensgarantien, Schutzmassnahmen und Verfahrensrechte zur Verfügung stehen.

Die Bestimmung beschränkt sich auf KI-Systeme, die «erhebliche» Auswirkungen auf die Ausübung der Menschenrechte haben, und gilt nicht für alle KI-Systeme im Sinne von Artikel 3 der Konvention.

Hinsichtlich der Anwendung der Bestimmung auf den Privatsektor wird auf die obigen Ausführungen zu Artikel 14 verwiesen (vgl. Ziff. 4.3.3.2). Das Erfordernis einer «erheblichen» Auswirkung wird die erfassten privatrechtlichen Sachverhalte weiter einschränken.

[192] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 100.

Diese Bestimmung betrifft die Verfahrensgarantien im Allgemeinen. Sie soll sicherstellen, dass die üblichen Verfahrensgarantien unabhängig vom Einsatz von KI gleich wirksam sind.

Im schweizerischen Recht ist dies insbesondere das Recht auf gleiche und gerechte Behandlung im Verfahren, das durch die Artikel 29–32 BV sowie durch Artikel 6 EMRK gewährleistet wird. Hervorzuheben ist insbesondere Folgendes:

- Eine Person hat das aus dem *Anspruch auf rechtliches Gehör* abgeleitete Recht, der Behörde *ihren Standpunkt darzulegen*, bevor diese eine sie betreffende Entscheidung fällt. Sie hat das Recht, sich an der Untersuchung zum Fall zu beteiligen, Tatsachen vorzubringen, Beweise anzubieten und sich zu den vorgebrachten Beweisen zu äussern.

  Das schweizerische Recht sieht in Artikel 21 Absätze 1 und 2 DSG eine Informationspflicht bei automatisierten Einzelentscheidungen sowie die Möglichkeit vor, dass die betroffene Person auf Wunsch ihren Standpunkt darlegen kann. Artikel 21 Absätze 3 und 4 DSG nennen mögliche Einschränkungen. So gelten die Rechte nach Artikel 21 Absatz 2 DSG für Bundesorgane nicht, wenn die betroffene Person nach Artikel 30 Absatz 2 VwVG nicht angehört werden muss.

  Wie bereits erwähnt, deckt der Anwendungsbereich der Bestimmungen des DSG jedoch nicht alle in Artikel 15 Absatz 1 der Konvention genannten Fälle ab, insbesondere nicht den Fall von teilautomatisierten Entscheidungen. Darüber hinaus bestimmt Artikel 1 Absatz 3 DSG, dass die Bearbeitung von Personendaten im Rahmen von Gerichtsverfahren oder von Verfahren, die durch Bundesrecht geregelt sind, sowie die Rechte der betroffenen Personen unter Ausschluss des DSG durch das anwendbare Verfahrensrecht geregelt werden. Diese Gesetze sehen jedoch ihrerseits Garantien des rechtlichen Gehörs vor.

- Eine wichtige Rolle spielt auch das Recht auf *Akteneinsicht*. Parteirechte laufen ins Leere, wenn die Verfahrensbeteiligten keine Kenntnis von den Unterlagen haben, die der Entscheidung der Behörde zugrunde liegen.

  Bei automatisierten Verfahren sollte die betroffene Person Zugang zu den spezifischen Informationen eines Falles haben, die von einem KI-System verarbeitet werden (*Input*), sowie zu dem daraus resultierenden Ergebnis (*Output*). Darüber hinaus sollte sie wissen können, wie diese Informationen maschinell verarbeitet werden. Unter Umständen sind diese Inhalte für die betroffene Person jedoch nur schwer verständlich. Trotzdem hat das Bundesgericht, wenn auch in einem anderen Zusammenhang, darauf hingewiesen, dass potenzielle faktische Schwierigkeiten, die sich aus der Menge

der einzusehenden Dokumente ergeben, dem Recht auf Akteneinsicht keinen Abbruch tun.[193] Bei KI-Systemen könnte jedoch eine Grenze erreicht sein, da die betroffene Person möglicherweise nicht in der Lage ist, aus den zur Verfügung gestellten Informationen Schlussfolgerungen zu ziehen.[194]

- Im Rahmen des Verwaltungsverfahren sind insbesondere die *Pflicht der Behörden zur Ermittlung des Sachverhalts von Amtes wegen* (Untersuchungsgrundsatz, Art. 12 VwVG) und die *Mitwirkungspflicht der Parteien* (Art. 13 VwVG) zu beachten.

  Bei voll- und teilautomatisierten Verfahren ist insbesondere sicherzustellen, dass alle relevanten Tatsachen berücksichtigt werden konnten.[195] Dies setzt einerseits die Richtigkeit und Vollständigkeit der Daten voraus, andererseits dass alle Daten, deren Relevanz für den spezifischen Kontext der Entscheidungsfindung zweifelhaft erscheint, unberücksichtigt bleiben.[196]

  Sachdaten können auch auf andere Weise beschafft werden, z. B. indem diese von der oder dem Beschwerdeführenden selber eingefordert werden. Die für die automatisierte Bearbeitung notwendige standardisierte Datenerfassung kann jedoch dazu führen, dass die oder der Beschwerdeführende nicht alle Angaben machen kann, die aus ihrer bzw. seiner Sicht für das Verfahren erforderlich wären. Wenn in einem Formular in bestimmten Feldern nur numerische Werte zulässig sind, können zusätzliche Erläuterungen, die für eine korrekte Beurteilung des Falles erforderlich wären, unter Umständen nicht gegeben werden.[197]

---

[193] BGE **144** II 427, 436, E. 3.2.3.

[194] In diesem Sinn: REGINA WEDER, Verfahrensgrundrechtliche Anforderungen an automatisierte Verwaltungsverfahren, in: Monika Simmler (Hrsg.), Smart Criminal Justice. Der Einsatz von Algorithmen in der Polizeiarbeit und Strafrechtspflege, Basel 2021, 237 ff., 251 f.

[195] NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 10 N 25 f.; vgl. auch NADJA BRAUN BINDER, Der Untersuchungsgrundsatz als Herausforderung vollautomatisierter Verfahren, zsis 2/2020, N 28.

[196] Europarat, L'administration et vous – Un manuel. Principes de droit administratif concernant les relations entre l'Administration et les personnes, Strassburg 2024, 18, abrufbar unter: https://www.coe.int/fr/web/cdcj/-/publication-of-the-new-handbook-the-administration-and-you-that-takes-into-account-the-increasing-use-of-ai (abgerufen am 26. August 2024).

[197] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (Fn. 39), 38; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (Fn. 28), 10 N 25 f.; NADJA BRAUN BINDER, Der Untersuchungsgrundsatz (Fn. 185), N 28.

- Die *Rechtsweggarantie* (Art. 29*a* BV) garantiert jeder Person das Recht auf Beurteilung durch eine richterliche Behörde. Daraus folgt, dass ein vollautomatisierter Entscheid im Einzelfall nicht von einer einzigen Instanz gefällt werden darf.[198]

Aus den obigen Ausführungen ergibt sich, dass die bestehenden Verfahrensgarantien im Zusammenhang mit KI-Systemen bereits vollumfänglich anwendbar sind. Sie werden in den Verwaltungs-, Zivil- und Strafprozessordnungen konkretisiert. Zudem bietet Artikel 21 DSG im erstinstanzlichen Verwaltungsverfahren bereits einen gewissen Schutz bei vollautomatisierten Einzelentscheidungen.

Abgesehen von den obigen Überlegungen zur Begründungspflicht (vgl. Ziff. 4.3.3.2) erscheinen die geltenden Regelungen ausreichend. Die Rechtsprechung wird die Konturen dieser Garantien im Zusammenhang mit KI-Systemen schärfen können. Bei Problemen oder Lücken können diejenigen Fälle geprüft werden, in denen der Einsatz von KI die Betroffenen daran hindern würde, ihre Rechte nach geltendem Recht gerichtlich geltend zu machen.

In jedem Fall sind mögliche entgegenstehende Interessen zu berücksichtigen, die den Anwendungsbereich dieser Garantien einschränken können, insbesondere das private Interesse des Entwicklers des Algorithmus, der sich auf das Geschäftsgeheimnis berufen könnte, oder ein anderes öffentliches Interesse, wie das Interesse an der Strafverfolgung. Im Verwaltungsverfahren ist es beispielsweise gemäss Artikel 27 VwVG bereits heute möglich, private Interessen, die dem Recht auf Akteneinsicht entgegenstehen, bei der Behandlung entsprechender Gesuche zu berücksichtigen.

Artikel 15 Absatz 2 – Information über Interaktion mit einem KI-System

Gemäss Artikel 15 Absatz 2 bemüht sich jede Vertragspartei sicherzustellen, dass Personen, die mit KI-Systemen interagieren, in einer dem Kontext angemessenen Weise darüber informiert werden, dass sie mit solchen Systemen und nicht mit einem Menschen interagieren. Damit soll insbesondere die Gefahr der Manipulation und Täuschung ausgeschlossen werden.[199]

Im schweizerischen Recht besteht für automatisierte Einzelentscheidungen eine entsprechende Verpflichtung im DSG. So sieht Artikel 21 Absatz 1 DSG vor, dass der Verantwortliche die betroffene Person über jede automatisierte Einzelentscheidung informieren muss. Ergeht die automatisierte Einzelentscheidung durch ein Bundesorgan, muss es die Entscheidung entsprechend kennzeichnen (Art. 21 Abs. 4 DSG). Artikel 25 Absatz 2 Buchstabe f DSG

[198] MICHAEL MONTAVON, Cyberadministration et protection des données – Etude théorique et pratique de la transition numérique en Suisse du point de vue de l'Etat, des citoyen-ne-s et des autorités de contrôle, Freiburg 2021, 667.

[199] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 104.

sieht im Rahmen des Auskunftsrechts vor, dass die betroffene Person über das Vorliegen einer automatisierten Einzelentscheidung informiert wird.

Wie bereits erwähnt deckt Artikel 21 DSG nicht alle Fälle ab, weil er nur in Fällen von automatisierten Einzelentscheidungen zur Anwendung gelangt. Artikel 15 Absatz 2 der Konvention würde hingegen eine Informationspflicht auch im Fall einer Interaktion mit KI-gestützten Chatbots auf den Websites der öffentlichen Verwaltung verlangen.[200]

Zusammenfassend kann festgestellt werden, dass die Umsetzung von Artikel 15 Absatz 2 in innerstaatliches Recht ein Tätigwerden des Gesetzgebers erfordert, um das im DSG bereits vorgesehene Auskunftsrecht zu erweitern. Eine solche Information ermöglicht es der betroffenen Person, ihre Verfahrensrechte wahrzunehmen.

Die Bestimmung lässt dem Gesetzgeber jedoch einen Spielraum, da die Formulierung «in einer dem Kontext angemessenen Weise» (französisch «en fonction du contexte») auch Situationen erfassen kann, in denen es offensichtlich ist, dass die Interaktion mit einer Maschine erfolgt und daher keine Information erforderlich ist. Nicht erfasst sind jedoch z. B. Situationen, in denen der eigentliche Zweck der Nutzung des KI-Systems durch die Information vereitelt würde, beispielsweise im Bereich der Strafverfolgung.

4.3.4 Kapitel V: Bewertung und Minderung von Risiken und nachteiligen Auswirkungen

Dieses Kapitel besteht aus einem einzigen Artikel (Art. 16) über den Rahmen für das Risiko- und Folgenmanagement.

Artikel 16 Absatz 1 sieht vor, dass jede Vertragspartei unter Berücksichtigung der Grundsätze in Kapitel III Massnahmen trifft oder aufrechterhält, um die von KI-Systemen ausgehenden Risiken zu ermitteln, zu bewerten, zu vermeiden und zu mindern, wobei die tatsächlichen und potenziellen Auswirkungen auf die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit zu berücksichtigen sind.

Diese Bestimmung verpflichtet zur Schaffung eines Rahmens für das Risiko- und Folgenmanagement. Damit sollen *ex ante*, d. h. vor dem Einsatz eines KI-Systems, und gegebenenfalls iterativ, während des gesamten Lebenszyklus des KI-Systems, die relevanten Risiken des KI-Systems für Menschenrechte, Demokratie und Rechtsstaatlichkeit anhand einer Methodik mit konkreten Kriterien bewertet werden. Dies ist ein wichtiges Instrument, um die Einhaltung der Anforderungen der Konvention sicherzustellen, insbesondere der in Kapitel III niedergelegten Grundsätze.

Artikel 16 stellt sicher, dass die Vertragsparteien bei der Bestimmung, der Analyse und der Bewertung der Risiken und Folgen von KI-Systemen einen gemeinsamen Ansatz verfolgen. Gleichzeitig geht er davon aus, dass die Vertragsparteien besser in der Lage sind, die ent-

[200] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 104.

sprechenden Regulierungsentscheidungen zu treffen, und lässt ihnen daher einen Ermessensspielraum bei der Umsetzung. Aus diesem Grund sieht Artikel 16 Absatz 2 vor, dass die zu treffenden Massnahmen gegebenenfalls abgestuft und differenziert sein müssen und den Kontext des geplanten Einsatzes von KI-Systemen sowie die Schwere und Wahrscheinlichkeit möglicher Folgen angemessen berücksichtigen müssen (vgl. Art. 16 Abs. 2 Bst. a und b).

Artikel 16 Absatz 2 enthält weitere Details zu den zu ergreifenden oder beizubehaltenden Massnahmen. Buchstabe c schreibt vor, dass im Rahmen des Risiko- und Folgenmanagements die Standpunkte der relevanten beteiligten Parteien, insbesondere der Personen, deren Rechte betroffen sein könnten, zu berücksichtigen sind. Andere Beteiligte können beispielsweise externe Fachexperten oder Vertreter der Zivilgesellschaft sein.[201]

Buchstabe d sieht den iterativen Charakter der Massnahmen in dem Sinne vor, dass sie während des gesamten Lebenszyklus der KI-Systeme wiederholt werden müssen. Buchstabe e legt fest, dass die Massnahmen auch eine Überwachung der Risiken und negativen Auswirkungen auf Menschenrechte, Demokratie und Rechtsstaatlichkeit umfassen müssen. Diese Merkmale ermöglichen es, die Auswirkungen von KI-Systemen auch *ex post*, d. h. nach deren Einsatz, zu identifizieren und zu bewerten. Buchstabe f enthält eine Verpflichtung zur Dokumentation der Risiken, der tatsächlichen und potenziellen Auswirkungen und des Risikomanagementansatzes. Buchstabe g stellt klar, dass der Rahmen für das Risikomanagement gegebenenfalls eine Vorabprüfung von KI-Systemen vor ihrer Bereitstellung zur erstmaligen Verwendung sowie bei wesentlichen Änderungen vorsehen muss.

Artikel 16 Absatz 3 verpflichtet die Vertragsparteien, Massnahmen zum angemessenen Umgang mit nachgewiesenen nachteiligen Auswirkungen von KI-Systemen zu ergreifen. Dadurch sollen solche Auswirkungen mithilfe geeigneter Massnahmen behoben werden können. Diese Phase ist zu dokumentieren und in den Rahmen für das Risiko- und Folgenmanagement gemäss Artikel 16 Absatz 2 zu integrieren.

Schliesslich sieht Artikel 16 Absatz 4 vor, dass eine Vertragspartei, die den Einsatz eines KI-Systems als mit der Achtung der Menschenrechte, der Demokratie und der Rechtsstaatlichkeit unvereinbar hält, prüft, ob ein Moratorium, ein Verbot oder andere geeignete Massnahmen erforderlich sind. Angesichts ihres einschneidenden Charakters sollten Massnahmen wie ein Moratorium oder ein Verbot nur dann erwogen werden, wenn eine Vertragspartei der Auffassung ist, dass ein bestimmter Einsatz von KI-Systemen ein unannehmbares Risiko darstellt und eine sorgfältige Prüfung ergibt, dass keine anderen Massnahmen zur Minderung dieses Risikos zur Verfügung stehen. Diese Massnahmen sollten auch geeignete Überprüfungsverfahren umfassen, damit sie aufgehoben werden können, sobald die betreffenden Risiken hinreichend gemindert wurden oder geeignete Massnahmen der Risikominderung verfügbar geworden sind.

Der Europarat wird eine nicht bindende Methodik entwickeln, um einen möglichen Weg für die Umsetzung der Verpflichtungen aus Artikel 16 aufzuzeigen. So soll den Vertragsparteien

[201] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 108.

ein Modell zur Verfügung gestellt werden, das sie bei der Entwicklung ihres eigenen rechtlichen Rahmens für das Risiko- und Folgenmanagement unterstützt.

Der Rahmen für das Risiko- und Folgenmanagement von KI-Systemen nach Artikel 16 der KI-Konvention mag der Datenschutz-Folgenabschätzung (DSFA) nach Artikel 22 DSG ähnlich erscheinen, es handelt sich jedoch um zwei unterschiedliche Instrumente.

Gemäss Artikel 22 Absatz 1 DSG hat der für die Bearbeitung Verantwortliche vorgängig eine DSFA durchzuführen, wenn die geplante Bearbeitung voraussichtlich ein hohes Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person mit sich bringt. Gemäss der Botschaft zum DSG gilt die Pflicht zur Durchführung einer DSFA unter bestimmten Voraussetzungen sowohl für private Datenbearbeiter als auch für Bundesorgane, weshalb in der Bestimmung nicht nur von einem hohen Risiko für die Persönlichkeit der betroffenen Person, sondern auch für deren Grundrechte die Rede ist.[202]

Die DSFA ist ein weniger umfassendes Instrument als der Rahmen für das Risiko- und Folgenmanagement in Artikel 16 der KI-Konvention. Zusätzlich unterscheidet sie sich auch im Anwendungsbereich: Obwohl die meisten KI-Systeme auch Personendaten bearbeiten, ist dies nicht immer der Fall. Darüber hinaus sind die Risiken im Rahmen des Risiko- und Folgenmanagement weiter gefasst als in der DSFA.

Für die Datenbearbeitung durch Bundesorgane sieht Artikel 22 Absatz 1 DSG die Durchführung einer DSFA vor, wenn die geplante Bearbeitung voraussichtlich *ein hohes Risiko für die Grundrechte* mit sich bringt. Die vom DSG erfassten Grundrechte sind in erster Linie das Recht auf persönliche Freiheit (Art. 10 Abs. 2 BV), das Recht auf Schutz der Privatsphäre (Art. 13 Abs. 1 BV) und das Recht auf informationelle Selbstbestimmung im Sinne von Artikel 13 Absatz 2 BV, das auch in Artikel 8 EMRK verankert ist.[203] In der Praxis wird bei einer DSFA nach Artikel 22 DSG in der Regel abgeschätzt, welche negativen Folgen die Bearbeitung von Personendaten mit einer gewissen Wahrscheinlichkeit für die betroffene Person haben könnte, insbesondere physische Folgen (z. B. können falsche Daten zu einer falschen medizinischen Behandlung führen), materielle Folgen (z. B. die Verweigerung einer Arbeitsstelle, die betrügerische Verwendung einer Kreditkarte) oder immaterielle Folgen (z. B. kann das Wissen, beobachtet zu werden, bei einer Person zu einer Verhaltensänderung führen).[204] Artikel 16 der KI-Konvention zielt darauf ab, die Risiken und Auswirkungen von KI-Systemen

---

[202] Botschaft Totalrevision DSG (Fn. 188), BBl **2017** 6941 7059.

[203] JULIEN FRANÇAIS, in: Petit commentaire LPD (Fn. 44), Art. 1 N 11; CR LPD-BERTIL COTTIER, Art. 1 N 19; BSK DSG/BGÖ-MATTHIAS R. SCHÖNBÄCHLER/URS MAURER-LAMBROU/SIMON KUNZ, Art. 1 N 18; MARCO FREY, in: Bruno Baeriswy/Kurt Pärli/Dominika Blonski (Hrsg.), Datenschutzgesetz, Stämpfli Handkommentar, Art. 1 N 29; DAVID ROSENTHAL, in: David Rosenthal/Yvonne Jöhri (Hrsg.), Handkommentar zum Datenschutzrecht, Art. 1 N 3 f.

[204] In diesem Sinn: DAVID ROSENTHAL/SAMIRA STUDER/ALEXANDRE LOMBARD (für die Übersetzung), La nouvelle loi sur la protection des données (Fn. 129), 58, N 149; CR LPD-PHILIPPE GILLIÉRON, Art. 22 N 24.

auf alle Menschenrechte sowie auf Demokratie und Rechtsstaatlichkeit zu bewerten, unabhängig von der Bearbeitung personenbezogener Daten.

Für den Privatsektor hätte die Umsetzung von Artikel 16 zur Folge, dass die betroffenen privaten Akteure zumindest in jenen Bereichen eine Grundrechte-Folgenabschätzung vornehmen müssten, in denen eine direkte oder indirekte horizontale Wirkung der Grundrechte zwischen Privaten besteht oder in Zukunft erkannt wird (vgl. die Überlegungen in Ziff. 4.2.3.1). Dies bedeutet beispielsweise, dass die Datenschutzgesetzgebung, die Normen zum Persönlichkeitsschutz, wie sie in ZGB und OR verankert sind, als auch das GIG oder das BehiG usw. zu beachten sind. Die KI-Konvention führt aber nicht zu einer Erweiterung des Schutzbereichs der Grundrechte.

Nach der hier vorgenommenen Auslegung lässt Artikel 16 den Vertragsstaaten einen gewissen Gestaltungsspielraum, einschliesslich der Möglichkeit, im Privatsektor in bestimmten Fällen weniger weit zu gehen. So wäre es beispielsweise denkbar, die Pflicht zur Durchführung einer Grundrechte-Folgenabschätzung auf gewisse Anbieter grundlegender Dienste zu beschränken, wie dies in der KI-Verordnung der EU vorgesehen ist (vgl. Art. 27 der Verordnung, Ziff. 5.2.7.3.3).

Aus den obigen Ausführungen ergibt sich, dass das schweizerische Recht eine Pflicht zur Durchführung einer Folgenabschätzung bei der Bearbeitung von Personendaten (DFAS) kennt. Diese Pflicht reicht jedoch nicht aus, um den in Artikel 16 der KI-Konvention beschriebenen Rahmen des Risiko- und Folgenmanagements zu erfüllen. Der Gesetzgeber müsste deshalb handeln. Dabei sind jedoch die Besonderheiten des schweizerischen Systems zu berücksichtigen, insbesondere in Bezug auf die Wirkung der Grundrechte zwischen Privaten.

Im schweizerischen Recht regelt Artikel 36 BV die Voraussetzungen für die Einschränkung von Grundrechten. Ein Eingriff durch ein KI-System ohne gesetzliche Grundlage oder mit Auswirkungen auf den Kerngehalt der Grundrechte wäre rechtswidrig. Für den öffentlichen Sektor ergibt sich daher aus Artikel 16 Absatz 4 der Konvention kein Handlungsbedarf. Ein Verbot oder ein Moratorium für bestimmte Anwendungen von KI-Systemen könnte im privaten Sektor sinnvoll sein; oder auch im öffentlichen Sektor im Rahmen einer politischen Entscheidung zum aktiven Eingreifen zwecks Verbots bestimmter Praktiken. Diese Frage könnte sich zum Beispiel bei den Technologien zur Emotionserkennung stellen.

Bei der Einführung eines Rahmens für das Risiko- und Folgenmanagement von KI-Systemen in das schweizerische Recht sollte zudem die bereits bestehende Datenschutzfolgenabschätzung (DSFA) berücksichtigt werden.

In der KI-Verordnung ist zu beachten, dass, sofern Anbieter von KI-Systemen bereits eine Datenschutzfolgenabschätzung durchführen müssen, diese zusammen mit der Folgenabschätzung gemäss KI-Verordnung durchgeführt wird (vgl. Ziff. 5.2.7.3.3). Bei einer Ratifizierung der KI-Konvention müsste somit auch im Schweizer Recht ein Koordinationsmechanismus geschaffen werden.

4.3.5 Kapitel VI: Umsetzung des Übereinkommens

4.3.5.1 Allgemeines

Kapitel VI enthält Verpflichtungen der Vertragsparteien, die sich ausschliesslich auf die Phase der Umsetzung der Konvention in innerstaatliches Recht beziehen.

4.3.5.2 Artikel 17 – Nichtdiskriminierung

Diese Bestimmung legt fest, dass die Umsetzung der KI-Konvention durch die Vertragsparteien in Übereinstimmung mit ihren völkerrechtlichen Verpflichtungen auf dem Gebiet der Menschenrechte ohne Diskriminierung zu erfolgen hat.

Artikel 17 verbietet somit jegliche Diskriminierung bei der *Umsetzung* der Konvention durch die Vertragsparteien. Der Begriff der Diskriminierung ist identisch mit dem des geltenden Völkerrechts und umfasst ein breites Spektrum von Diskriminierungsgründen im Zusammenhang mit den persönlichen Merkmalen einer Person, ihrer Situation oder ihrer Zugehörigkeit zu einer Gruppe.[205]

In der schweizerischen Rechtsordnung verbietet Artikel 8 Absatz 2 BV bereits jegliche Diskriminierung. Insbesondere haben die rechtsanwendenden Behörden die gesetzlichen Bestimmungen im Lichte der Nichtdiskriminierung auszulegen.

4.3.5.3 Artikel 18 – Rechte von Menschen mit Behinderungen und von Kindern

Diese Bestimmung verpflichtet die Vertragsparteien, im Einklang mit ihrem innerstaatlichen Recht und den geltenden völkerrechtlichen Verpflichtungen die besondere Schutzbedürftigkeit im Zusammenhang mit der Achtung der Rechte von Menschen mit Behinderungen und der Rechte von Kindern gebührend zu berücksichtigen. Sie verweist daher auf die Bestimmungen und den Rechtsrahmen des Übereinkommens über die Rechte von Menschen mit Behinderungen[206] und des Übereinkommens über die Rechte des Kindes[207] sowie auf das geltende innerstaatliche Recht jeder Vertragspartei in Bezug auf die Rechte von Menschen mit Behinderungen und die Rechte von Kindern.

Ziel ist eine grösstmögliche Berücksichtigung der besonderen Bedürfnisse und der Schutzbedürftigkeit von Menschen mit Behinderungen und von Kindern im Hinblick auf die Achtung ihrer Rechte, einschliesslich der Vermittlung digitaler Kompetenz.[208]

---

[205] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 114.

[206] SR **0.109**.

[207] SR **0.107**.

[208] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 117.

Die Schweiz hat sowohl das Übereinkommen über die Rechte von Menschen mit Behinderungen als auch das Übereinkommen über die Rechte des Kindes ratifiziert und sich damit verpflichtet, die in diesen Übereinkommen enthaltenen Verpflichtungen in allen Lebensbereichen umzusetzen.

Im Falle einer Ratifizierung müsste die Schweiz bei der Umsetzung der Verpflichtungen aus den Übereinkommen die besonderen Bedürfnisse von Menschen mit Behinderungen und von Kindern berücksichtigen.

4.3.5.4 Artikel 19 – Öffentliche Konsultation

Diese Bestimmung fordert die Vertragsparteien auf, die Gesellschaft als Ganzes in die Diskussion über wichtige, durch den Einsatz von KI-Systemen aufgeworfene Fragen einzubeziehen, insbesondere hinsichtlich ihrer sozialen, wirtschaftlichen, rechtlichen, ethischen und ökologischen Auswirkungen.

Der Ausdruck «gegebenenfalls» überlässt es den Parteien, die Modalitäten des Einbezugs der Öffentlichkeit zu gestalten.[209] Die Bestimmung empfiehlt öffentliche Gespräche und Konsultationen verschiedener Interessenträger.

Die Schweiz verfügt bereits über Instrumente, die eine Beteiligung der Öffentlichkeit bei wichtigen Fragen im Zusammenhang mit KI-Systemen ermöglichen. So sehen insbesondere die Gesetzesentwürfe des Bundes in der Regel eine externe Vernehmlassung vor (vgl. VIG). Eine solche wäre auch für Gesetzesentwürfe im Bereich der KI vorzusehen. Das Vernehmlassungsverfahren bezweckt die Mitwirkung der Kantone, der politischen Parteien und der interessierten Kreise an der Meinungsbildung und Entscheidfindung des Bundes (Art. 2 VIG). Gemäss Artikel 4 Absatz 1 VIG kann jede Person oder Organisation an einem Vernehmlassungsverfahren teilnehmen und eine Stellungnahme einreichen.

Im spezifischen Kontext der KI-Systeme sind auch andere Mechanismen zur Beteiligung interessierter Kreise in der Schweiz zu nennen, insbesondere die «*Plateforme Tripartite Suisse* für digitale Gouvernanz und künstliche Intelligenz»[210]. Dabei handelt es sich um ein nationales Informationsnetzwerk für den Austausch über Themen im Zusammenhang mit der Digitalisierung. Diese Plateforme Tripartite wird vom BAKOM betreibt und steht allen Interessierten aus dem Privatsektor, der Zivilgesellschaft, der Wissenschaft und aus allen Verwaltungsebenen offen. Sie soll es ermöglichen, die Bedürfnisse, Anliegen und Erwartungen der verschiedenen Teilnehmer (Stakeholder) zu verstehen und diese bei der Entwicklung geeigneter Regeln im digitalen Bereich, einschliesslich KI, zu berücksichtigen.

---

[209] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 121.

[210] Vgl. www.bakom.admin.ch > Das BAKOM > Internationale Aktivitäten > Informationsgesellschaft International / WSIS > Die Plateforme Tripartite Suisse für digitale Gouvernanz und künstliche Intelligenz.

Aus den obigen Ausführungen geht hervor, dass die Schweiz bereits über angemessene Instrumente zur Erfüllung der Verpflichtungen aus Artikel 19 der KI-Konvention verfügt.

4.3.5.5 Artikel 20 – Digitale Kompetenzen und Fähigkeiten

Diese Bestimmung enthält die Verpflichtung der Vertragsparteien, angemessene digitale Kompetenzen und Fähigkeiten in allen Bevölkerungsgruppen zu fördern und zu unterstützen, einschliesslich spezifischer Spitzenkompetenz für Personen, die für die Identifizierung, Bewertung, Vermeidung und Minderung von Risiken im Zusammenhang mit KI-Systemen zuständig sind.

Die Begriffe der digitalen Kompetenzen und Fähigkeiten beziehen sich auf die Kompetenz, digitale Technologien, einschliesslich KI-Technologien, effizient zu nutzen, zu verstehen und mit ihnen zu interagieren. Dies sind die grundlegenden Fähigkeiten, die die breite Bevölkerung benötigt, um die Chancen und Risiken von KI-Systemen zu verstehen.[211]

Digitale Kompetenzen und Fähigkeiten sind besonders wichtig für diejenigen, die KI bei ihrer Arbeit nutzen. Dazu gehören beispielsweise Mitarbeitende, die ein KI-System in einem Entscheidungsfindungsprozess einsetzen, oder Personen, die für die Beschaffung von KI-Systemen in öffentlichen Verwaltungen zuständig sind. Die beteiligten Personen müssen in der Lage sein, die Herausforderungen zu verstehen, die KI-Systeme mit sich bringen. Solche Kompetenzen ermöglichen es beispielsweise, der Automatisierungs- oder Bestätigungsverzerrung entgegenzuwirken, die dazu führt, dass Menschen Maschinen und technologischen Artefakten mehr Vertrauen schenken als ihrem eigenen, möglicherweise widersprüchlichen Urteilsvermögen, und daher algorithmische Ergebnisse blind und ohne sie zu hinterfragen bestätigen.

Artikel 20 bezieht sich auf die spezifischen Fachkenntnisse der Personen, die gemäss Artikel 16 der Konvention am Rahmen für das Risiko- und Folgenmanagement beteiligt sind. Es versteht sich von selbst, dass diese Personen über die spezifischen Fähigkeiten verfügen müssen, die für die ordnungsgemässe Durchführung der in dieser Bestimmung vorgesehenen Risiko- und Folgenanalyse erforderlich sind.[212]

Der Bund könnte in seinen Kompetenzbereichen Massnahmen ergreifen, zum Beispiel bestimmte Anforderungen an die Ausbildung derjenigen Mitarbeitenden vorsehen, die zur Erfüllung ihrer Aufgaben KI-Systeme einsetzen.

4.3.5.6 Artikel 21 – Schutz der bestehenden Menschenrechte

Diese Bestimmung soll die Koexistenz der KI-Konvention mit anderen internationalen Menschenrechtsübereinkommen und -instrumenten sicherstellen. Sie besagt, dass keine Bestim-

[211] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 122.

[212] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 124.

mung des Übereinkommens so ausgelegt werden darf, dass sie eine einschränkende, abweichende oder nachteilige Wirkung auf die Menschenrechte oder andere damit zusammenhängende Rechte und Rechtspflichten hat, die durch das innerstaatliche Recht einer Vertragspartei oder durch eine andere einschlägige internationale Übereinkunft garantiert sind.

Jede Bezugnahme in der Konvention auf innerstaatliches Recht ist daher so auszulegen, dass sie sich auf Fälle beschränkt, in denen das innerstaatliche Recht ein höheres Schutzniveau für die Menschenrechte vorsieht als das einschlägige Völkerrecht.[213]

Hinsichtlich der Umsetzung in innerstaatliches Recht ist zu diesem Artikel nichts anzumerken.

4.3.5.7 Artikel 22 – Umfassenderer Schutz

Diese Bestimmung schützt die Bestimmungen des innerstaatlichen Rechts und bestehender oder künftiger verbindlicher völkerrechtlicher Übereinkommen, die einen zusätzlichen Schutz für Tätigkeiten im Lebenszyklus von KI-Systemen vorsehen, der über den durch die Konvention gewährleisteten Schutz hinausgeht. Dieser Schutz darf durch die Konvention nicht eingeschränkt werden.[214]

Hinsichtlich der Umsetzung in innerstaatliches Recht ist zu diesem Artikel nichts anzumerken.

4.4 Kapitel VII: Nachfolgemechanismus und Zusammenarbeit

4.4.1 Allgemeines

Dieses Kapitel enthält die Bestimmungen, die den in Artikel 1 Absatz 3 der Konvention vorgesehenen Nachfolgemechanismus bilden. Ziel dieses Mechanismus ist es, die wirksame Umsetzung der KI-Konvention durch die Vertragsparteien zu gewährleisten. Von den vorgesehenen Instrumenten sind die ersten drei auf internationaler Ebene (Art. 23–25) und das vierte auf nationaler Ebene (Art. 26) angesiedelt.

4.4.2 Artikel 23 – Konferenz der Vertragsparteien

Dieser Artikel sieht die Einrichtung eines Organs für die Konvention vor, namentlich die Konferenz der Vertragsparteien, die sich aus Vertretern der Vertragsparteien zusammensetzt. Die Einrichtung dieses Organs gewährleistet die gleichberechtigte Beteiligung aller Vertragsparteien am Entscheidungs- und Überwachungsprozess der Konvention und stärkt die Zusammenarbeit zwischen den Vertragsparteien, um eine angemessene und wirksame Implementierung der Konvention sicherzustellen.[215]

---

[213] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 126.

[214] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 127.

[215] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 130.

Dieses Organ wird seine eigene Geschäftsordnung verabschieden (Art. 23 Abs. 4). Es wird von der Generalsekretärin oder dem Generalsekretär des Europarats je nach Bedarf einberufen, in jedem Fall aber, wenn die Mehrheit der Vertragsparteien oder des Ministerkomitees die Einberufung verlangt (Art. 23 Abs. 3).

Die Konferenz der Vertragsparteien verfügt über die üblichen Überwachungsbefugnisse. Hervorzuheben ist insbesondere die Möglichkeit, Änderungen der Konvention durch Vorschläge gemäss Artikel 28 der Konvention vorzunehmen. Darüber hinaus hat die Vertragsparteienkonferenz eine allgemeine beratende Funktion in Bezug auf die Konvention. Sie kann spezifische Empfehlungen zu allen Fragen der Auslegung oder der Anwendung der Konvention abgeben, beispielsweise durch Vorschläge zur Auslegung verschiedener in der Konvention verwendeter Rechtsbegriffe. Diese Empfehlungen sind zwar nicht rechtsverbindlich, können aber als Ausdruck eines gemeinsamen Verständnisses der Vertragsparteien zu einer bestimmten Frage angesehen werden, das von den Vertragsparteien bei der Umsetzung der Konvention nach Treu und Glauben zu berücksichtigen ist.[216]

Im Falle eines Beitritts der Schweiz zur Konvention muss sie eine aus einer oder mehreren Personen bestehende Vertretung für die Konferenz der Vertragsparteien ernennen.

4.4.3 Artikel 24 – Berichterstattungspflicht

Zur Erleichterung der Zusammenarbeit und zur regelmässigen Unterrichtung über die Umsetzung der Konvention legt jede Vertragspartei der Konferenz der Vertragsparteien innerhalb von zwei Jahren nach dem Tag, an dem sie Vertragspartei geworden ist, und danach in regelmässigen Abständen einen Bericht vor, in dem die von ihr ergriffenen Massnahmen zur Umsetzung von Artikel 3 Absatz 1 Buchstaben a und b (vgl. Ziff. 4.2.3.1) im Einzelnen aufgeführt sind.

Die Konferenz der Vertragsparteien legt die Form und das Verfahren für die Vorlage des Berichts in Übereinstimmung mit ihrer Geschäftsordnung fest.

Wenn die Schweiz der Konvention beitritt, muss sie gemäss dieser Bestimmung und gemäss den von der Konferenz der Vertragsparteien festgelegten Modalitäten Bericht erstatten.

4.4.4 Artikel 25 – Internationale Zusammenarbeit

Diese Bestimmung legt die Verpflichtung der Vertragsparteien fest, bei der Umsetzung der Ziele der Konvention zusammenzuarbeiten.

Darüber hinaus werden die Vertragsparteien ermutigt, gegebenenfalls Nichtvertragsstaaten dabei zu unterstützen, im Einklang mit den Bestimmungen der Konvention zu handeln und dieser beizutreten (Abs. 1).

[216] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 132, Bst. c.

Darüber hinaus tauschen die Vertragsparteien relevante und nützliche Informationen über Aspekte von KI aus, die erhebliche positive oder negative Auswirkungen auf die Ausübung der Menschenrechte, das Funktionieren der Demokratie und die Achtung der Rechtsstaatlichkeit haben können, einschliesslich der im privaten Sektor festgestellten Risiken und Auswirkungen. Dabei werden sie ermutigt, gegebenenfalls relevante Interessengruppen und Staaten, die nicht Vertragsparteien der Konvention sind, einzubeziehen (Abs. 2).

Schliesslich werden die Vertragsparteien ermutigt, die Zusammenarbeit zu verstärken, gegebenenfalls auch mit relevanten Teilnehmergruppen (z. B. Vertreterinnen und Vertreter von Nichtregierungsorganisationen), um Risiken und nachteilige Auswirkungen auf Menschenrechte, Demokratie und Rechtsstaatlichkeit im Zusammenhang mit Tätigkeiten während des Lebenszyklus von KI-Systemen zu vermeiden und zu mindern.

Diese Bestimmung ist eher vage und lässt den Parteien relativ freie Hand bei der Art der geplanten Zusammenarbeit. Es handelt sich hier eher um einen Informationsaustausch als um eine operative Zusammenarbeit zwischen den Behörden.

4.4.5 Artikel 26 – Wirksame Aufsichtsmechanismen

Artikel 26 verpflichtet die Vertragsparteien, einen oder mehrere wirksame Mechanismen zur Aufsicht über die Einhaltung der Verpflichtungen aus der Konvention einzurichten oder zu bestimmen. Dies bedeutet, dass die Vertragsparteien die Mechanismen ihrer jeweiligen Rechtsordnungen überprüfen und gegebenenfalls deren Aufgaben neu definieren oder sogar gänzlich neue Strukturen schaffen müssen.

Gemäss dem erläuternden Bericht müssen diese Mechanismen funktionell von der Exekutive und der Legislative unabhängig sein. Dieser Begriff umfasst verschiedene Arten der funktionellen Unabhängigkeit. Dabei kann es sich um Kontrollfunktionen handeln, die in bestimmte Regierungsorgane integriert sind und die Entwicklung und den Einsatz von KI-Systemen beurteilen oder überwachen.

Die betreffenden Stellen müssen auch über die zur wirksamen Erfüllung ihrer Aufgaben erforderlichen Befugnisse, Fachkenntnisse, einschliesslich technischer Kenntnisse und Fähigkeiten, und sonstigen Ressourcen verfügen. Im Falle einer gemeinsamen Aufsicht durch mehrere Stellen verlangt das Übereinkommen eine wirksame Zusammenarbeit.[217]

Wirksame Aufsichtsmechanismen sollten auch im privaten Sektor vorgesehen werden, wo eine direkte oder indirekte horizontale Wirkung der Grundrechte zwischen Privaten besteht oder in Zukunft erkannt wird (vgl. Ziff. 4.2.3.1).

Diese Bestimmung verlangt ein Tätigwerden des Gesetzgebers, um entweder eine oder mehrere bestehende Behörden mit der Aufsicht über die Einhaltung der Konvention zu betrauen oder eine neue Behörde zu schaffen. Im schweizerischen Recht gibt es verschie-

[217] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 141 ff.

dene unabhängige Aufsichtsbehörden (z. B. EDÖB), aber keine, die den gesamten Anwendungsbereich der Konvention abdeckt. Der Gesetzgeber sollte die Organisation und die Eingriffsbefugnisse der Aufsichtsbehörde(n) im Zusammenhang mit der Konvention regeln. Wären mehrere Aufsichtsbehörden zuständig, sollte zudem die Koordination zwischen ihnen geregelt werden.

4.5 Kapitel VIII: Schlussbestimmungen

Die Bestimmungen der Artikel 27 bis 36 lehnen sich mit wenigen Ausnahmen im Wesentlichen an das Muster für die Schlussbestimmungen der im Rahmen des Europarats geschlossenen Übereinkommen, Zusatzprotokolle und Änderungsprotokolle an, das vom Ministerkomitee des Europarats anlässlich der 1291. Versammlung der Ministerdelegierten vom 5. Juli 2017 verabschiedet wurde.[218]

Besonders hervorzuheben ist Artikel 30 über die Unterzeichnung und das Inkrafttreten der Konvention. Nach Absatz 1 liegt das Übereinkommen für die Mitgliedstaaten des Europarats, für die Nichtmitgliedstaaten des Europarats, die an der Ausarbeitung der Konvention beteiligt waren (Argentinien, Australien, Kanada, Costa Rica, der Heilige Stuhl, Israel, Japan, Mexiko, Peru, die Vereinigten Staaten von Amerika und Uruguay) sowie für die Europäische Union zur Unterzeichnung auf.

Nach Artikel 30 Absatz 3 tritt die Konvention am ersten Tag des Monats in Kraft, der auf einen Zeitabschnitt von drei Monaten nach dem Tag folgt, an dem fünf Unterzeichnerstaaten, darunter mindestens drei Mitgliedstaaten des Europarats, erklärt haben, durch die Konvention gebunden zu sein (durch Ratifikation, Annahme oder Genehmigung, vgl. Art. 30 Abs. 2).

Nach Inkrafttreten der Konvention können andere Nichtmitgliedstaaten, die sich nicht an der Ausarbeitung der Konvention beteiligt haben, vom Ministerkomitee des Europarats eingeladen werden, dem Übereinkommen gemäss Artikel 31 beizutreten.

4.6 Zwischenfazit

Mit der vorliegenden Analyse der KI-Konvention konnte der gesetzgeberische Handlungsbedarf im schweizerischen Recht für den Fall einer Ratifikation der Konvention durch die Schweiz skizziert werden.

Unabhängig von der Frage der Ratifizierung ermöglichte diese Analyse auch eine Beurteilung des Schutzniveaus des schweizerischen Rechts im Hinblick auf die Herausforderungen, die KI-Systeme heute für Menschenrechte, Demokratie und Rechtsstaatlichkeit im Allgemeinen darstellen, da die KI-Konvention die grossen, international anerkannten Herausforderungen beleuchtet. Die KI-Konvention stellt somit eine Art Leitfaden für die Überprüfung des aktuellen Stands des schweizerischen Rechts in diesen Bereichen dar.

Es ist darauf hinzuweisen, dass die Analyse in Bezug auf bestimmte Aspekte vertieft werden muss, um die Gesamtheit der bestehenden einschlägigen Rechtsvorschriften einzubeziehen.

---

[218] Erläuternder Bericht zur KI-Konvention (Fn. 20), N 145.

Die vorliegende Analyse ist nur ein erster Schritt im Hinblick auf eine umfassendere Untersuchung, die gegebenenfalls im Rahmen der weiteren Arbeiten nach dem Grundsatzentscheid des Bundesrates zur Regulierung der KI durchgeführt wird.

Die Analyse hat folgendes Zwischenfazit hervorgebracht:

- Ziel der KI-Konvention ist der Schutz der Menschenrechte, einer funktionierenden Demokratie und der Rechtsstaatlichkeit. Die KI-Konvention richtet sich grundsätzlich an Staaten und enthält Normen, die nicht unmittelbar anwendbar sind.

- Die KI-Konvention schreibt keine genauen und konkreten Massnahmen zur Erreichung ihrer Ziele vor. Die Vertragsparteien haben daher einen grossen Ermessensspielraum bei der Umsetzung der Konvention.

  So legt die KI-Konvention beispielsweise keine allgemeinen Zulassungsvoraussetzungen für KI-Systeme fest, sondern bleibt auf einer allgemeineren Ebene. Zum Verhältnis zwischen der KI-Konvention und der KI-Verordnung siehe Kapitel Ziff. 5.3.4.

- Für den Handlungsbedarf des Gesetzgebers im Falle einer Ratifikation hat die Analyse drei Konstellationen ergeben:

  - Hinsichtlich einiger Bestimmungen der Konvention scheint das schweizerische Recht ein Schutzniveau zu bieten, das den Anforderungen der Konvention genügt, so dass ein Tätigwerden des Gesetzgebers derzeit nicht erforderlich erscheint. Dies gilt beispielsweise für Artikel 5 (Integrität demokratischer Prozesse und Achtung der Rechtsstaatlichkeit) und Artikel 19 (Öffentliche Konsultation).

  - Hinsichtlich anderer Bestimmungen der Konvention enthält das schweizerische Recht bereits einschlägige Bestimmungen, die jedoch im Vergleich zu den Verpflichtungen der Konvention nicht weit genug gehen. Zur Umsetzung der Konventionsbestimmungen wären daher Anpassungen notwendig. Zu erwähnen sind insbesondere Artikel 8 (Transparenz und Aufsicht), Artikel 13 (Sichere Innovation), Artikel 14 (Rechtsmittel) und Artikel 15 (Verfahrensgarantien).

    Die Umsetzung einzelner Schlüsselprinzipien, wie z. B. jenes in Artikel 8 (Transparenz und Aufsicht), ermöglicht es zudem, die Wirksamkeit des bereits bestehenden rechtlichen Rahmens des schweizerischen Rechts zu verbessern, z. B. in den Bereichen Verantwortlichkeit, Gleichstellung und Nichtdiskriminierung sowie Datenschutz.

  - Für andere Bestimmungen sieht das schweizerische Recht keine entsprechenden Normen vor. Dies gilt insbesondere für Artikel 16 (Rahmen für das Risiko- und Folgenmanagement) und Artikel 26 (Wirksame Aufsichtsmechanismen). Hier könnte der Gesetzgeber erwägen, neue Massnahmen zu ergreifen, da das schweizerische Recht nur punktuelle Aspekte regelt.

Zusammenfassend kann festgehalten werden, dass der geltende Rechtsrahmen zwar relevante Bestimmungen enthält, aber in vielen Fällen ergänzt werden müsste. Daraus lässt sich schliessen, dass im Falle einer Ratifizierung der KI-Konvention durch die Schweiz Anpassungen notwendig wären.

- Über Art und Umfang der zu treffenden Massnahmen hat gegebenenfalls der Gesetzgeber zu entscheiden. Bei der Ausarbeitung eines Gesetzgebungsvorhabens selbst ist nach den Grundsätzen der Gesetzgebungstechnik und unter Berücksichtigung der zu treffenden politischen Entscheidungen auch festzulegen, wie die neuen Normen in das geltende Recht integriert werden sollen. So ist zum Beispiel zu prüfen, ob durch die neuen Massnahmen ein oder mehrere bestehende Gesetze ergänzt werden sollen, oder ob ein neues Gesetz zu erlassen ist. Hinsichtlich der Regelungsdichte sind insbesondere die Vorgaben von Artikel 164 BV zu beachten.

  Es zeichnet sich jedoch bereits jetzt ab, dass aufgrund der möglicherweise betroffenen Rechtsbereiche voraussichtlich mehrere Gesetze angepasst werden müssen. Bereits jetzt ist ein erheblicher Koordinationsbedarf absehbar. Ein solcher besteht insbesondere mit dem DSG.

- Die Analyse hat auch gezeigt, dass die Verhältnismässigkeit ein wichtiger Aspekt bei der Umsetzung ist. Die Tragweite der Verpflichtungen aus der Konvention sollte an die Schwere möglicher Verletzungen angepasst werden. Auch hier ist die Frage, wie diese Abstufung konkret aussehen sollte, in erster Linie von gesetzgeberischer und politischer Natur.

- Hinsichtlich des Privatsektors hat die Analyse gezeigt, dass sich der Anwendungsbereich der Konvention auf Fälle beschränken würde, in denen eine direkte oder indirekte horizontale Wirkung der Grundrechte zwischen Privaten besteht oder in Zukunft erkannt wird (vgl. Ziff. 4.2.3.1). Die KI-Konvention lässt Raum für Besonderheiten des schweizerischen Rechts.

5 Verordnung der Europäischen Union zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz

5.1 Struktur des Kapitels und Methode

In diesem Kapitel wird zunächst der Inhalt der KI-Verordnung präsentiert (vgl. Ziff. 5.2).

Danach folgt eine rechtliche Würdigung der KI-Verordnung, in der einige Aspekte, die aus Sicht des Schweizer Rechts relevant sind, analysiert werden (vgl. Ziff. 5.3).

In diesem Rahmen wird auf die rechtlichen Auswirkungen der KI-Verordnung auf die Schweizer Akteure (vgl. Ziff. 5.3.1), das Verhältnis zwischen der KI-Verordnung und dem Abkommen zwischen der Schweizerischen Eidgenossenschaft und der Europäischen Union über die gegenseitige Anerkennung von Konformitätsbewertungen (MRA) (vgl. Ziff. 5.3.2) und das Verhältnis zum Angemessenheitsbeschluss der Europäischen Kommission im Bereich Datenschutz (vgl. Ziff. 5.3.3) eingegangen. Das Verhältnis zur KI-Konvention wird ebenfalls thematisiert (vgl. Ziff. 5.3.4). Anschliessend werden weitere ausgewählte Aspekte erörtert (vgl. Ziff. 5.3.5), bevor das Kapitel mit einem Zwischenfazit abgeschlossen wird (vgl. Ziff. 5.4).

Aus methodischer Sicht weicht die Präsentation von der vorstehenden Analyse der KI-Konvention des Europarats ab (vgl. Ziff. 4). Die KI-Verordnung ist eine EU-interne Regelung und ist für die Schweiz nicht bindend. Grundsätzlich hat diese Verordnung derzeit im Schweizer Recht keine Entsprechung. Eine Analyse, in der jede ihrer Bestimmungen in Bezug auf das Schweizer Recht untersucht wird, wie es bei der KI-Konvention gemacht wurde, erscheint daher derzeit nicht angebracht. Stattdessen soll der Inhalt der KI-Verordnung präsentiert und gewisse relevante Aspekte hervorgehoben werden, um so eine allgemeine Würdigung aus Sicht des Schweizer Rechts vornehmen zu können. Da die Verordnung in der EU erst vor Kurzem verabschiedet wurde und noch viele Fragen hinsichtlich ihrer Anwendung geklärt werden müssen, wird hier nur auf die wichtigen Herausforderungen eingegangen, die für die Schweiz derzeit relevant sind.

5.2 Inhalt der Verordnung

5.2.1 Ausgangslage

Am 21. April 2021 hat die Europäische Kommission ein Legislativpaket zu KI vorgestellt, das einen Vorschlag für eine Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und einen Vorschlag für eine Verordnung über Maschinen und Geräte (neue Konformitätsverfahren, die u. a. das Internet der Dinge regeln) enthielt.[219]

Aus Sicht der Europäischen Kommission will die EU mit dieser Verordnung ihre Rolle als Pionierin bei der Regulierung der digitalen Wirtschaft und als Vorreiterin bei der Festlegung von

[219] Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates vom 21. April 2021 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM/2021/206 final, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A52021PC0206 (abgerufen am 26. August 2024); Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über Maschinenprodukte, COM(2021) 202 final, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A52021PC0202 (abgerufen am 26. August 2024).

Standards festigen, die Einfluss nehmen sollen auf die internationale Debatte zum Umgang mit KI.

Die KI-Verordnung stellt die erste horizontale und direkt verbindliche KI-Regulierung auf regionaler Ebene dar. Sie deckt insbesondere die Entwicklung, das Inverkehrbringen, die Inbetriebnahme und die Verwendung von KI-Systemen in der EU ab. Sie wurde vom Europäischen Parlament am 13. März 2024 und vom Rat am 21. Mai 2024 verabschiedet, am 12. Juni 2024 unterzeichnet und im Amtsblatt der EU am 12. Juli 2024 veröffentlicht. Sie ist am 1. August 2024 in Kraft getreten.[220]

Diese Verordnung ist für die Schweiz nicht bindend.

5.2.2 Ziele der Regelung

Gemäss Artikel 1 Absatz 1 besteht der Zweck der Verordnung darin, das Funktionieren des Binnenmarkts zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen KI zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der EU-Charta verankerten Grundrechte, einschliesslich Demokratie, Rechtsstaatlichkeit und Umweltschutz, vor schädlichen Auswirkungen von KI-Systemen in der EU zu gewährleisten und die Innovation zu unterstützen.

Das Ziel besteht also darin, einen einheitlichen Rechtsrahmen, insbesondere für die Entwicklung, das Inverkehrbringen, die Inbetriebnahme und die Verwendung von KI-Systemen in der EU, zu schaffen.[221] Die Verordnung soll gewährleisten, dass die in der EU in Verkehr gebrachten KI-Systeme sicher sind und den geltenden europäischen Normen entsprechen, namentlich im Bereich der Produktsicherheit.[222] Sie soll auch den grenzüberschreitenden Verkehr KI-gestützter Waren und Dienstleistungen ermöglichen und eine Fragmentierung des Markts verhindern.

Die Wendung «und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta verankerten Grundrechte [...] zu gewährleisten» (vgl. Art. 1 Abs. 1) zeigt, dass der europäische Gesetzgeber über diesen auf der Produktsicherheit basierenden Regulierungsansatz auch andere Interessen schützen wollte, insbesondere die Grundrechte der betroffenen Personen. Dies führt zu einer Verordnung mit hybriden Zielen (vgl. Ziff. 5.3.4).

5.2.3 Risikobasierter Ansatz

Die KI-Verordnung weist einen risikobasierten Ansatz auf. So nimmt sie eine Einstufung der KI-Systeme nach ihrem Risiko für die Gesundheit, die Sicherheit und die Grundrechte vor.

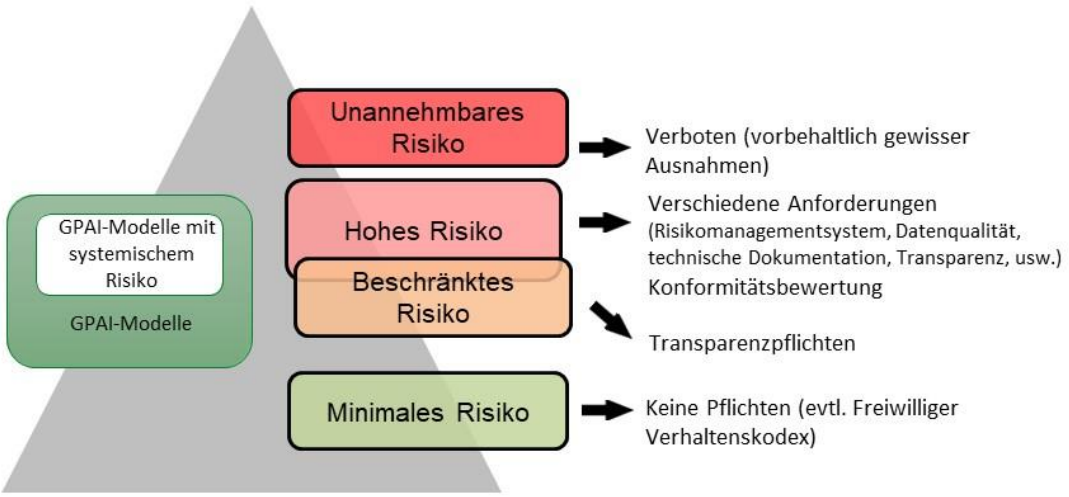
---

[220] Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Fn. 3).

[221] DAVID ROSENTHAL, Der EU AI Act – Verordnung über künstliche Intelligenz, Jusletter 5. August 2024, 5 N 6.

[222] MARTINA ARIOLI, Risikomanagement nach der EU-Verordnung über Künstliche Intelligenz, Jusletter IT 4. Juli 2024, 4 N 8 f.

Gestützt auf diese Einstufung sieht die Verordnung verschiedene Pflichten für den Marktzugang vor.[223]



Quelle: BJ

Gemäss diesem Ansatz sind KI-Systeme, die unannehmbare Risiken aufweisen, bis auf einige Ausnahmen verboten (vgl. Ziff. 5.2.6).

KI-Systeme, die als hochriskant eingestuft werden, sind zulässig, müssen aber viele Anforderungen erfüllen, damit sie in der EU in Verkehr gebracht werden können (vgl. Ziff. 5.2.7). Insbesondere müssen sie einer Konformitätsbewertung unterzogen werden (vgl. Ziff. 5.2.7.4). In ihrer Folgenabschätzung der Verordnung[224] kam die Europäische Kommission zum Schluss, dass die Hochrisiko-KI-Anwendungen schätzungsweise nur 5 bis 15 Prozent aller Anwendungen auf dem Markt ausmachen.

KI-Systeme, die nur ein beschränktes Risiko aufweisen, unterliegen lediglich Transparenzpflichten (vgl. Ziff. 5.2.8). Wenn diese Systeme jedoch die Kriterien erfüllen, um als hochriskant eingestuft zu werden, gelten die für diese Systeme vorgesehenen Pflichten ebenfalls.

---

[223] Vgl. ANGELA MÜLLER, Der Artificial intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz, Zeitschrift für Europarecht 1/2022, 1 ff., 7 ff. und 15 f. (im Zusammenhang mit einer früheren Version der KI-Verordnung).

[224] Commission staff working document, Impact assessment – Accompanying the Proposal for a Regulation of the European Parliament and of the Council, Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative act, SWD(2021) 84 final, 68 (abrufbar unter: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0084 (abgerufen am 26. August 2024).

Fällt ein KI-System unter keine der oben erwähnten Kategorien, so wird es als KI-System mit minimalem Risiko betrachtet. Gemäss der Europäischen Kommission weist die Mehrheit der KI-Systeme ein minimales Risiko auf. Die Verordnung sieht für diese Systeme daher kein staatliches Eingreifen vor. Sie erlaubt die freie Verwendung von Anwendungen wie Videospielen mit integrierten KI-Systemen oder KI-basierten Spam-Filtern. In diesem Fall werden die betroffenen Akteure angehalten, Verhaltenskodizes zu erstellen.

Die Verordnung regelt auch die KI-Modelle mit allgemeinem Verwendungszweck (Englisch: «general purpose AI model», im Folgenden: GPAI-Modelle), die im ursprünglichen Verordnungsvorschlag nicht enthalten waren. Die Verordnung sieht nun Pflichten für alle GPAI-Modelle und zusätzliche Pflichten für GPAI-Modelle vor, die systemische Risiken beinhalten (vgl. Ziff. 5.2.10).

5.2.4 Begriffsbestimmungen

In Artikel 3 der KI-Verordnung werden gewisse Begriffe definiert (insgesamt 68). Diese Analyse konzentriert sich auf die Begriffe «KI-System» und «KI-Modell mit allgemeinem Verwendungszweck».

5.2.4.1 System der künstlichen Intelligenz

Gemäss Artikel 3 Nummer 1 der KI-Verordnung ist ein KI-System «ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können».

Die Verordnung will zukunftssicher sein und sowohl aktuelle als auch künftige technologische Entwicklungen im KI-Bereich abdecken. So lehnt sich die Begriffsbestimmung der EU für das KI-System eng an die Arbeiten der internationalen Organisationen an, die im KI-Bereich tätig sind, wie die OECD, um die Rechtssicherheit zu gewährleisten sowie die internationale Konvergenz und die allgemeine Akzeptanz zu fördern.

Angesichts der Ähnlichkeiten dieser Definition des KI-Systems mit jener in der KI-Konvention, die sich ebenfalls an den Arbeiten der OECD orientiert, wird für weitere Erläuterungen auf die Ausführungen zur Begriffsbestimmung der KI-Konvention verwiesen (vgl. Ziff. 4.2.2.1).

Ausserdem wird die Europäische Kommission Leitlinien für die Anwendung der Definition eines KI-Systems erarbeiten (vgl. Art. 96 Abs. 1 Bst. f).

5.2.4.2 Modell der künstlichen Intelligenz mit allgemeinem Verwendungszweck

In der Verordnung wird zwischen KI-Modellen mit allgemeinem Verwendungszweck und KI-Systemen unterschieden.

Gemäss Artikel 3 Nummer 63 ist ein KI-Modell mit allgemeinem Verwendungszweck «ein KI-Modell – einschliesslich der Fälle, in denen ein solches KI-Modell mit einer grossen Datenmenge unter umfassender Selbstüberwachung trainiert wird –, das eine erhebliche allgemeine Verwendbarkeit aufweist und in der Lage ist, unabhängig von der Art und Weise seines Inverkehrbringens ein breites Spektrum unterschiedlicher Aufgaben kompetent zu

erfüllen, und das in eine Vielzahl nachgelagerter Systeme oder Anwendungen integriert werden kann, ausgenommen KI-Modelle, die vor ihrem Inverkehrbringen für Forschungs- und Entwicklungstätigkeiten oder die Konzipierung von Prototypen eingesetzt werden».

Die Begriffsbestimmung des KI-Modells mit allgemeinem Verwendungszweck beruht auf den wesentlichen funktionalen Merkmalen dieser Art von Modell, insbesondere auf der allgemeinen Verwendbarkeit und der Fähigkeit, ein breites Spektrum unterschiedlicher Aufgaben kompetent zu erfüllen. Bespiele dafür sind «GPT» und «Llama».[225]

Aus dieser Begriffsbestimmung geht hervor, dass KI-Modelle mit allgemeinem Verwendungszweck wesentliche Komponenten von KI-Systemen sind, aber für sich allein keine KI-Systeme darstellen. Damit KI-Modelle mit allgemeinem Verwendungszweck zu KI-Systemen werden, ist die Hinzufügung weiterer Komponenten, beispielsweise einer Nutzerschnittstelle, erforderlich. Daher sind diese Modelle in der Regel in KI-Systeme integriert und Teil davon.

Wenn ein KI-System auf einem KI-Modell mit allgemeinem Verwendungszweck beruht, spricht man von KI-System mit allgemeinem Verwendungszweck. Dabei handelt es sich um ein System, das in der Lage ist, einer Vielzahl von Zwecken sowohl für die direkte Verwendung als auch für die Integration in andere KI-Systeme zu dienen (vgl. Art. 3 Nr. 66).

5.2.5 Geltungsbereich

5.2.5.1 Umfang des Geltungsbereichs

Die KI-Verordnung hat einen weiten Geltungsbereich. Sie gilt sowohl für Private als auch für Behörden.[226]

Um den Geltungsbereich zu verstehen, ist zunächst zu beachten, dass in der Verordnung zwischen folgenden Rollen unterschieden wird:

- Anbieter: Dabei handelt es sich um jede natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System oder ein KI-Modell mit allgemeinem Verwendungszweck entwickelt oder entwickeln lässt *und* es unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringt oder das KI-System unter ihrem eigenen Namen oder ihrer Handelsmarke in Betrieb nimmt, sei es entgeltlich oder unentgeltlich. Er kann in der EU oder in einem Drittland niedergelassen sein (vgl. Art. 3 Nr. 3).

  Unter «Inverkehrbringen» ist die erstmalige Bereitstellung eines KI-Systems oder eines KI-Modells mit allgemeinem Verwendungszweck auf dem EU-Markt zu verstehen (vgl. Art. 3 Nr. 9).

[225] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 3 N 5.

[226] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 2 N 2; 6 N 11.

Mit «Inbetriebnahme» ist die Bereitstellung eines KI-Systems zum Erstgebrauch direkt an den Betreiber oder zum Eigengebrauch entsprechend seiner Zweckbestimmung gemeint (vgl. Art. 3 Nr. 11).

- Betreiber: Hier handelt es sich um jede natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung verwendet, es sei denn, diese Verwendung erfolgt im Rahmen einer persönlichen und nicht beruflichen Tätigkeit (vgl. Art. 3 Nr. 4).

- Einführer: Jede in der EU ansässige oder niedergelassene natürliche oder juristische Person, die ein KI-System in Verkehr bringt, das den Namen oder die Handelsmarke einer in einem Drittland niedergelassenen natürlichen oder juristischen Person trägt (vgl. Art. 3 Nr. 6).

- Händler: Jede natürliche oder juristische Person in der Lieferkette, die ein KI-System auf dem EU-Markt bereitstellt, mit Ausnahme des Anbieters oder des Einführers (vgl. Art. 3 Nr. 7).

- Akteur: ein Anbieter, Produkthersteller, Betreiber, Bevollmächtigter, Einführer oder Händler (vgl. Art. 3 Nr. 8).

Die KI-Verordnung gilt in erster Linie für die in der EU oder einem Drittland niedergelassenen oder ansässigen Anbieter, die in der EU KI-Systeme in Verkehr bringen oder in Betrieb nehmen oder KI-Modelle mit allgemeinem Verwendungszweck in Verkehr bringen (vgl. Art. 2 Abs. 1 Bst. a).

Sie gilt auch für Betreiber von KI-Systemen, die ihren Sitz in der EU haben oder sich in der EU befinden (vgl. Art. 2 Abs. 1 Bst. b).

Weiter gilt die Verordnung für Anbieter und Betreiber von KI-Systemen, die ihren Sitz in einem Drittland haben oder sich in einem Drittland befinden, wenn die vom KI-System hervorgebrachte Ausgabe in der EU verwendet wird (Art. 2 Abs. 1 Bst. c). Mit dieser Bestimmung soll verhindert werden, dass Massnahmen zur Umgehung der Rechtsvorschriften umgesetzt werden könnten. Auf die Frage, welche Auswirkungen die KI-Verordnung auf die Schweizer Akteure hat, wird weiter unten eingegangen (vgl. Ziff. 5.3.1).

Die Verordnung gilt ebenfalls für Einführer und Händler von KI-Systemen (Art. 2 Abs. 1 Bst. d), Produkthersteller, die ein KI-System zusammen mit ihrem Produkt unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringen oder in Betrieb nehmen (Art. 2 Abs. 1 Bst. e) und Bevollmächtigte von Anbietern, die nicht in der EU niedergelassen sind (Art. 2 Abs. 1 Bst. f).

Ferner gilt sie für betroffene Personen, die sich in der EU befinden (Art. 2 Abs. 1 Bst. g). Dies betrifft beispielsweise die Rechtsbehelfe gemäss den Artikeln 85 ff. der Verordnung.

5.2.5.2 Ausnahmen

Die Verordnung sieht mehrere Ausnahmen von ihrer Anwendung vor. Dies sind namentlich Folgende:

- KI-Systeme, die ausschliesslich für militärische Zwecke, Verteidigungszwecke oder Zwecke der nationalen Sicherheit in Verkehr gebracht, in Betrieb genommen oder, mit oder ohne Änderungen, verwendet werden, unabhängig von der Art der Einrichtung, die diese Tätigkeiten ausübt, sind vom Geltungsbereich der Verordnung ausgenommen (Art. 2 Abs. 3). Der Bereich der nationalen Sicherheit ist nicht Bestandteil des EU-Rechts, sondern liegt in der ausschliesslichen Verantwortung der Mitgliedstaaten (Art. 4 Abs. 2 EUV). Der Ausschluss des Militär- und Verteidigungsbereichs gründet sich auf Artikel 4 Absatz 2 EUV sowie auf die Besonderheiten der Verteidigungspolitik der Mitgliedstaaten und der gemeinsamen Verteidigungspolitik der EU gemäss Titel V Kapitel 2 EUV, die dem Völkerrecht unterliegen. Wird ein KI-System jedoch vorübergehend oder ständig ausserhalb der erwähnten Bereiche für andere Zwecke verwendet, etwa für zivile oder humanitäre Zwecke oder für Zwecke der Strafverfolgung oder öffentlichen Sicherheit, so fällt dieses System in den Geltungsbereich der Verordnung (vgl. E. 24 der Verordnung).

- Die Verordnung gilt weder für Behörden in Drittländern noch für internationale Organisationen, die gemäss Artikel 2 Absatz 1 in den Geltungsbereich der Verordnung fallen, soweit diese Behörden oder Organisationen KI-Systeme im Rahmen internationaler Übereinkünfte im Bereich der Strafverfolgung und justiziellen Zusammenarbeit mit der EU oder mit einem oder mehreren Mitgliedstaaten verwenden, sofern Garantien hinsichtlich des Schutzes der Grundrechte und der Grundfreiheiten von Personen gewährt werden (Art. 2 Abs. 4).

- Für KI-Systeme, die als Hochrisiko-KI-Systeme gemäss Artikel 6 Absatz 1 eingestuft sind und im Zusammenhang mit Produkten stehen, die unter die in Anhang I Abschnitt B aufgeführten Harmonisierungsrechtsvorschriften fallen (z. B. Zivilluftfahrt, land- und forstwirtschaftliche Fahrzeuge, Schiffsausrüstung), gelten nur Artikel 6 Absatz 1 und die Artikel 102–109 und 112 der Verordnung (vgl. Art. 2 Abs. 2).

- Die Verordnung gilt nicht für KI-Systeme oder KI-Modelle, einschliesslich ihrer Ausgabe, die für den alleinigen Zweck der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden (Art. 2 Abs. 6). Forschungs-, Test- und Entwicklungstätigkeiten fallen nicht in den Geltungsbereich der Verordnung, müssen aber im Einklang mit dem geltenden EU-Recht durchgeführt werden. Tests unter Realbedingungen sind von diesem Ausschluss nicht betroffen (Art. 2 Abs. 8 und Art. 60).

- Die Verordnung gilt nicht für die Pflichten von Betreibern, die natürliche Personen sind und KI-Systeme im Rahmen einer ausschliesslich persönlichen und nicht beruflichen Tätigkeit verwenden (Art. 2 Abs. 10).

- Ferner gilt die Verordnung ebenfalls nicht für KI-Systeme, die unter freien und quelloffenen Lizenzen bereitgestellt werden, es sei denn, sie werden als Hochrisiko-KI-Systeme oder als ein KI-System, das unter Artikel 5 (Verbotene Praktiken) oder 50 (Transparenzpflichten unterliegende KI-Systeme) fällt, in Verkehr gebracht oder in Betrieb genommen (vgl. Art. 2 Abs. 12).

5.2.6 Verbotene Praktiken

Gemäss dem oben beschriebenen risikobasierten Ansatz (vgl. Ziff. 5.2.3) verbietet das Kapitel II der KI-Verordnung acht Praktiken, die als unannehmbares Risiko betrachtet werden (vgl. Art. 5 Abs. 1):[227]

- KI-Systeme, die Techniken der unterschwelligen Beeinflussung einsetzen: Verboten sind das Inverkehrbringen, die Inbetriebnahme oder die Verwendung von KI-Systemen, die Techniken der unterschwelligen Beeinflussung oder absichtlich manipulative oder täuschende Techniken mit dem Ziel oder der Wirkung einsetzen, das Verhalten einer Person oder einer Gruppe von Personen wesentlich zu verändern, indem ihre Fähigkeit, eine fundierte Entscheidung zu treffen, deutlich beeinträchtigt wird, wodurch sie veranlasst wird, eine Entscheidung zu treffen, die sie andernfalls nicht getroffen hätte, und zwar in einer Weise, die dieser Person, einer anderen Person oder einer Gruppe von Personen erheblichen Schaden zufügt oder mit hinreichender Wahrscheinlichkeit zufügen wird (Art. 5 Abs. 1 Bst. a).

- KI-Systeme, die eine Vulnerabilität oder Schutzbedürftigkeit ausnutzen: Verboten sind Systeme, die eine Vulnerabilität oder Schutzbedürftigkeit einer natürlichen Person oder einer Gruppe von Personen aufgrund ihres Alters, einer Behinderung oder der sozialen oder wirtschaftlichen Situation mit dem Ziel oder der Wirkung ausnutzen, das Verhalten dieser Person oder einer dieser Gruppe angehörenden Person in einer Weise wesentlich zu verändern, die dieser Person oder einer anderen Person erheblichen Schaden zufügt oder mit hinreichender Wahrscheinlichkeit zufügen wird (Art. 5 Abs. 1 Bst. b).

- Soziale Bewertung durch öffentliche und private Akteure: Verboten sind Systeme, die natürliche Personen oder Gruppen von Personen auf der Grundlage ihres sozialen Verhaltens oder bekannter, abgeleiteter oder vorhergesagter persönlicher Eigenschaften oder Persönlichkeitsmerkmale bewerten oder einstufen, *wenn* die soziale Bewertung zu einem oder beiden der folgenden Ergebnisse führt (Art. 5 Abs. 1 Bst. c):

  - Schlechterstellung oder Benachteiligung in sozialen Zusammenhängen, die in keinem Zusammenhang zu den Umständen stehen, unter denen die Daten ursprünglich erzeugt oder erhoben wurden;

  - Schlechterstellung oder Benachteiligung in einer Weise, die im Hinblick auf ihr soziales Verhalten oder dessen Tragweite ungerechtfertigt oder unverhältnismässig ist.

- «Predictive Policing»: Verboten sind KI-Systeme, um das Risiko, dass eine natürliche Person eine Straftat begeht, ausschliesslich auf der Grundlage ihres Profilings oder der Bewertung ihrer persönlichen Merkmale und Eigenschaften zu beurteilen oder vorherzusagen. Das Verbot gilt nicht, wenn das System dazu verwendet wird, eine durch Menschen durchgeführte Bewertung zu unterstützen (Art. 5 Abs. 1 Bst. d).

---

[227] Für kritische Erwägungen s. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 18 N 42 ff.

- Massenhafte Sammlung von Bildern zum Zweck der Gesichtserkennung: Verboten sind KI-Systeme, die Datenbanken zur Gesichtserkennung durch das ungezielte Auslesen von Gesichtsbildern aus dem Internet oder von Überwachungsaufnahmen erstellen oder erweitern (Art. 5 Abs. 1 Bst. e).

- Erkennung von Emotionen: Verboten sind KI-Systeme, mit denen Emotionen einer Person am Arbeitsplatz und in Bildungseinrichtungen abgeleitet werden können, es sei denn, die Verwendung des KI-Systems erfolgt aus medizinischen Gründen oder Sicherheitsgründen (Art. 5 Abs. 1 Bst. f).

- Systeme zur biometrischen Kategorisierung: Systeme, mit denen Personen auf der Grundlage ihrer biometrischen Daten, wie des Gesichts oder der Fingerabdrücke, kategorisiert werden, um ihre politischen Einstellungen, ihre Gewerkschaftszugehörigkeit, ihre religiösen oder weltanschaulichen Überzeugungen, ihre Rasse, ihr Sexualleben oder ihre sexuelle Ausrichtung abzuleiten, sind verboten (Art. 5 Abs. 1 Bst. g).

- Biometrische Echtzeit-Fernidentifizierung in öffentlich zugänglichen Räumen zu Strafverfolgungszwecken: Ein solches Vorgehen ist grundsätzlich verboten (Art. 5 Abs. 1 Bst. h). Der Begriff «Echtzeit» bedeutet, dass die Erfassung biometrischer Daten, der Abgleich und die Identifizierung ohne erhebliche Verzögerung erfolgen (vgl. Art. 3 Nr. 42).

  Die Verwendung dieser Technik wird als besonders starker Eingriff in die Rechte und Freiheiten der Personen betrachtet, da sie die Privatsphäre der Bevölkerung beeinträchtigen, ein Gefühl der ständigen Überwachung wecken und indirekt von der Ausübung der Versammlungsfreiheit und anderer Grundrechte abhalten kann. Zudem können die technischen Ungenauigkeiten von KI-Systemen, die für die biometrische Fernidentifizierung natürlicher Personen bestimmt sind, zu verzerrten Ergebnissen führen und eine diskriminierende Wirkung haben (E. 32 der Verordnung).

  Die Verwendung dieser Systeme ist nur in drei Situationen zulässig (vgl. Art. 5 Abs. 1 Bst. h): i) gezielte Suche nach bestimmten Opfern von Entführung, Menschenhandel oder sexueller Ausbeutung sowie die Suche nach vermissten Personen; ii) Abwenden einer konkreten, erheblichen und unmittelbaren Gefahr für das Leben oder die Sicherheit natürlicher Personen oder einer tatsächlichen und bestehenden oder tatsächlichen und vorhersehbaren Gefahr eines Terroranschlags; iii) Aufspüren oder Identifizieren von Personen, die beschuldigt oder verdächtigt werden, eine in Anhang II aufgeführte Straftat begangen zu haben, die im betreffenden Mitgliedstaat mit einer Freiheitsstrafe oder einer Freiheit beschränkenden sichernden Massnahme im Höchstmass von mindestens vier Jahren bedroht ist.

  Artikel 5 Absätze 2 und 3 legt die Voraussetzungen fest, unter denen die biometrische Fernidentifizierung in diesen ausserordentlichen Situationen eingesetzt werden darf. Insbesondere muss jede Verwendung von einer Justizbehörde oder einer unabhängigen Verwaltungsbehörde genehmigt werden. Eine solche Genehmigung muss grundsätzlich vor der Verwendung des Systems zur Identifizierung einer Person oder mehrerer Personen eingeholt werden, ausser in begründeten dringenden Fällen. In diesen Fällen ist es möglich, das System ohne Genehmigung zu verwenden, sofern diese unverzüglich, spätestens jedoch innerhalb von 24 Stunden beantragt wird. Die Genehmigung wird nur gewährt, wenn die Verwendung für das Erreichen eines der in Artikel 5

Absatz 1 Buchstabe h genannten Ziele notwendig und verhältnismässig ist und insbesondere wenn sie auf das unbedingt erforderliche Mass beschränkt bleibt.

Zudem wird die Verwendung dieser Systeme – ausser in Ausnahmefällen – nur gestattet, wenn die Strafverfolgungsbehörde eine Folgenabschätzung im Hinblick auf die Grundrechte gemäss Artikel 27 der Verordnung durchgeführt und das System gemäss Artikel 49 der Verordnung in der Datenbank registriert hat (vgl. Art. 5 Abs. 2 *in fine*) (vgl. Ziff. 5.2.7.3.3).

Die nationale Marktüberwachungsbehörde (vgl. Ziff. 5.2.14) und die nationale Datenschutzbehörde müssen über jeden Einsatz eines biometrischen Echtzeit-Fernidentifizierungssystems informiert werden und der Europäischen Kommission jährlich einen Bericht über die Verwendung dieser Systeme vorlegen. Auf der Grundlage der Berichte der Mitgliedstaaten veröffentlicht die Europäische Kommission jährlich einen Gesamtbericht (Art. 5 Abs. 6 und 7).

Innerhalb der in den Artikeln 5 Absatz 1 Buchstabe h und 5 Absätze 2 und 3 aufgeführten Grenzen können die Mitgliedstaaten in ihrem nationalen Recht detaillierte Vorschriften, namentlich für die Beantragung und Erteilung der Genehmigungen, vorsehen (vgl. Art. 5 Abs. 5).

5.2.7 Hochrisiko-Systeme der künstlichen Intelligenz

5.2.7.1 Einstufung

Gemäss der KI-Verordnung dürfen Hochrisiko-KI-Systeme nur in Verkehr gebracht, in Betrieb genommen oder verwendet werden, wenn sie bestimmte Bedingungen erfüllen.

Die Verordnung sieht zwei Kategorien von Hochrisiko-KI-Systemen vor:

- KI-Systeme, die als Sicherheitsbauteil eines Produkts (z. B. Spielzeug, Medizinprodukte) verwendet werden, das unter die in Anhang I aufgeführten Harmonisierungsrechtsvorschriften der EU fällt, sowie KI-Systeme, die selber solche Produkte sind, sofern diese Produkte einer Konformitätsbewertung durch Dritte im Hinblick auf ihr Inverkehrbringen oder ihre Inbetriebnahme gemäss den Harmonisierungsrechtsvorschriften der EU unterzogen werden müssen (Art. 6 Abs. 1).

- Die in Anhang III genannten KI-Systeme (Art. 6 Abs. 2), also KI-Systeme, die in acht spezifischen Bereichen eingesetzt werden:[228]

[228] Für kritische Erwägungen s. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 22 N 51 ff.

- Biometrie[229];
- kritische Infrastruktur;
- allgemeine und berufliche Bildung;
- Beschäftigung, Personalmanagement und Zugang zur Selbstständigkeit;
- Zugänglichkeit und Inanspruchnahme grundlegender privater und grundlegender öffentlicher Dienste und Leistungen;
- Strafverfolgung;
- Migration, Asyl und Grenzkontrolle;
- Rechtspflege und demokratische Prozesse.

Die Europäische Kommission ist befugt, diese Liste zu ändern (vgl. Art. 7).

Allerdings gilt ein KI-System, obwohl es in Anhang III genannt ist, in bestimmten Fällen, wenn es die Entscheidungsfindung nicht wesentlich beeinflusst oder die geschützten rechtlichen Interessen nicht beeinträchtigt, nicht als hochriskant (vgl. Art. 6 Abs. 3 und E. 53):

- Das KI-System ist dazu bestimmt, eine eng gefasste Verfahrensaufgabe durchzuführen. Dabei kann es sich beispielsweise um ein System handeln, das unstrukturierte Daten in strukturierte Daten umwandelt, oder ein KI-System, das zur Erkennung von Duplikaten unter einer grossen Zahl von Anwendungen eingesetzt wird (E. 53 der Verordnung).

- Das KI-System ist dazu bestimmt, das Ergebnis einer zuvor abgeschlossenen menschlichen Tätigkeit zu verbessern, beispielsweise wenn ein KI-System zum Ziel hat, die in zuvor verfassten Dokumenten verwendete Sprache zu verbessern (E. 53 der Verordnung).

- Das KI-System ist dazu bestimmt, Entscheidungsmuster oder Abweichungen von früheren Entscheidungsmustern zu erkennen, und ist nicht dazu gedacht, die zuvor abgeschlossene menschliche Bewertung ohne eine angemessene menschliche Überprüfung zu ersetzen oder zu beeinflussen. Das Risiko wird als geringer betrachtet, da die Verwendung des KI-Systems einer zuvor abgeschlossenen menschlichen Bewertung folgt, die das KI-System ohne angemessene menschliche Überprüfung nicht ersetzen oder beeinflussen soll (E. 53).

[229] Diese Kategorie umfasst keine KI-Systeme, die bestimmungsgemäss für die biometrische Verifizierung, wozu die Authentifizierung gehört, verwendet werden, deren einziger Zweck darin besteht zu bestätigen, dass eine bestimmte natürliche Person die Person ist, für die sie sich ausgibt, und die Identität einer natürlichen Person zu dem alleinigen Zweck zu bestätigen, Zugang zu einem Dienst zu erhalten, ein Gerät zu entriegeln oder Sicherheitszugang zu Räumlichkeiten zu erhalten. Biometrische Systeme, die ausschliesslich dazu bestimmt sind, Massnahmen zur Cybersicherheit und zum Schutz personenbezogener Daten durchführen zu können, sollten nicht als Hochrisiko-KI-Systeme gelten (vgl. Anhang III Nr. 1 Bst. a und E. 54).

- Das KI-System ist dazu bestimmt, eine Aufgabe auszuführen, die eine Bewertung, die für die Zwecke der in Anhang III aufgeführten Anwendungsfälle relevant ist, lediglich vorbereitet (E. 53).

Ungeachtet dieser vier Fälle gilt ein in Anhang III genanntes KI-System immer als hochriskant, wenn es ein Profiling natürlicher Personen vornimmt (vgl. Art. 6 Abs. 3 *in fine*).

Um die Nachvollziehbarkeit und Transparenz der Systeme zu gewährleisten, die gemäss den oben genannten Kriterien ausgenommen sind, muss ein Anbieter, der der Auffassung ist, dass von einem in Anhang III aufgeführten KI-System kein hohes Risiko ausgeht, seine Bewertung dokumentieren, bevor dieses System in Verkehr gebracht oder in Betrieb genommen wird. Zudem ist dieser Anbieter verpflichtet, das System in der EU-Datenbank zu registrieren. Er muss diese Dokumentation den zuständigen nationalen Behörden auf Anfrage zur Verfügung stellen (Art. 6 Abs. 4).

Die Europäische Kommission muss nach Konsultation des Europäischen Gremiums für Künstliche Intelligenz weitere Orientierungshilfen für die praktische Umsetzung von Artikel 6 der Verordnung bereitstellen (Ziff. 5.2.13). Sie erarbeitet Leitlinien für die Umsetzung und stellt eine umfassende Liste von Anwendungsfällen für KI-Systeme bereit, die hochriskant oder nicht hochriskant sind (vgl. Art. 6 Abs. 5).

5.2.7.2 Anforderungen an Hochrisiko-KI-Systeme

Hochrisiko-KI-Systeme müssen die in Abschnitt 2 des Kapitels III festgelegten Anforderungen, also die Artikel 8 bis 15, erfüllen. Auf die Frage, welche Pflichten im Zusammenhang mit der Einhaltung der Grundanforderungen für welche Akteure (Anbieter, Betreiber oder andere) gelten, wird in den Artikeln 16 ff. der Verordnung eingegangen (vgl. Ziff. 5.2.7.3).

Es müssen folgende Anforderungen eingehalten werden:

- Einrichtung eines Risikomanagementsystems (Art. 9): Das Risikomanagementsystem hat zum Ziel, die bekannten und vernünftigerweise vorhersehbaren Risiken zu ermitteln und zu mindern, die von einem Hochrisiko-KI-System für die Gesundheit, die Sicherheit und die Grundrechte ausgehen können, wenn es entsprechend seiner Zweckbestimmung verwendet wird (Art. 9 Abs. 2 Bst. a). Weiter dient es der Abschätzung und Bewertung der Risiken, die bei einer vernünftigerweise vorhersehbaren Fehlanwendung entstehen können (Art. 9 Abs. 2 Bst. b). Es handelt sich um einen kontinuierlichen iterativen Prozess, der während des gesamten Lebenszyklus eines Hochrisiko-KI-Systems geplant und durchgeführt wird. Das Risikomanagementsystem umfasst auch das Ergreifen geeigneter und gezielter Risikomanagementmassnahmen zur Bewältigung der ermittelten Risiken (Art. 9 Abs. 2 Bst. d).[230]

- Daten und Daten-Governance (Art. 10): Für Hochrisiko-KI-Systeme, in denen Techniken eingesetzt werden, mit denen KI-Modelle mit Daten trainiert werden, gelten im Bereich Daten und Daten-Governance besondere Bedingungen. Insbesondere müssen

[230] Für eine Vertiefung s. MARTINA ARIOLI, Risikomanagement (Fn. 222), 6 N 19 ff.

die Trainings-, Validierungs- und Testdatensätze den Anforderungen von Artikel 10 Absätze 2 bis 5 der Verordnung entsprechen.

Die verwendeten Daten müssen namentlich hochwertig sein, um sicherzustellen, dass das System bestimmungsgemäss funktioniert und nicht zur Ursache für Diskriminierung wird. Dies erfordert, dass alle verwendeten Daten im Hinblick auf die Zweckbestimmung des Systems relevant, hinreichend repräsentativ und so weit wie möglich fehlerfrei und vollständig sind. Alle Daten müssen zudem geeignete statistische Merkmale haben, insbesondere bezüglich der Personen oder Personengruppen, auf die das KI-System bestimmungsgemäss angewandt werden soll, und der Minderung möglicher Verzerrungen in allen Daten besondere Beachtung beimessen. Zudem müssen Massnahmen vorgesehen werden, mit denen sichergestellt wird, dass die verarbeiteten Daten geschützt und Gegenstand angemessener Sicherheitsvorkehrungen sind. Für personenbezogene Daten, einschliesslich sensibler Daten, sind ausserdem besondere Vorschriften vorgesehen.

- Technische Dokumentation (Art. 11): Um die Nachvollziehbarkeit der Hochrisiko-KI-Systeme, die Überprüfung der Einhaltung der Anforderungen der Verordnung sowie die Kontrolle ihres Betriebs und die Beobachtung nach dem Inverkehrbringen zu ermöglichen, braucht es umfassende Informationen darüber, wie diese Systeme entwickelt wurden und wie sie während ihrer gesamten Lebensdauer funktionieren. Artikel 11 sieht daher die Pflicht vor, eine technische Dokumentation in klarer und verständlicher Form bereitzustellen, mit der die Konformität des KI-Systems beurteilt werden kann und die zumindest die in Anhang IV genannten Angaben enthält. Dabei handelt es sich namentlich um allgemeine Merkmale, die Fähigkeiten und Grenzen des Systems, die Algorithmen, die Daten, die Trainingsmethoden, die Tests und die verwendeten Validierungsverfahren sowie die Dokumentation des Risikomanagementsystems.

- Aufzeichnungspflichten (Art. 12): Die Technik der Hochrisiko-KI-Systeme muss die automatische Aufzeichnung von Ereignissen (Protokollierung) während des gesamten Lebenszyklus des Systems ermöglichen. Eines der Ziele besteht darin, die Beobachtung nach dem Inverkehrbringen gemäss Artikel 72 zu erleichtern und den Betrieb des Systems gemäss Artikel 26 Absatz 5 der Verordnung zu überwachen (vgl. Art. 12 Abs. 2).

- Transparenz und Bereitstellung von Informationen für die Betreiber (Art. 13): Hochrisiko-KI-Systeme müssen so konzipiert und entwickelt werden, dass die Betreiber die Ausgaben eines Systems angemessen interpretieren und verwenden können. Sie müssen mit geeigneten Informationen in Form einer Betriebsanleitung bereitgestellt werden.

- Menschliche Aufsicht (Art. 14): Hochrisiko-KI-Systeme müssen so konzipiert und entwickelt werden, dass sie während der Dauer ihrer Verwendung von natürlichen Personen wirksam beaufsichtigt werden können. Diese wirksame Aufsicht dient der Verhinderung oder Minimierung der Risiken für Gesundheit, Sicherheit oder Grundrechte, die entstehen können, wenn ein Hochrisiko-KI-System im Einklang mit seiner Zweckbestimmung oder im Rahmen einer vernünftigerweise vorhersehbaren Fehlanwendung verwendet wird. Es müssen den Risiken angemessene Massnahmen ergriffen werden. Sie können vom Anbieter direkt in das System eingebaut werden oder von diesem bestimmt und

vom Betreiber umgesetzt werden (vgl. Art. 14 Abs. 3). Dabei kann es sich beispielsweise um die Möglichkeit handeln, den Betrieb eines KI-Systems mit einer Stopptaste zu unterbrechen (vgl. Art. 14 Abs. 4 Bst. e).

Die den in Anhang III Nummer 1 Buchstabe a genannten Systeme unterliegen einer erweiterten menschlichen Aufsicht. So darf der Betreiber keine Massnahmen oder Entscheidungen allein aufgrund des vom System hervorgebrachten Identifizierungsergebnisses treffen, solange die Identifizierung nicht von mindestens *zwei* natürlichen Personen, die über die notwendige Kompetenz, Ausbildung und Befugnis verfügen, getrennt überprüft und bestätigt wurde. Diese Pflicht gilt jedoch nicht für Hochrisiko-KI-Systeme, die in den Bereichen Strafverfolgung, Migration, Grenzkontrolle oder Asyl verwendet werden, wenn die Umsetzung dieser Anforderung nach EU-Recht oder nationalem Recht unverhältnismässig wäre (Art. 14 Abs. 5 *in fine*).

- Genauigkeit, Robustheit und Cybersicherheit (Art. 15): Es müssen technische Massnahmen ergriffen werden, um ein angemessenes Mass an Genauigkeit, Robustheit und Cybersicherheit der Hochrisiko-KI-Systeme zu gewährleisten.

5.2.7.3 Pflichten der Anbieter und anderer Beteiligter

5.2.7.3.1 Pflichten der Anbieter

Die Verordnung legt den Schwerpunkt auf die zentrale Rolle der Anbieter, denen sie die Verantwortung für die Konformität der Hochrisiko-KI-Systeme überträgt.[231]

Die meisten Pflichten im Zusammenhang mit Hochrisiko-KI-Systemen entfallen daher auf die Anbieter von solchen Systemen. Dieses Unterkapitel konzentriert sich auf die wichtigsten Pflichten:

- Gemäss Artikel 16 Buchstabe a stellen die Anbieter sicher, dass ihre Systeme die in Kapitel III Abschnitt 2 (Art. 8–15 der Verordnung, vgl. Ziff. 5.2.7.2) festgelegten Anforderungen erfüllen. Auf begründete Anfrage einer zuständigen nationalen Behörde haben sie nachzuweisen, dass das KI-System diese Anforderungen erfüllt (vgl. Art. 16 Bst. k).

- Die Anbieter müssen identifizierbar sein (Art. 16 Bst. b).

- Gemäss Artikel 16 Buchstabe c und Artikel 17 müssen die Anbieter von Hochrisiko-KI-Systemen über ein *Qualitätsmanagementsystem* verfügen, das das in Artikel 9 genannte Risikomanagementsystem und verschiedene andere Aspekte umfasst, wie ein Konzept zur Einhaltung der Regulierungsvorschriften, Systeme für die Aufzeichnung der Dokumentation, Ressourcenmanagement, einen internen Rechenschaftsrahmen

[231] Für eine Kritik zu diesem Ansatz vgl. ANGELA MÜLLER, Der Artificial intelligence Act der EU (Fn. 223), 18.

usw. Wenn in der sektorspezifischen Gesetzgebung bereits ein Qualitätsmanagementmechanismus vorgesehen ist, können diese Aspekte in das Qualitätsmanagementsystem gemäss dieser Gesetzgebung integriert werden (vgl. Art. 17 Abs. 3).

- Die Artikel 18 und 19 sehen Pflichten für die Aufbewahrung der Dokumentation und der automatisch erzeugten Protokolle vor.

- Artikel 20 sieht die Pflicht vor, bei Verdacht, dass ein in Verkehr gebrachtes KI-System nicht der Verordnung entspricht, Massnahmen zu ergreifen.

  Die Anbieter von Hochrisiko-KI-Systemen müssen auch über ein «System zur Beobachtung nach dem Inverkehrbringen» verfügen (vgl. Art. 3 Nrn. 25 und 72), damit den möglichen Risiken im Zusammenhang mit KI-Systemen, die nach dem Inverkehrbringen oder der Inbetriebnahme «dazulernen», effizient begegnet werden kann (E. 155).

- Gemäss Artikel 21 ist der Anbieter zudem verpflichtet, mit den zuständigen Behörden zusammenzuarbeiten.

- Die Anbieter müssen sicherstellen, dass das Hochrisiko-KI-System dem Konformitätsbewertungsverfahren gemäss Artikel 43 unterzogen wird, bevor es in Verkehr gebracht oder in Betrieb genommen wird (Art. 16 Bst. f). Weiter haben sie eine Konformitätserklärung gemäss Artikel 47 auszustellen, am System eine CE-Kennzeichnung («Conformité Européenne»-Kennzeichnung) anzubringen (Art. 48) und das System ordnungsgemäss zu registrieren (vgl. Art. 49).

  Für diese Aspekte wird auf die Ausführungen unten verwiesen (vgl. Ziff. 5.2.7.4).

Liegen besondere Umstände vor, so gehen die in den Artikeln 16 ff. vorgesehenen Pflichten auf den Händler, Einführer oder Betreiber oder auf sonstige Dritte über. Dies ist beispielsweise der Fall, wenn einer dieser Akteure die Zweckbestimmung eines KI-Systems so verändert, dass dieses zu einem Hochrisiko-KI-System wird (z. B. ein System in der Art von ChatGPT wird für die Analyse der Lebensläufe von Bewerberinnen und Bewerbern verwendet).[232] In diesen Fällen gilt der «ursprüngliche» Anbieter nicht mehr als Anbieter im Sinne der Verordnung. Für die anderen Fälle wird auf Artikel 25 der Verordnung verwiesen.

5.2.7.3.2 Pflichten der Einführer und der Händler

Die Einführer müssen sicherstellen, dass ein Hochrisiko-KI-System der Verordnung entspricht, bevor sie es in Verkehr bringen (vgl. Art. 23). Sie müssen beispielsweise überprüfen, ob der Anbieter das entsprechende Konformitätsbewertungsverfahren gemäss Artikel 43 durchgeführt und die technische Dokumentation gemäss Artikel 11 und Anhang IV erstellt hat. Weiter müssen sie ihren Namen und die Anschrift angeben, unter der sie kontaktiert werden können (vgl. Art. 23 Abs. 3).

---

[232] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 10 N 25.

Die Händler ihrerseits müssen sicherstellen, dass das System mit der erforderlichen CE-Kennzeichnung versehen ist, ihm eine Kopie der EU-Konformitätserklärung und der Betriebsanleitung beigefügt ist und der Anbieter und gegebenenfalls der Einführer dieses Systems ihre in Artikel 16 Buchstaben b und c sowie in Artikel 23 Absatz 3 festgelegten jeweiligen Pflichten erfüllt haben (vgl. Art. 24).

5.2.7.3.3 Pflichten der Betreiber

Gemäss der Verordnung sind die Betreiber nach den Anbietern die zweite Gruppe von Akteuren, die im Zusammenhang mit den Hochrisiko-KI-Systemen am meisten Pflichten haben.

Die Verordnung legt zunächst die allgemeinen Pflichten fest. Dazu zählen insbesondere:

- Die Pflicht, geeignete technische und organisatorische Massnahmen zu treffen, um eine vorschriftsgemässe Verwendung der Hochrisiko-KI-Systeme sicherzustellen (Art. 26 Abs. 1).

- Die Betreiber müssen zudem gewährleisten, dass die menschliche Aufsicht natürlichen Personen, die über die erforderliche Kompetenz, Ausbildung und Befugnis verfügen, übertragen wird (Art. 26 Abs. 2).

- Wenn der Betreiber die Möglichkeit hat, die Eingabedaten zu kontrollieren, muss er sicherstellen, dass diese der Zweckbestimmung des Hochrisiko-KI-Systems entsprechen und ausreichend repräsentativ sind (Art. 26 Abs. 4).

- Die Betreiber sind verpflichtet, den Betrieb des Hochrisiko-KI-Systems zu überwachen. Sie haben eine Informationspflicht, wenn ein Risiko im Sinne von Artikel 79 vorliegt oder ein schwerwiegender Vorfall festgestellt wurde (Art. 26 Abs. 5).

- Sie bewahren die von ihrem Hochrisiko-KI-System automatisch erzeugten Protokolle auf, soweit diese Protokolle ihrer Kontrolle unterliegen (Art. 26 Abs. 6).

Anschliessend legt die Verordnung spezifische Pflichten je nach Betreiber fest, namentlich:

- Betreiber, die Arbeitgeber sind, informieren die Arbeitnehmervertreterinnen und -vertreter und die betroffenen Arbeitnehmenden darüber, dass sie der Verwendung eines Hochrisiko-KI-Systems unterliegen werden, und zwar vor seiner Inbetriebnahme oder Verwendung (Art. 26 Abs. 7).

- Betreiber, bei denen es sich um Behörden oder Organe, Einrichtungen oder sonstige Stellen der EU handelt, müssen den Registrierungspflichten gemäss Artikel 49 nachkommen (Art. 26 Abs. 8). Sie müssen insbesondere die Verwendung der in Anhang III aufgeführten Systeme registrieren, mit Ausnahme der unter Nummer 2 genannten (die auf nationaler Ebene registriert werden) (vgl. Art. 49 Abs. 3 und 5). Die Registrierung der in Anhang III Nummern 1, 6 und 7 genannten Hochrisiko-KI-Systeme erfolgt in den Bereichen Strafverfolgung, Migration, Asyl und Grenzkontrolle in einem nicht öffentlichen Teil der Datenbank, auf den nur die Europäische Kommission und bestimmte nationale Behörden Zugriff haben (Art. 49 Abs. 4).

- Betreiber eines Hochrisiko-KI-Systems zur nachträglichen[233] biometrischen Fernidentifizierung im Rahmen von Ermittlungen zur gezielten Suche einer Person, die der Begehung einer Straftat verdächtigt wird oder aufgrund einer solchen verurteilt wurde, müssen vorab oder unverzüglich, spätestens jedoch innerhalb von 48 Stunden bei einer Justizbehörde oder einer Verwaltungsbehörde eine Genehmigung für die Verwendung des Systems beantragen (vgl. Art. 26 Abs. 10).

  Weiter legt die Verordnung fest, dass diese Systeme in keinem Fall zu Strafverfolgungszwecken in nicht zielgerichteter Weise und ohne jeglichen Zusammenhang mit einer Straftat, einem Strafverfahren, einer tatsächlichen und bestehenden oder tatsächlichen und vorhersehbaren Gefahr einer Straftat oder der Suche nach einer bestimmten vermissten Person verwendet werden dürfen. Ausserdem muss sichergestellt werden, dass die Strafverfolgungsbehörden keine ausschliesslich auf der Grundlage der Ausgabe solcher Systeme beruhende Entscheidung treffen, aus der sich eine nachteilige Rechtsfolge für eine Person ergibt.

  Keine Genehmigung ist erforderlich, wenn das System zur erstmaligen Identifizierung einer potenziell verdächtigen Person auf der Grundlage objektiver und nachprüfbarer Tatsachen, die in unmittelbarem Zusammenhang mit der Straftat stehen, verwendet wird.

  Ferner müssen die Betreiber dieser Systeme den zuständigen Marktüberwachungsbehörden und den nationalen Datenschutzbehörden Jahresberichte über ihre Verwendung vorlegen, wovon die Offenlegung sensibler operativer Daten im Zusammenhang mit der Strafverfolgung ausgenommen ist (Art. 26 Abs. 10).

- Wenn der Betreiber ein in Anhang III aufgeführtes Hochrisiko-KI-System verwendet, um Entscheidungen betreffend natürliche Personen zu treffen oder solche Entscheidungen zu unterstützen, muss er die betroffenen Personen informieren (Art. 26 Abs. 11). Diese Pflicht gilt unbeschadet von Artikel 50 (Transparenzpflichten, vgl. Ziff. 5.2.8). Auf die zu Strafverfolgungszwecken verwendeten Systeme ist Artikel 13 der Richtlinie (EU) 2016/680[234] anwendbar.

  Zusätzlich hat die betroffene Person das Recht, eine Erläuterung im Zusammenhang mit automatisierten Einzelentscheidungen zu erhalten (vgl. Art. 86, Ziff. 5.2.15).

---

[233] Der Begriff «nachträglich» bedeutet, dass es eine zeitliche Verzögerung zwischen der Erfassung biometrischer Daten, dem Abgleich und der Identifizierung gibt (vgl. Art. 3 Nrn. 42 und 43).

[234] Richtlinie (EU) 2016/680 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates, ABl. L 119/89, 4. Mai 2016.

- Schliesslich haben bestimmte Betreiber die Pflicht, eine Abschätzung der Folgen von Hochrisiko-KI-Systemen auf die Grundrechte durchzuführen.[235] Diese Pflicht ist in Artikel 27 geregelt:

  - Adressaten der Pflicht: Eine Folgenabschätzung durchführen müssen 1) Einrichtungen des öffentlichen Rechts oder private Einrichtungen, die öffentliche Dienste[236] erbringen, 2) (öffentliche oder private) Betreiber von Hochrisiko-KI-Systemen gemäss Anhang III Nummer 5 Buchstaben b und c[237].

  - Anwendungsbereich: Die Grundrechte-Folgenabschätzung muss vor der Inbetriebnahme der Hochrisiko-KI-Systeme gemäss Artikel 6 Absatz 2 der Verordnung, also der in Anhang III genannten Systeme, durchgeführt werden. Von dieser Pflicht ausgenommen sind KI-Systeme, die im Bereich kritischer Infrastruktur verwendet werden (vgl. Anhang III Nummer 2).

  - Inhalt: Die Grundrechte-Folgenabschätzung muss dem Betreiber ermöglichen, die spezifischen Risiken für die Grundrechte des Einzelnen, die wahrscheinlich berührt sein werden, zu ermitteln und die Massnahmen festzulegen, die zu ergreifen sind, wenn diese Risiken eintreten (vgl. Art. 27 Abs. 1 Bst. a–f). Um die Betreiber bei der Durchführung der Abschätzung zu unterstützen, erarbeitet das Büro für Künstliche Intelligenz ein Muster für einen Fragebogen, auch mithilfe eines automatisierten Instruments (Art. 27 Abs. 5).

    Wenn gewisse Aspekte der Abschätzung bereits durch die Datenschutz-Folgenabschätzung (Art. 35 DSGVO oder Art. 27 der Richtlinie (EU) 2016/680[238]) abgedeckt sind, kann die Grundrechte-Folgenabschätzung die Datenschutz-Folgenabschätzung ergänzen (vgl. 27 Abs. 4).

---

[235] Zum Unterschied zu den Risikomanagementsystemen (Art. 9) vgl. MARTINA ARIOLI, Risikomanagement (Fn. 222), 14 N 48 f.

[236] Es ist nicht klar, ob der Begriff «öffentliche Dienste» auch rein privatrechtliche Tätigkeiten umfasst, aber in Anbetracht der Erwägungen scheint dies der Fall zu sein (vgl. E. 96). Erwähnt sind die im öffentlichen Interesse erbrachten Dienste, namentlich in den Bereichen Bildung, Gesundheitsversorgung, Sozialdienste, Wohnungswesen und Justizverwaltung, vgl. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 30 N 58.

[237] Hier handelt es sich um Betreiber von KI-Systemen, die bestimmungsgemäss für die Kreditwürdigkeitsprüfung und Bonitätsbewertung natürlicher Personen verwendet werden, mit Ausnahme von KI-Systemen, die zur Aufdeckung von Finanzbetrug verwendet werden (Anhang III Nr. 5 Bst. b); Betreiber von KI-Systemen, die bestimmungsgemäss für die Risikobewertung und Preisbildung in Bezug auf natürliche Personen im Fall von Lebens- und Krankenversicherungen verwendet werden (Anhang III Nr. 5 Bst. c).

[238] Richtlinie (EU) 2016/680 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten im Bereich der Strafverfolgung (Fn. 234).

- Überwachung: Wurde die Abschätzung durchgeführt, so teilt der Betreiber die Ergebnisse der Marktüberwachungsbehörde mit (vgl. Ziff. 5.2.14). In bestimmten Fällen sind Ausnahmen möglich (vgl. Art. 27 Abs. 3).

5.2.7.4 Konformitätsbewertung

5.2.7.4.1 Verfahren

Gemäss Artikel 16 Buchstabe f der Verordnung stellen die Anbieter von Hochrisiko-KI-Systemen sicher, dass diese Systeme dem Konformitätsbewertungsverfahren gemäss Artikel 43 unterzogen werden, bevor sie in Verkehr gebracht oder in Betrieb genommen werden.

Artikel 43 regelt die Konformitätsbewertung, wobei je nach System unterschiedliche Verfahren zur Anwendung gelangen:[239]

- Bei den Hochrisiko-KI-Systemen, die unter die Harmonisierungsrechtsvorschriften der EU fallen (vgl. Anhang I Abschnitt A), wird die Einhaltung der Anforderungen an Hochrisiko-KI-Systeme (Art. 8 ff., vgl. Ziff. 5.2.7.2) in das in diesen Vorschriften vorgesehene Konformitätsbewertungsverfahren einbezogen, um den Aufwand für die Akteure zu verringern und Doppelspurigkeiten zu vermeiden. Die Stellen, die gemäss diesen sektorspezifischen Vorschriften notifiziert wurden, sind berechtigt, die Konformität von Hochrisiko-KI-Systemen zu bewerten (vgl. Art. 43 Abs. 3).

  Die Anwendbarkeit der Anforderungen der KI-Verordnung muss die spezifische Logik, die Methodik oder die allgemeine Struktur der in den sektorspezifischen Rechtsvorschriften vorgesehenen Konformitätsbewertung unberührt lassen.

- Bei Hochrisiko-KI-Systemen, die nicht unter eine Harmonisierungsregelung (vgl. Anhang III) fallen, sieht die KI-Verordnung zwei Verfahrensarten für die Konformitätsbewertung vor: eine Konformitätsbewertung *auf der Grundlage einer internen Kontrolle* (gemäss Anhang VI) und in bestimmten Fällen eine Konformitätsbewertung *durch eine Drittpartei* (notifizierte Stelle[240]) (gemäss Anhang VII).

  Dieses Vorgehen soll dazu beitragen, bei den Anbietern, die zurzeit keinen Pflichten im Zusammenhang mit den Produkten unterstehen, eine Konformitätskultur zu schaffen

---

[239] Die Verfahren basieren auf dem neuen Rechtsrahmen («New Legislative Framework», NLF), s. dazu MARTINA ARIOLI, Risikomanagement (Fn. 222), 4 N 8 f.

[240] Damit KI-Systeme, falls vorgeschrieben, Konformitätsbewertungen durch Dritte unterzogen werden können, müssen die notifizierten Stellen gemäss den Rechtsvorschriften von den zuständigen nationalen Behörden notifiziert werden, sofern sie eine Reihe von Anforderungen erfüllen, insbesondere in Bezug auf Unabhängigkeit, Kompetenz, Nichtvorliegen von Interessenkonflikten und geeignete Anforderungen an die Cybersicherheit. Die zuständigen nationalen Behörden müssen die Notifizierung dieser Stellen der Europäischen Kommission und den anderen Mitgliedstaaten übermitteln (vgl. Art. 28 ff. der Verordnung).

und zu stärken, ohne dass von Anfang an langwierige Verfahren vorgeschrieben werden, die die Innovation unverhältnismässig hemmen könnten.

Daher muss die Konformitätsbewertung in der Regel vom Anbieter in eigener Verantwortung durchgeführt werden («self-assessment»). In diesem Sinn sieht Artikel 43 Absatz 2 der Verordnung vor, dass bei den in Anhang III Nummern 2 bis 8 aufgeführten Hochrisiko-KI-Systemen ein Konformitätsbewertungsverfahren auf der Grundlage einer internen Kontrolle durchgeführt werden muss, das keine Beteiligung einer notifizierten Stelle vorsieht.

Wenn erforderlich, namentlich zur Vermeidung und Minimierung von Risiken für die Sicherheit und die Grundrechte, ist die Europäische Kommission befugt, einzugreifen und diese Systeme der Konformitätsbewertung oder Teilen davon durch eine notifizierte Stelle zu unterstellen (Art. 43 Abs. 6).

Bei KI-Systemen, die bestimmungsgemäss zu biometrischen Zwecken verwendet werden (vgl. Anhang III Nr. 1), sieht Artikel 43 Absatz 1 der Verordnung einen detaillierteren Mechanismus vor.

In beiden Fällen misst die KI-Verordnung den harmonisierten Normen grosse Bedeutung bei (für diesen Punkt wird auf Ziff. 5.2.11 verwiesen).

In Artikel 46 der Verordnung sind die Voraussetzungen geregelt, unter denen die Marktüberwachungsbehörde eine Genehmigung erteilt, mit der Hochrisiko-KI-Systeme in Abweichung vom Konformitätsbewertungsverfahren in Verkehr gebracht oder in Betrieb genommen werden dürfen. Ausnahmen sind beispielsweise möglich aus aussergewöhnlichen Gründen der öffentlichen Sicherheit. Für Hochrisiko-KI-Systeme im Zusammenhang mit Produkten, die den in Anhang I Abschnitt A aufgeführten Harmonisierungsrechtsvorschriften der EU unterstehen, gelten nur die in diesen Vorschriften festgelegten Ausnahmen von den Konformitätsbewertungsverfahren (vgl. Art. 46 Abs. 7).

5.2.7.4.2 Konformitätserklärung, CE-Kennzeichnung und Registrierung

Wurde im Rahmen einer internen Kontrolle oder einer Bewertung durch eine notifizierte Stelle festgestellt, dass das Hochrisiko-KI-System den Anforderungen der Verordnung entspricht und wurde eine Konformitätsbescheinigung ausgestellt (vgl. Art. 44), so muss der Anbieter eine schriftliche EU-Konformitätserklärung erstellen, die die in Anhang V vorgesehenen Informationen enthält. Diese bestätigt, dass das Hochrisiko-KI-System die Anforderungen der Verordnung erfüllt. Mit dieser Erklärung übernimmt der Anbieter also die Verantwortung für die Erfüllung dieser Anforderungen (vgl. Art. 47 Abs. 4).

Wenn bereits in sektorspezifischen Rechtsvorschriften (Harmonisierungsrechtsvorschriften) eine EU-Konformitätserklärung vorgesehen ist, wird, um Doppelverfahren zu vermeiden, eine einzige Erklärung ausgestellt (Art. 47 Abs. 3).

Hochrisiko-KI-Systeme müssen mit einer CE-Kennzeichnung («Conformité Européenne»-Kennzeichnung) versehen werden, aus der ihre Konformität mit der Verordnung hervorgeht, damit sie frei im Binnenmarkt verkehren können (vgl. Art. 48). Wenn Hochrisiko-KI-Systeme

unter andere Rechtsvorschriften der EU fallen, in denen die CE-Kennzeichnung ebenfalls vorgesehen ist, bedeutet die CE-Kennzeichnung, dass die Hochrisiko-KI-Systeme auch die Anforderungen dieser anderen Rechtsvorschriften erfüllen.

Schliesslich müssen die Anbieter und bestimmte Betreiber vor dem Inverkehrbringen, der Inbetriebnahme oder der Verwendung eines in Anhang III (mit Ausnahme von Nr. 2) aufgeführten Hochrisiko-KI-Systems sich selber und ihre Systeme gemäss ihren diesbezüglichen Pflichten registrieren (vgl. Art. 49), und zwar in der von der Europäischen Kommission in Zusammenarbeit mit den Mitgliedstaaten geführten Datenbank (vgl. Art. 71) (die in Anhang III Nr. 2 aufgeführten Hochrisiko-KI-Systeme [Infrastruktur] werden hingegen auf nationaler Ebene registriert). Die bereitzustellenden Informationen sind in Anhang VIII aufgeführt.

5.2.8 Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme

In Kapitel IV sieht die Verordnung Transparenzpflichten für Anbieter und Betreiber in bestimmten Fällen vor (Art. 50 ff.). Diese Pflichten gelten für alle KI-Systeme, unabhängig davon, ob es sich um Hochrisiko-KI-Systeme handelt oder nicht. Bei Hochrisiko-KI-Systemen ergänzen diese Pflichten die anderen Transparenzpflichten, die für diese Systeme bereits vorgesehen sind (Art. 50 Abs. 6):

- KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind: Die *Anbieter* müssen sicherstellen, dass die KI-Systeme so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren, beispielsweise einer Roboter-Software (Chatbot), es sei denn, dies ist aufgrund der Umstände und des Kontexts der Nutzung offensichtlich. Diese Pflicht gilt nicht für Systeme, die beispielsweise zur Aufdeckung von Straftaten verwendet werden (Art. 50 Abs. 1).

- KI-Systeme, einschliesslich KI-Systeme mit allgemeinem Verwendungszweck, die Audio-, Bild-, Video- oder Textinhalte erzeugen: Die *Anbieter* haben dafür zu sorgen, dass die Ausgaben des KI-Systems in einem maschinenlesbaren Format gekennzeichnet und als künstlich erzeugt oder manipuliert erkennbar sind. Die Verordnung sieht eine Ausnahme für KI-Systeme vor, die eine unterstützende Funktion für die Standardbearbeitung ausführen oder die vom Betreiber bereitgestellten Eingabedaten nicht wesentlich verändern, sowie für Systeme, die gesetzlich zur Aufdeckung von Straftaten zugelassen sind (Art. 50 Abs. 2).[241]

- Emotionserkennungssysteme oder Systeme zur biometrischen Kategorisierung: Die *Betreiber* müssen die von einem solchen System betroffenen natürlichen Personen über seinen Betrieb und die Verarbeitung personenbezogener Daten informieren. KI-Systeme, die rechtmässig zur Aufdeckung, Verhütung, Ermittlung und Verfolgung von Straftaten verwendet werden, sind von dieser Pflicht ausgenommen (Art. 50 Abs. 3).

- Deepfakes: Die *Betreiber* von KI-Systemen die Deepfakes erzeugen müssen offenlegen, dass der Inhalt künstlich erzeugt oder manipuliert wurde. Die Verwendung solcher

[241] Für praktische Erwägungen vgl. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 34 N 69.

Techniken im gesetzlichen Rahmen zur Aufdeckung, Verhütung, Ermittlung und Verfolgung von Straftaten, ist hingegen erlaubt. Deepfakes, die in einem künstlerischen, kreativen, satirischen, fiktionalen oder analogen Kontext verwendet werden, dürfen in einer Weise offengelegt werden, die die Darstellung oder den Genuss des Werks nicht beeinträchtigt (Art. 50 Abs. 4).

- Bei Texten, die veröffentlicht werden, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren, muss offengelegt werden, dass diese künstlich erzeugt oder manipuliert wurden, es sei denn, sie wurden einer menschlichen Überprüfung oder redaktionellen Kontrolle unterzogen und eine natürliche oder juristische Person trägt die redaktionelle Verantwortung für ihre Veröffentlichung (Art. 50 Abs. 4).

Die Verordnung sieht vor, dass die in Artikel 50 Absätze 1 bis 4 genannten Informationen den betreffenden natürlichen Personen spätestens zum Zeitpunkt der ersten Interaktion oder Aussetzung in klarer und eindeutiger Weise bereitgestellt werden müssen.

Das Büro für Künstliche Intelligenz (vgl. Ziff. 5.2.13) fördert die Ausarbeitung von Praxisleitfäden auf EU-Ebene, um die Umsetzung der Pflichten in Bezug auf die Feststellung und Kennzeichnung künstlich erzeugter oder manipulierter Inhalte zu erleichtern. Die Europäische Kommission kann solche Praxisleitfäden mittels Durchführungsrechtsakte genehmigen, oder einen Durchführungsrechtsakt erlassen, in dem gemeinsame Vorschriften für die Umsetzung dieser Pflichten festgelegt werden, wenn sie einen Praxisleitfaden für nicht angemessen hält (Art. 50 Abs. 7).

5.2.9 Andere Systeme der künstlichen Intelligenz

Wenn ein KI-System weder eine verbotene Praxis noch ein Hochrisiko-KI-System ist und auch nicht unter die Transparenzbestimmungen fällt (Art. 50), bestehen gemäss der KI-Verordnung grundsätzlich keine Pflichten für die Anbieter, Betreiber und anderen beteiligten Stellen.

Die Verordnung sieht jedoch vor, dass das Büro für Künstliche Intelligenz und die Mitgliedstaaten die Erarbeitung von Verhaltenskodizes fördern und erleichtern, mit denen die freiwillige Anwendung einiger oder aller der in Kapitel III Abschnitt 2 genannten Anforderungen (also die für die Hochrisiko-KI-Systeme geltenden Vorschriften) auf diese anderen Systeme unterstützt werden soll (Art. 95 Abs. 1). Die Verhaltenskodizes sollen auch andere Aspekte fördern, wie ethische Grundsätze, ein respektvoller Umgang mit den ökologischen Ressourcen, die KI-Kompetenz und die Verhinderung negativer Auswirkungen auf schutzbedürftige Personen oder Gruppen schutzbedürftiger Personen (Art. 95 Abs. 2).

Zudem sieht Artikel 4 der Verordnung eine Art allgemeine Pflicht für alle Anbieter und Betreiber von KI-Systemen vor, Massnahmen zu ergreifen, um nach besten Kräften sicherzustellen,

dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Mass an KI-Kompetenz verfügen («AI literacy»).[242]

5.2.10 Modelle der künstlichen Intelligenz mit allgemeinem Verwendungszweck

5.2.10.1 Allgemeine Bemerkungen

Wie oben ausgeführt (vgl. Ziff. 5.2.4), unterscheidet die KI-Verordnung zwischen KI-Systemen und GPAI-Modellen. Das Kapitel V der Verordnung ist den GPAI-Modellen gewidmet.

Die GPAI-Modelle waren im Vorschlag für eine KI-Verordnung der Europäischen Kommission vom 4. April 2021 nicht enthalten. Das Aufkommen dieser Modelle mit der Lancierung von ChatGPT im November 2022 hat jedoch die Verhandlungen zwischen dem Europäischen Parlament und dem Rat wesentlich mitbestimmt. Der endgültige Regulierungsansatz, auf den sich die Mitgesetzgeber geeinigt haben, umfasst neue Pflichten für Anbieter von «einfachen» GPAI-Modellen und solchen mit systemischen Risiken.

5.2.10.2 Pflichten für alle Modelle der künstlichen Intelligenz mit allgemeinem Verwendungszweck

Die Anbieter[243] von GPAI-Modellen haben unabhängig von deren Risiko namentlich folgende Pflichten (vgl. Art. 53):

- Sie müssen die technische Dokumentation ihres Modells erstellen und regelmässig aktualisieren. Die Elemente, die die Dokumentation mindestens enthalten muss, sind in den Anhängen XI und XII beschrieben.

  Die Dokumentation muss insbesondere die Einzelheiten des Trainings- und Testverfahrens und der Ergebnisse der Bewertung enthalten. Auf Anfrage muss sie dem Büro für Künstliche Intelligenz und den zuständigen nationalen Behörden zur Verfügung gestellt werden (vgl. Art. 53 Abs. 1 Bst. a).

- Die Anbieter von GPAI-Modellen müssen auch Informationen und die Dokumentation vorbereiten und aktualisieren, die Anbietern von KI-Systemen zur Verfügung gestellt werden, die das GPAI-Modell in ihre Systeme integrieren wollen, um sie insbesondere dabei zu unterstützen, die Fähigkeiten und Grenzen des GPAI-Modells zu verstehen und die Verordnung einzuhalten (vgl. Art. 53 Abs. 1 Bst. b).

- Weiter sind die Anbieter von GPAI-Modellen verpflichtet, interne Vorschriften zur Einhaltung des Urheberrechts der EU einzuführen (Art. 53 Abs. 1 Bst. c).[244]

---

[242] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 36 N 72.

[243] Für die GPAI-Modelle gibt es die Rolle des Betreibers nicht; vgl. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 31 N 62.

[244] Zur Auslegung dieser Pflicht vgl. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 33 N 65.

- Die Anbieter müssen ausserdem eine hinreichend detaillierte Zusammenfassung der für das Training des GPAI-Modells verwendeten Inhalte erstellen und veröffentlichen. Diese Zusammenfassung muss Parteien mit berechtigtem Interesse, einschliesslich der Inhaber von Urheberrechten, die Ausübung und Durchsetzung ihrer Rechte nach dem EU-Recht ermöglichen. Das Büro für Künstliche Intelligenz erarbeitet eine Vorlage, die den Anbietern helfen soll, die erforderliche Zusammenfassung in verbaler Form bereitzustellen (Art. 53 Abs. 1 Bst. d).

- Anbieter von GPAI-Modellen, die sich ausserhalb der EU befinden, aber in der EU GPAI-Modelle in Verkehr bringen, müssen einen Bevollmächtigten in der EU benennen (vgl. Art. 54).

GPAI-Modelle, die im Rahmen einer freien und quelloffenen Lizenz bereitgestellt werden, sind von den in Artikel 53 Absatz 1 Buchstaben a und b und Artikel 54 aufgeführten Pflichten ausgenommen. Diese Ausnahmeregelung gilt, wenn die Parameter der KI-Modelle, einschliesslich Gewichte, Informationen über die Modellarchitektur und Informationen über die Modellnutzung, öffentlich zugänglich sind. Sie gilt jedoch nicht für GPAI-Modelle, die ein systemisches Risiko beinhalten (vgl. Ziff. 5.2.10.3; Art. 53 Abs. 2 und 54 Abs. 6).

Anbieter von GPAI-Modellen können sich bis zur Veröffentlichung einer harmonisierten Norm auf Praxisleitfäden im Sinne von Artikel 56 stützen, um die Einhaltung ihrer Pflichten nachzuweisen. Die Einhaltung der harmonisierten Normen begründet für die Anbieter die Vermutung der Konformität (Art. 53 Abs. 4). Das Büro für Künstliche Intelligenz hat den Auftrag, die Schaffung von Praxisleitfäden unter Berücksichtigung internationaler Ansätze zu fördern und zu erleichtern.

5.2.10.3 Pflichten für Modelle der künstlichen Intelligenz mit allgemeinem Verwendungszweck mit systemischem Risiko

5.2.10.3.1 Begriff und Verfahren

Für GPAI-Modelle mit systemischem Risiko gelten gemäss dem Grundsatz, dass eine Steigerung der Fähigkeiten auch die Risiken erhöht, zusätzliche Pflichten.

Ein GPAI-Modell wird grundsätzlich als Modell mit systemischen Risiken angesehen, wenn es über «Fähigkeiten mit hohem Wirkungsgrad» verfügt (vgl. Art. 51). Als Kriterium verwendet die Verordnung die kumulierte Menge der für sein Training verwendeten Berechnungen, gemessen in Gleitkommaoperationen («floating point operations», FLOP) (vgl. E. 111 der Verordnung). Um zu bestimmen, ob es sich um ein GPAI-Modell mit systemischen Risiken handelt, wurde ein Grenzwert von 10^25 FLOP festgelegt (vgl. Art. 51 Abs. 2).

Dieser Grenzwert und dieses Kriterium können im Laufe der Zeit angepasst werden, um technologischen und industriellen Veränderungen, wie algorithmischen Verbesserungen oder erhöhter Hardware-Effizienz, Rechnung zu tragen. Das wissenschaftliche Gremium Sachverständiger (vgl. Ziff. 5.2.13) kann dem Büro für Künstliche Intelligenz eine qualifizierte Warnung übermitteln, wenn es Grund zur Annahme hat, dass ein GPAI-Modell ein systemisches Risiko auf EU-Ebene aufweist (Art. 90).

Die Vorschriften für die Einstufung eines GPAI-Modells als Modell mit systemischem Risiko sind in den Artikeln 51 und 52 der Verordnung enthalten. Wenn also ein KI-Modell mit allgemeinem Verwendungszweck die Bedingung nach Artikel 51 Absatz 1 Buchstabe a erfüllt (es

verfügt über Fähigkeiten mit hohem Wirkungsgrad), so teilt der betreffende Anbieter dies der Europäischen Kommission unverzüglich mit. Er kann gleichzeitig nachzuweisen versuchen, dass das Modell trotz allem keine systemischen Risiken aufweist (vgl. Art. 52 Abs. 2).

Die Europäische Kommission kann ein KI-Modell mit allgemeinem Verwendungszweck auch selber als Modell mit systemischen Risiken bezeichnen (Art. 52 Abs. 4). In diesem Fall kann der Anbieter eine Neubewertung beantragen.

Die Europäische Kommission veröffentlicht eine Liste der KI-Modelle mit allgemeinem Verwendungszweck, die ein systemisches Risiko aufweisen, und hält diese unter Beachtung der Vorschriften über geistiges Eigentum und der Geschäftsgeheimnisse auf dem neusten Stand (Art. 52 Abs. 6).

5.2.10.3.2 Zusätzliche Pflichten

Zusätzlich zu den Pflichten, die für alle GPAI-Modelle gelten (vgl. Art. 53 und 54, Ziff. 5.2.10.2), haben die Anbieter von GPAI-Modellen mit systemischen Risiken insbesondere folgende Pflichten (vgl. Art. 55):

- Angriffstests («adversarial testing»): Die Anbieter müssen die erforderlichen Bewertungen des Modells, insbesondere vor seinem ersten Inverkehrbringen, durchführen, wozu auch die Durchführung und Dokumentation von Angriffstests bei Modellen gehören, gegebenenfalls auch im Rahmen interner oder unabhängiger externer Tests, um systemische Risiken zu ermitteln und zu mindern.

- Minderung systemischer Risiken: Die Anbieter müssen mögliche systemische Risiken auf EU-Ebene, die sich aus der Entwicklung, dem Inverkehrbringen oder der Verwendung des Modells ergeben können, bewerten und mindern.

- Meldung von Vorfällen: Wenn trotz der Bemühungen, Risiken im Zusammenhang mit einem GPAI-Modell mit systemischen Risiken zu ermitteln und vorzubeugen, die Entwicklung oder Verwendung des Modells einen schwerwiegenden Vorfall verursacht, muss der Anbieter des Modells dem Vorfall unverzüglich nachgehen und dem Büro für Künstliche Intelligenz und den zuständigen nationalen Behörden alle einschlägigen Informationen und mögliche Korrekturmassnahmen mitteilen.

- Cybersicherheit: Die Anbieter müssen während des gesamten Lebenszyklus des Modells ein angemessenes Mass an Cybersicherheit für das Modell und seine physische Infrastruktur gewährleisten.

Die Anbieter können die Einhaltung dieser Pflicht über Praxisleitfäden nachweisen, deren Bereitstellung vom Büro für Künstliche Intelligenz bis zur Veröffentlichung einer harmonisierten Norm erleichtert wird. Die Einhaltung harmonisierter Normen begründet für die Anbieter die Vermutung der Konformität (Art. 55 Abs. 2; Ziff. 5.2.11).

5.2.11 Die harmonisierten Normen und ihre Rolle in der KI-Verordnung[245]

Die Normung spielt bei der Umsetzung der KI-Verordnung eine zentrale Rolle.[246] So können die Anbieter der betroffenen KI-Systeme und GPAI-Modelle durch die Einhaltung der harmonisierten Normen die Konformität mit den Anforderungen der Verordnung nachweisen.

Um zu verstehen, wie diese Normen mit der KI-Verordnung zusammenhängen, werden im Folgenden der Begriff «harmonisierte Normen» (vgl. Ziff. 5.2.11.1), das Normungssystem der EU (vgl. Ziff. 5.2.11.2) und der Mechanismus der Konformitätsvermutung (vgl. Ziff. 5.2.11.3) vorgestellt. Anschliessend werden die Normungsarbeiten im KI-Bereich erläutert (vgl. Ziff. 5.2.11.4).

5.2.11.1 Die harmonisierten Normen

Eine «harmonisierte Norm» ist eine europäische Norm, die auf der Grundlage eines Auftrags der Europäischen Kommission zur Durchführung von Harmonisierungsrechtsvorschriften der EU angenommen wurde (vgl. Art. 2 Abs. 1 Bst. c der Verordnung [EU] 1025/2012[247]). Diese technischen Normen sollen die europäische Rechtssetzung unterstützen. Dabei handelt es sich um nicht rechtsverbindliche Regeln, Leitlinien oder Merkmale, die von Fachpersonen in den europäischen Normungsorganisationen erarbeitet wurden und die fast alle Gebiete des modernen Wirtschafts- und Alltagslebens betreffen. Technische Normen regeln vielfältige materielle und immaterielle Gegenstände wie Produkte, Verfahren, Messmethoden, Prozesse und Dienstleistungen und kommen in nahezu allen Branchen und Fachgebieten zum Einsatz.[248]

5.2.11.2 Das Normungssystem der EU

Das System der europäischen Normungsorganisationen bilden das Europäische Komitee für Normung (CEN) und das Europäische Komitee für elektrotechnische Normung (CENELEC)

---

[245] Dieses Kapitel wurde auf der Grundlage von Beiträgen der DV verfasst.

[246] Vgl. E. 121 der KI-Verordnung; s. auch MARTINA ARIOLI, Risikomanagement (Fn. 222), 12 N 38 ff.

[247] Verordnung (EU) Nr. 1025/2012 des Europäischen Parlaments und des Rates vom 25. Oktober 2012 zur europäischen Normung, zur Änderung der Richtlinien 89/686/EWG und 93/15/EWG des Rates sowie der Richtlinien 94/9/EG, 94/25/EG, 95/16/EG, 97/23/EG, 98/34/EG, 2004/22/EG, 2007/23/EG, 2009/23/EG und 2009/105/EG des Europäischen Parlaments und des Rates und zur Aufhebung des Beschlusses 87/95/EWG des Rates und des Beschlusses Nr. 1673/2006/EG des Europäischen Parlaments und des Rates, ABl. L 316/12, 14. November 2012.

[248] SECO, Förderung der Normungsorganisationen im Bereich der Digitalisierung: Akteure und Erkenntnisse in ausgewählten Themenbereichen, Bericht an den Bundesrat, 16. August 2022, 6, abrufbar unter: www.admin.ch > Dokumentation > Medienmitteilungen > Publikationshinweis > Bericht Förderung der Normierungsorganisationen im Bereich der Digitalisierung: Akteure und Erkenntnisse in ausgewählten Themenbereichen (abgerufen am 26. August 2024); für eine Zusammenfassung der bestehenden Normen auf internationaler und europäischer Ebene s. MÉLANIE GORNET/WINSTON MAXWELL, Normes techniques et éthique de l'IA. CNIA 2023 – Conférence Nationale en Intelligence Artificielle, Strassburg 2023, abrufbar unter: https://hal.science/hal-04121843 (abgerufen am 26. August 2024).

zusammen mit dem Europäischen Institut für Telekommunikationsnormen (ETSI). Das CEN ist verantwortlich für die europäischen Normen (EN) in allen technischen Bereichen ausser der Elektrotechnik und der Telekommunikation. Das CENELEC ist zuständig für die europäische Normung im Bereich Elektrotechnik (ENEC-Kennzeichnung) und das ETSI im Bereich der Telekommunikation.

Aus dem 1985 von der EU eingeführten «New Approach»-Konzept auf dem Gebiet der Produktevorschriften ergibt sich eine Verknüpfung zwischen Gesetzgebung und Normung. Der Gesetzgeber beschränkt sich auf das Formulieren von grundlegenden Anforderungen und Schutzzielen, die technisch durch Normen konkretisiert werden. Dieses Vorgehen soll einerseits den EU-Rechtssetzungsmechanismus entlasten, andererseits die Anpassung an den Stand der Technik und des Wissens ermöglichen.

Diese harmonisierten Normen werden von den europäischen Normungsorganisationen (CEN, CENELEC, ETSI) aufgrund eines von der EU-Kommission erteilten Mandates erarbeitet und im Amtsblatt der EU veröffentlicht. Nach ihrer Veröffentlichung im Amtsblatt der EU begründet ihre Anwendung eine Konformitätsvermutung. Die Normung spielt daher bei der Umsetzung der KI-Verordnung eine zentrale Rolle.

5.2.11.3 Die Konformitätsvermutung

5.2.11.3.1 Im Allgemeinen

Bei Produkten, die in Übereinstimmung mit harmonisierten Normen hergestellt worden sind, wird davon ausgegangen, dass sie die entsprechenden wesentlichen Anforderungen der einschlägigen Rechtsvorschriften erfüllen. Dieses Verfahren wird auch «Konformitätsbewertungsverfahren» genannt und führt bei positivem Ergebnis zur Berechtigung, eine «Konformitätserklärung» zu unterzeichnen und auf dem Produkt das CE-Zeichen anzubringen.

Entsprechend der spezifischen EU-Richtlinie oder EU-Verordnung kann ein Inverkehrbringer das Konformitätsbewertungsverfahren im Rahmen der Selbstkontrolle durchführen oder er muss akkreditierte Prüfstellen («Certified Bodies») und Benannte Stellen («Notified Bodies») zuziehen. Im Bedarfsfall können auch andere Normen berücksichtigt werden; in diesem Fall muss aber der Nachweis geführt werden, dass auch mit diesen Normen die «grundlegenden Sicherheits- und Gesundheitsanforderungen» erfüllt werden. Bei anderen Lösungen muss der Stand der Technik und des Wissens nachgewiesen werden.

5.2.11.3.2 In der KI-Verordnung

Gemäss Artikel 40 Absatz 1 der KI-Verordnung wird bei Hochrisiko-KI-Systemen oder bei GPAI-Modellen, die mit harmonisierten Normen oder Teilen davon übereinstimmen, deren Fundstellen gemäss der Verordnung 1025/2012[249] im Amtsblatt der EU veröffentlicht wurden, eine Konformität mit den Anforderungen gemäss Kapitel III Abschnitt 2 (im Zusammenhang mit Hochrisiko-KI-Systemen; Art. 8 ff. der KI-Verordnung) oder gegebenenfalls mit den Pflich-

[249] Verordnung (EU) 1025/2012 zur europäischen Normung (Fn. 247).

ten gemäss Kapitel V Abschnitte 2 und 3 der Verordnung (im Zusammenhang mit GPAI-Modellen) vermutet, soweit diese Anforderungen oder Verpflichtungen von den Normen abgedeckt sind.

Mit dieser Bestimmung wird somit eine Vermutung der Konformität mit den Anforderungen der Verordnung eingeführt, wenn die harmonisierten Normen eingehalten werden.

Damit diese harmonisierten Normen erarbeitet werden, erteilt die Europäische Kommission den europäischen Normungsorganisationen einen Normungsauftrag (vgl. Ziff. 5.2.11.4).

5.2.11.4 Die Normungsarbeiten im KI-Bereich

Am 22. Mai 2023 erteilte die Europäische Kommission dem CEN und dem CENELEC den Auftrag[250], die nötigen technischen Normen für die Umsetzung der KI-Verordnung auszuarbeiten. Konkret hat die Europäische Kommission die beiden Komitees aufgefordert, bis am 30. April 2025 zehn neue europäische Normen im KI-Bereich zu erarbeiten. Die Normen betreffen die zehn folgenden Bereiche:

- Risikomanagement für KI-Systeme
- Governance und Qualität von Datensätzen, die zur Entwicklung von KI-Systemen verwendet werden
- Aufzeichnung durch Protokollierungsfunktionen von KI-Systemen
- Transparenz und Bereitstellung von Information an Nutzerinnen und Nutzer von KI-Systemen
- Menschliche Aufsicht über KI-Systeme
- Spezifikationen für die Genauigkeit von KI-Systemen
- Spezifikationen für die Robustheit von KI-Systemen
- Spezifikationen für die Cybersicherheit von KI-Systemen
- Qualitätsmanagementsysteme für Anbieter von KI-Systemen, einschliesslich Verfahren zur Beobachtung nach dem Inverkehrbringen
- Konformitätsbewertung für KI-Systeme

[250] Durchführungsbeschluss der Kommission vom 22. Mai 2023 über einen Normungsauftrag an das Europäische Komitee für Normung und das Europäische Komitee für elektrotechnische Normung zur Unterstützung der Unionspolitik im Bereich der künstlichen Intelligenz, C(2023) 3215 final, abrufbar unter: https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=de (abgerufen am 26. August 2024).

Die zehn oben erwähnten technischen Normen werden von Fachpersonen in einem gemischten technischen Komitee, dem «Joint CEN-CENELEC Technical Committee 21 'Artificial Intelligence'» (CEN-CLC/JTC 21), ausgearbeitet.

Ist dieses Komitee nicht in der Lage, die von der Europäischen Kommission verlangten Normen innerhalb der festgelegten Frist (bis am 30. April 2025) auszuarbeiten, so kann die Kommission gemäss Artikel 41 der Verordnung gemeinsame Spezifikationen erlassen. Dieser Mechanismus ermöglicht somit der Kommission, einen gewissen Druck auf die Normungsorganisationen auszuüben.

5.2.12 Massnahmen zur Innovationsförderung

Die KI-Verordnung versucht, die Innovation mit der Sicherheit der KI-Systeme in Einklang zu bringen, indem ein System von «KI-Reallaboren» («regulatory sandboxes») geschaffen wird, also rechtliche Rahmenwerke, die beschränkte Innovationstests unter gesetzlicher Kontrolle ermöglichen.[251]

Die Mitgliedstaaten sind verpflichtet, mindestens ein solches KI-Reallabor auf nationaler Ebene einzurichten, das zwei Jahre nach dem Inkrafttreten der Verordnung einsatzbereit sein sollte. Sie können dieser Verpflichtung auch durch Beteiligung an einem bestehenden Reallabor nachkommen, sofern diese Beteiligung die nationale Abdeckung der teilnehmenden Mitgliedstaaten in gleichwertigem Mass gewährleistet (vgl. Art. 57).

Das Ziel dieser Reallabore besteht darin, eine kontrollierte Umgebung bereitzustellen, um die Innovation zu fördern und die Entwicklung, das Training, das Testen und die Validierung innovativer KI-Systeme für einen begrenzten Zeitraum vor ihrem Inverkehrbringen oder ihrer Inbetriebnahme zu erleichtern (vgl. Art. 57 Abs. 5). Die zuständigen nationalen Behörden müssen das Büro für Künstliche Intelligenz und das Gremium für Künstliche Intelligenz (vgl. Ziff. 5.2.13) über die Einrichtung eines Reallabors unterrichten. Das Büro für Künstliche Intelligenz veröffentlicht eine Liste der geplanten und bestehenden Reallabore (Art. 57 Abs. 15).

Die zuständigen nationalen Behörden haben die Aufgabe, innerhalb der KI-Reallabore Anleitung, Aufsicht und Unterstützung bereitzustellen, um Risiken im Zusammenhang mit Grundrechten, Gesundheit und Sicherheit sowie Risikominderungsmassnahmen und deren Wirksamkeit hinsichtlich der Pflichten und Anforderungen der Verordnung zu ermitteln. Auf Anfrage der Anbieter oder zukünftigen Anbieter kann die zuständige Behörde einen Abschlussbericht vorlegen, in dem sie die im Reallabor durchgeführten Tätigkeiten und die entsprechenden Ergebnisse im Einzelnen darlegt. Die Anbieter können diese Unterlagen nutzen, um im Rahmen des Konformitätsbewertungsverfahrens die Einhaltung der Verordnung nachzuweisen (vgl. Art. 57 Abs. 7).

Die am KI-Reallabor beteiligten Anbieter und zukünftigen Anbieter bleiben nach dem Recht der EU und der Mitgliedstaaten für Schäden haftbar, die Dritten infolge der Erprobung im Re-

[251] Für eine Kritik zur Wirksamkeit dieser Massnahmen vgl. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 4 N 5.

allabor entstehen. Wenn jedoch der Anbieter den Plan und die Beteiligungsbedingungen beachtet hat und der Anleitung der zuständigen nationalen Behörden in gutem Glauben gefolgt ist, werden von den Behörden keine Geldbussen für Verstösse gegen die KI-Verordnung verhängt (Art. 57 Abs. 12). Die Verordnung schafft somit einen starken Anreiz für die Beteiligung an einem KI-Reallabor.

Schliesslich sieht die Verordnung unter bestimmten Voraussetzungen auch die Möglichkeit vor, Hochrisiko-KI-Systeme unter Realbedingungen in und ausserhalb eines KI-Reallabors zu testen (Art. 57, 58 und 60 f. der Verordnung). Die Durchführung eines Tests unter Realbedingungen muss von der Marktüberwachungsbehörde genehmigt werden (vgl. Art. 60).

5.2.13 Governance

Mit der Verordnung wird ein mehrstufiger Governance-Rahmen geschaffen, der zum Ziel hat, die Anwendung der Verordnung auf nationaler Ebene zu koordinieren und zu unterstützen, auf EU-Ebene Kapazitäten aufzubauen und Interessenträger im KI-Bereich zu integrieren.

Auf EU-Ebene sieht die Verordnung folgende Governance-Organe vor:

- Büro für Künstliche Intelligenz («AI Office», Art. 64): Dieses innerhalb der Europäischen Kommission geschaffene Büro soll zur Umsetzung, Beobachtung und Überwachung von KI-Systemen und GPAI-Modellen sowie der KI-Governance beitragen (vgl. Art. 3 Nr. 47). Es stellt insbesondere Muster bereit, die bei der Anwendung der Verordnung helfen (vgl. z. B. Art. 27 Abs. 5, der vorsieht, dass das Büro für Künstliche Intelligenz ein Muster für einen Fragebogen, auch mithilfe eines automatisierten Instruments, ausarbeitet, das die Betreiber bei ihrer Pflicht, eine Folgenabschätzung durchzuführen, unterstützen soll).

- Europäisches Gremium für Künstliche Intelligenz («European AI Board», Art. 65): Dieses setzt sich aus Vertreterinnen und Vertretern der Mitgliedstaaten zusammen und hat die Aufgabe, die Europäische Kommission und die Mitgliedstaaten zu beraten und zu unterstützen, um die einheitliche und wirksame Anwendung der Verordnung zu erleichtern. Der Europäische Datenschutzbeauftragte nimmt als Beobachter teil.

- Beratungsforum («Advisory Forum», Art. 67): Dieses stellt technisches Fachwissen bereit und berät das Gremium für Künstliche Intelligenz und das Büro für Künstliche Intelligenz. Es hat die Aufgabe, die Teilnahme der Interessenträger an der Umsetzung und Anwendung der Verordnung sicherzustellen. Das Forum setzt sich unter anderem aus Interessenträger der Industrie, der Universitäten, der Zivilgesellschaft sowie aus der Agentur der EU für Grundrechte, der Agentur der EU für Cybersicherheit, dem CEN, dem CENELEC und dem ETSI zusammen.

- Wissenschaftliches Gremium unabhängiger Sachverständiger («Scientific Panel of Independent Experts», Art. 68): Dieses setzt sich aus Sachverständigen der Wissenschaft zusammen und soll die Umsetzung und Anwendung der Verordnung unterstützen, namentlich die Durchsetzungstätigkeiten des Büros für Künstliche Intelligenz in Bezug auf GPAI-Modelle (vgl. Ziff. 5.2.10). Die Mitgliedstaaten können das Gremium um Hilfe für ihre Durchsetzungstätigkeiten im Rahmen der Verordnung bitten (Art. 69).

Auf nationaler Ebene schaffen oder benennen die Mitgliedstaaten mindestens eine notifizierende Behörde und mindestens eine Marktüberwachungsbehörde als zuständige nationale Behörden (Art. 70):

- Notifizierende Behörde: Dies ist die nationale Behörde, die dafür zuständig ist, die erforderlichen Verfahren für die Bewertung, Benennung und Notifizierung von Konformitätsbewertungsstellen und für deren Überwachung einzurichten und durchzuführen, also derjenigen Stellen, die für die Konformitätsbewertung von Hochrisiko-KI-Systemen zuständig sind (vgl. Art. 3 Nrn. 19–22; Ziff. 5.2.7.4).

- Marktüberwachungsbehörde: Hier handelt es sich um die nationale Behörde, die die Tätigkeiten durchführt und die Massnahmen ergreift, die in der Verordnung (EU) 2019/1020 über Marktüberwachung und die Konformität von Produkten[252] vorgesehen sind (vgl. Art. 3 Nr. 26) (vgl. Ziff. 5.2.14).

  Jeder Mitgliedstaat muss eine Marktüberwachungsbehörde benennen, die als zentrale Anlaufstelle für die Verordnung fungiert (Art. 70 Abs. 2).

5.2.14 Überwachung und Durchsetzung

Den Mitgliedstaaten kommt bei der Anwendung und Durchsetzung der Verordnung eine Schlüsselrolle zu. Jeder Mitgliedstaat benennt eine oder mehrere zuständige nationale Behörden, die die Anwendung und Umsetzung der Vorschriften überwachen und Marktüberwachungstätigkeiten durchführen.

Für die Überwachung verweist die KI-Verordnung auf die Verordnung (EU) 2019/1020[253] (vgl. Art. 74). Die KI-Verordnung enthält jedoch wichtige Präzisierungen, insbesondere zur Organisation und Koordinierung der Marktüberwachungsbehörden. So beispielsweise:

- Bei Hochrisiko-KI-Systemen und damit in Zusammenhang stehenden Produkten, auf die die in Anhang I Abschnitt A aufgeführten Harmonisierungsrechtsvorschriften der EU Anwendung finden, gilt als Marktüberwachungsbehörde im Sinne der KI-Verordnung die in diesen Vorschriften für die Marktüberwachung benannte Behörde (Art. 74 Abs. 3).

- Bei Hochrisiko-KI-Systemen, die in Finanzinstituten verwendet werden, gilt gemäss dem Recht der EU im Bereich der Finanzdienstleistungen die für die Finanzaufsicht

[252] Verordnung (EU) 2019/1020 des Europäischen Parlaments und des Rates vom 20. Juni 2019 über Marktüberwachung und die Konformität von Produkten sowie zur Änderung der Richtlinie 2004/42/EG und der Verordnungen (EG) Nr. 765/2008 und (EU) Nr. 305/2011, ABl. L 169/1, 25. Juni 2019.

[253] Verordnung (EU) 2019/1020 über Marktüberwachung und die Konformität von Produkten (Fn. 252).

über diese Institute benannte nationale Behörde als Marktüberwachungsbehörde, sofern ein Zusammenhang mit der Erbringung dieser Finanzdienstleistungen besteht (Art. 74 Abs. 6).

- Für die in Anhang III Nummer 1 (Biometrie) genannten Hochrisiko-KI-Systeme, sofern diese Systeme für Strafverfolgungszwecke, Grenzmanagement sowie Justiz und Demokratie eingesetzt werden, und für die in Anhang III Nummern 6 bis 8 genannten Hochrisiko-KI-Systeme benennen die Mitgliedstaaten als Marktüberwachungsbehörden die für den Datenschutz zuständigen Aufsichtsbehörden (Art. 74 Abs. 8).[254]

Grundsätzlich gewähren die Anbieter den Marktüberwachungsbehörden uneingeschränkten Zugang zur Dokumentation sowie zu den für die Entwicklung von Hochrisiko-KI-Systemen verwendeten Trainings-, Validierungs- und Testdatensätzen (Art. 74 Abs. 12). Unter bestimmten Bedingungen können die Überwachungsbehörden Zugang zum Quellcode eines Hochrisiko-KI-Systems erhalten (vgl. Art. 74 Abs. 13).

In der KI-Verordnung werden die Interventionsbefugnisse und die Verfahren beschrieben, die die Überwachungsbehörden je nach Umständen einsetzen können (Art. 79 ff.) Wenn ein Hochrisiko-KI-System beispielsweise ein Risiko für die Gesundheit, die Sicherheit oder die Grundrechte von Personen darstellt, führt die Überwachungsbehörde selber eine Konformitätsbewertung unter Berücksichtigung aller in der Verordnung festgelegten Anforderungen durch. Stellt sie fest, dass die Voraussetzungen nicht erfüllt sind, fordert sie den betreffenden Akteur auf, alle geeigneten Korrekturmassnahmen zu ergreifen (Art. 79).

In einer Art Generalklausel ermächtigt die Verordnung die Marktüberwachungsbehörde zudem, für jedes KI-System, das zwar den Vorgaben der Verordnung entspricht, aber dennoch ein Risiko für die Gesundheit oder Sicherheit von Personen, für die Grundrechte oder für andere Aspekte des Schutzes öffentlicher Interessen darstellt, besondere Massnahmen anzuordnen (vgl. Art. 82 Abs. 1).[255]

Für die Organe, Einrichtungen und sonstigen Stellen der UE, die in den Geltungsbereich der Verordnung fallen, wird der Europäische Datenschutzbeauftragte als zuständige Marktüberwachungsbehörde benannt (Art. 74 Abs. 9 und E. 156).

---

[254] Vgl. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 6 N 9, der der Ansicht ist, dass die Mitgliedstaaten, zusätzlich zu den in Art. 74 Abs. 8 der KI-Verordnung genannten Situationen, in ungefähr der Hälfte der Fälle die Datenschutzbehörden als Überwachungsbehörde bezeichnen werden. S. auch Erklärung des Europäischen Datenschutzausschusses vom 16. Juli 2024, in der die Datenschutzbehörden den Wunsch äussern, mit der Anwendung der KI-Verordnung in Bezug auf die Hochrisiko-KI-Systeme beauftragt zu werden. Die Erklärung ist abrufbar unter: https://www.edpb.europa.eu/news/news/2024/edpb-adopts-statement-dpas-role-ai-act-framework-eu-us-data-privacy-framework-faq_en (abgerufen am 26. August 2024).

[255] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 30 N 59.

Bei GPAI-Modellen verfügt die Europäische Kommission über das Büro für Künstliche Intelligenz über ausschliessliche Befugnisse zur Beaufsichtigung und Durchsetzung der Verordnung (vgl. Art. 88).

Ferner überträgt die KI-Verordnung den für den Schutz der Grundrechte zuständigen Behörden gewisse Befugnisse. Gemäss Artikel 77 Absatz 1 sind die nationalen Behörden oder öffentlichen Stellen, die die Einhaltung des EU-Rechts zum Schutz der Grundrechte beaufsichtigen oder durchsetzen, befugt, sämtliche Dokumentation im Zusammenhang mit der Verwendung der in Anhang III genannten Hochrisiko-KI-Systeme anzufordern, sofern dies für die wirksame Ausübung ihrer Aufträge im Rahmen ihrer Befugnisse notwendig ist. Die Mitgliedstaaten erstellen eine Liste dieser Behörden oder Stellen (Art. 77 Abs. 2). Wenn dies zur Feststellung, ob gegen die Pflichten verstossen wurde, notwendig ist, kann die Behörde oder die Stelle die Durchführung eines Tests des Hochrisiko-KI-Systems beantragen (vgl. Art. 77 Abs. 3).

5.2.15 Individuelle Rechte

In einem separaten Abschnitt mit dem Titel «Rechtsbehelfe» (Art. 85 f.) enthält die KI-Verordnung zwei Bestimmungen, die individuelle Rechte betreffen:

- Recht auf Beschwerde: Jede natürliche oder juristische Person, die Grund zur Annahme hat, dass gegen die Bestimmungen der Verordnung verstossen wurde, kann bei der betreffenden Marktüberwachungsbehörde Beschwerde einreichen. Diese Beschwerden werden gemäss der Verordnung (EU) 2019/1020[256] und der Durchführung von Überwachungstätigkeiten berücksichtigt. Dieses Recht besteht unbeschadet anderer verwaltungsrechtlicher oder gerichtlicher Rechtsbehelfe (Art. 85).

- Recht auf Erläuterung der Entscheidungsfindung im Einzelfall: Jede betroffene Person hat unter folgenden Bedingungen das Recht, vom Betreiber eines KI-Systems eine klare und aussagekräftige Erläuterung zur Rolle des KI-Systems im Entscheidungsprozess und zu den wichtigsten Elementen der getroffenen Entscheidung zu erhalten (Art. 86 Abs. 1):

    - Es handelt sich um ein Hochrisiko-KI-System, das in Anhang III aufgeführt ist, mit Ausnahme der unter Nummer 2 genannten Systeme (kritische Infrastruktur).

    - Die Entscheidung wurde vom Betreiber auf der Grundlage der Ausgaben dieses Systems getroffen.

    - Die Entscheidung hat rechtliche Auswirkungen oder beeinträchtigt die Person in ähnlicher Art erheblich auf eine Weise, die ihrer Ansicht nach negative Auswirkungen auf ihre Gesundheit, ihre Sicherheit oder ihre Grundrechte hat.

[256] Verordnung (EU) 2019/1020 über Marktüberwachung und die Konformität von Produkten (Fn. 252).

Dieses Recht kann nicht geltend gemacht werden, wenn das EU-Recht oder das nationale Recht Ausnahmen oder Beschränkungen vorsieht. Schliesslich gilt Artikel 86 nur insoweit, als das Recht gemäss Absatz 1 nicht anderweitig im EU-Recht festgelegt ist.

5.2.16 Sanktionen

Die Verordnung sieht für Verstösse gegen die in der KI-Verordnung festgelegten Pflichten erhebliche finanzielle Sanktionen vor.

Insbesondere werden Verstösse gegen die verbotenen KI-Praktiken mit Geldbussen von bis zu 35 Millionen Euro oder, im Fall von Unternehmen, von bis zu 7 Prozent des gesamten weltweiten Jahresumsatzes des vorangegangenen Geschäftsjahrs geahndet, je nachdem, welcher Betrag höher ist (vgl. Art. 99 Abs. 3).

Verstösse gegen die Bestimmungen über die Akteure oder die notifizierten Stellen (vgl. z. B die Pflichten der Anbieter nach Art. 16) werden mit Geldbussen von bis zu 15 Millionen Euro oder, im Fall von Unternehmen, von bis zu 3 Prozent des gesamten weltweiten Jahresumsatzes des vorangegangenen Geschäftsjahres geahndet, je nachdem, welcher Betrag höher ist (vgl. Art. 99 Abs. 4). Gegen Anbieter von KI-Modellen mit allgemeinem Verwendungszweck können Bussen im gleichen Umfang verhängt werden (vgl. Art. 101).

Werden notifizierten Stellen oder zuständigen nationalen Behörden auf deren Auskunftsersuchen hin falsche, unvollständige oder irreführende Informationen bereitgestellt, so werden Geldbussen von bis zu 7,5 Millionen Euro oder, im Fall von Unternehmen, von bis zu 1 Prozent des gesamten weltweiten Jahresumsatzes des vorangegangenen Geschäftsjahres verhängt, je nachdem, welcher Betrag höher ist (vgl. Art. 99 Abs. 5).

Für KMU, einschliesslich Start-up-Unternehmen, belaufen sich die vorstehend genannten Bussen auf höchstens die in Artikel 99 Absätze 3 bis 5 genannten Prozentsätze oder Beträge (vgl. Art. 99 Abs. 6).

5.2.17 Inkrafttreten und Geltungsbeginn

Die KI-Verordnung gilt nach Ablauf von zwei Jahren nach ihrem Inkrafttreten am 1. August 2024 (Art. 113). Die Verordnung ist für ihre Adressatinnen (Anbieter, Betreiber, betroffene Personen usw.) direkt anwendbar, ohne dass sie vorgängig in das Recht der Mitgliedstaaten übertragen werden muss (Art. 288 AEUV).[257] Es gibt jedoch Ausnahmen von der Frist von zwei Jahren. So beispielsweise:

- Die Kapitel I (Allgemeine Bestimmungen) und II (Verbotene Praktiken) gelten bereits sechs Monate nach dem Inkrafttreten, also ab dem 2. Februar 2025.

- Die Vorschriften für die GPAI-Modelle gelten, mit Ausnahme von Artikel 101, bereits zwölf Monate nach dem Inkrafttreten, also ab dem 2. August 2025. Das Gleiche gilt für

[257] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 2 N 2.

die Vorschriften für die Stellen, die Konformitätserklärungen ausstellen können, die Vorgaben für die Schaffung der in der Verordnung vorgesehenen Behörden und Stellen (z. B. das Büro für Künstliche Intelligenz) und die Bestimmungen über die Sanktionen.

- Artikel 6 Absatz 1 der Verordnung und die entsprechenden Pflichten (Hochrisiko-KI-Systeme im Rahmen der Harmonisierungsrechtsvorschriften der EU für Produkte) gilt erst 36 Monate nach dem Inkrafttreten, also ab dem 2. August 2027.

5.3 Würdigung

5.3.1 Rechtliche Auswirkungen auf die Schweizer Akteure

5.3.1.1 Betroffene Akteure

Wie oben erwähnt (vgl. Ziff. 5.2.5), gilt die KI-Verordnung für Anbieter, die sich in der EU *oder in einem Drittland* befinden oder dort niedergelassen sind und in der EU KI-Systeme in Verkehr bringen oder in Betrieb nehmen oder GPAI-Modelle in Verkehr bringen (Art. 2 Abs. 1 Bst. a). Das Inverkehrbringen oder die Inbetriebnahme betrifft gemäss den Begriffsbestimmungen (vgl. Art. 3 Nrn. 9 und 11) nur die erstmalige Bereitstellung beziehungsweise die erste Verwendung auf dem EU-Markt.

Gemäss Artikel 2 Absatz 1 Buchstabe c gilt die Verordnung auch für Anbieter und Betreiber von KI-Systemen, die ihren Sitz *in einem Drittland* haben oder sich *in einem Drittland* befinden, wenn die vom KI-System hervorgebrachte Ausgabe in der EU verwendet wird.

Mit diesen Bestimmungen soll verhindert werden, dass die Verordnung umgangen wird, indem KI-Systeme in einem Drittland entwickelt werden und das vom KI-System erzeugte Ergebnis schliesslich in der EU verwendet wird.

Ein Teil der Lehre weist jedoch darauf hin, dass die Auslegung von Artikel 2 Absatz 1 Buchstaben a und c nicht ganz unproblematisch ist: Da der Begriff «Anbieter» das Inverkehrbringen auf dem EU-Markt voraussetzt (vgl. Art. 3 Nrn. 3 und 9), bedeutet dies, dass ein Anbieter, der das KI-System nicht auf dem EU-Markt in Verkehr bringt, nicht als Anbieter im Sinne der Verordnung gilt, auch wenn sein KI-System letztlich den Weg in die EU findet oder dort Wirkung erzielt. Daher würde Artikel 2 Absatz 1 Buchstabe c, gemäss dem jeder Anbieter, der sich in einem Drittland befindet, der Verordnung untersteht, wenn die von seinem KI-System erzeugten Ergebnisse in der EU verwendet werden, den Begriff «Anbieter» *de facto* erweitern. In diesem Fall ist die Anwendung der Verordnung durch den blossen Umstand gegeben, dass das Ergebnis in der EU verwendet wird.[258] Es muss beobachtet werden, wie die beiden Bestimmungen künftig miteinander in Einklang gebracht werden. Es scheint auf jeden Fall, dass die Absicht des europäischen Gesetzgebers eher in Richtung einer umfassenden Anwendung der Verordnung geht.

Für die Betreiber, die sich in einem Drittland befinden, ist das Kriterium die Verwendung des Ergebnisses in der EU (vgl. Art. 2 Abs. 1 Bst. c). Als Beispiel kann der Fall eines Schweizer

[258] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 9 N 24; s. auch MARTINA ARIOLI, Risikomanagement (Fn. 222), 15 N 50 f.

Unternehmens angeführt werden, das seinen Kunden in der EU Texte zustellt, die mit einem KI-System generiert wurden.[259]

In den Erwägungen der KI-Verordnung wird auch der Fall erwähnt, in dem ein in der EU niedergelassener Akteur bestimmte Dienstleistungen im Zusammenhang mit einer Tätigkeit, die von einem als hochriskant eingestuften KI-System ausgeübt werden soll, an einen in einem Drittland niedergelassenen Akteur vergibt. Unter diesen Umständen könnte das vom Akteur in einem Drittland betriebene KI-System Daten verarbeiten, die rechtmässig in der EU erhoben und aus der EU übermittelt wurden, und dem vertraglichen Akteur in der EU die aus dieser Verarbeitung resultierende Ausgabe dieses KI-Systems liefern. Die KI-Verordnung sollte daher auch für diesen Fall gelten (vgl. E. 22).

5.3.1.2 Auswirkungen

Wenn Schweizer Akteure von der Verordnung betroffen sind, müssen sie die darin vorgesehenen Pflichten für das betreffende KI-System oder GPAI-Modell einhalten.

Weist das KI-System ein minimales Risiko auf, so gibt es keine Pflichten, sondern höchstens eine freiwillige Anwendung der Verhaltenskodizes (vgl. Ziff. 5.2.9). Wenn das KI-System von einem der in Artikel 50 aufgeführten Fälle erfasst ist, müssen die entsprechenden Transparenzpflichten eingehalten werden (vgl. Ziff. 5.2.8).

Für Hochrisiko-KI-Systeme gelten die Pflichten von Kapitel III Abschnitt 2 (vgl. Ziff. 5.2.7). Insbesondere müssen Schweizer Anbieter von Hochrisiko-KI-Systemen eine Konformitätsbewertung ihrer Produkte durchführen. Für weitere Ausführungen zu diesem Punkt wird auf die Erläuterungen unten im Zusammenhang mit dem MRA verwiesen (vgl. Ziff. 5.3.2).

KI-Systeme mit unannehmbaren Risiken sind schlicht und einfach verboten (abgesehen von den unter Ziff. 5.2.6 genannten Ausnahmen). Bei GPAI-Modellen haben sich die Anbieter an die Pflichten gemäss den Artikeln 51 ff. der KI-Verordnung zu halten (vgl. Ziff. 5.2.10).

Weiter ist zu beachten, dass die in der Schweiz niedergelassenen Anbieter von Hochrisiko-KI-Systemen und GPAI-Modellen vor dem Inverkehrbringen dieser Systeme und Modelle in der EU schriftlich einen in der EU niedergelassenen Bevollmächtigten benennen müssen (vgl. Art. 22 und 54 der KI-Verordnung). Bei Hochrisiko-KI-Systemen muss der Bevollmächtigte insbesondere sicherstellen, dass das KI-System in der EU-Datenbank gemäss Artikel 49 registriert ist.

Auf begründete Anfrage der zuständigen Behörden beziehungsweise des Büros für Künstliche Intelligenz muss der Bevollmächtigte alle Informationen übermitteln, die erforderlich sind, um die Konformität eines Hochrisiko-KI-Systems oder eines GPAI-Modells nachzuweisen (vgl. Art. 22 Abs. 3 Bst. c und d und 54 Abs. 3 Bst. c und d der KI-Verordnung). Zudem muss der Bevollmächtigte mit diesen bei allen ergriffenen Massnahmen zusammenarbeiten. Wenn ein Bevollmächtigter eines in der Schweiz niedergelassenen Anbieters aufgefordert würde, Informationen bereitzustellen, die sich in der Schweiz befinden, und diese den zuständigen

[259] Für weitere Beispiele s. DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 14 N 32.

europäischen Behörden ohne Bewilligung übermitteln würde, obwohl dafür die Schweizer Behörden zuständig wären, würde sich die Frage eines allfälligen Verstosses gegen Artikel 271 StGB stellen. Diese Bestimmung stellt Handlungen, die ohne Bewilligung für einen fremden Staat vorgenommen werden, unter Strafe.

Die Verordnung sieht weiter vor, dass der Bevollmächtigte den Auftrag beendet, wenn er der Auffassung ist oder Grund zur Annahme hat, dass der Anbieter gegen seine Pflichten gemäss der Verordnung verstösst.

Anbieter von Hochrisiko-KI-Systemen geben zudem auf dem Hochrisiko-KI-System oder, falls dies nicht möglich ist, auf seiner Verpackung oder in der beigefügten Dokumentation ihren Namen, ihren eingetragenen Handelsnamen oder ihre eingetragene Handelsmarke und ihre Kontaktanschrift an (vgl. Art. 16 Bst. b).

5.3.2 Verhältnis zum Abkommen zwischen der Schweiz und der EU über die gegenseitige Anerkennung von Konformitätsbewertungen[260]

5.3.2.1 Funktionsweise des Abkommens

Mit dem Abkommen zwischen der Schweiz und der EU über die gegenseitige Anerkennung von Konformitätsbewertungen (MRA)[261] soll die gegenseitige Anerkennung von Konformitätsbewertungen für die Vermarktung von Industrieprodukten (z. B. Maschinen, Medizinprodukte, Aufzüge) in Sektoren sichergestellt werden, in denen im EU-Recht harmonisierte Vorschriften existieren und eine zwingende Konformitätsbewertung vorgesehen ist.[262] Das Abkommen ermöglicht einen erleichterten Zugang zum EU-Binnenmarkt für die Produktsektoren, die es abdeckt, und reduziert Zeit und Kosten für die Vermarktung der Produkte. Konkret leistet das MRA einen Beitrag zum Abbau der technischen Handelshemmnisse. Zum einen geschieht dies durch die Harmonisierung der Vorschriften der Schweiz und der EU und zum anderen durch die Durchführung einer einzigen Konformitätsbewertung für den Zutritt zum EU-Markt beziehungsweise zum Schweizer Markt, die auf der Grundlage der technischen Vorschriften der Schweiz oder der EU von einer durch das Abkommen anerkannten Konformitätsbewertungsstelle ausgestellt wird. Das Abkommen beruht auf der Gleichwertigkeit der Gesetzgebungen der Schweiz und der EU. Bei jeder substanziellen Revision der technischen Vorschriften der EU in einem durch das Abkommen abgedeckten Sektor, muss die schweizerische Gesetzgebung angepasst werden, um die Gleichwertigkeit mit den Rechtsvorschriften der EU zu erhalten, und das entsprechende Kapitel des Abkommens aktualisiert werden.

Seit einigen Jahren lehnt die EU jedoch aufgrund institutioneller Fragen in den Beziehungen zwischen der Schweiz und der EU eine Aktualisierung des MRA ab. 2017 hat die EU eine

---

[260] Dieses Kapitel wurde auf der Grundlage von Beiträgen des BAKOM verfasst.

[261] Abkommen zwischen der Schweizerischen Eidgenossenschaft und der Europäischen Gemeinschaft über die gegenseitige Anerkennung von Konformitätsbewertungen, SR **0.946.526.81**.

[262] Botschaft zur Genehmigung der sektoriellen Abkommen zwischen der Schweiz und der EG, BBl **1999** 6128, 6212.

neue Regulierung der Medizinprodukte erlassen, die im Mai 2021 in Kraft getreten ist. Die Schweiz hat eine Gesetzgebung verabschiedet, die mit jener der EU im Einklang steht. Doch seit dem 26. Mai 2021 weigert sich die EU, das MRA im Bereich der Medizinprodukte zu aktualisieren, weil bei den Verhandlungen zu den institutionellen Fragen keine Fortschritte erzielt werden. Der Schweiz wird deshalb bei der Konformitätsbewertung nicht mehr die gegenseitige Anerkennung gewährt und Schweizer Anbieter von Medizinprodukten sind beim Zugang zum EU-Markt mit vielfältigen Hindernissen konfrontiert, wie der Notwendigkeit, die Konformität der Produkte durch eine Konformitätsbewertungsstelle der EU bewerten zu lassen. In Zukunft könnte diese Blockade auch andere Sektoren des MRA betreffen, die in der EU einer umfassenden Revision unterzogen wurden oder derzeit unterzogen werden (Maschinen, Bauprodukte, Spielzeug). Dies ist jedoch abhängig vom Ausgang der laufenden Verhandlungen zwischen der Schweiz und der EU.

5.3.2.2 Auswirkungen der KI-Verordnung auf die vom Abkommen betroffenen Schweizer Akteure

Die KI-Verordnung regelt die KI-spezifischen Aspekte in der Mehrheit der vom MRA abgedeckten Produktsektoren. Sie gilt zusätzlich zu den sektorspezifischen Rechtsvorschriften für Produkte und sieht spezifische zusätzliche Anforderungen im Zusammenhang mit KI vor.

In der Verordnung werden «Hochrisiko-KI-Systeme» als KI-Systeme definiert, die als Sicherheitsbauteile von Produkten verwendet werden sollen oder selbst Produkte sind, die unter die in Anhang I der Verordnung aufgeführten Rechtsvorschriften fallen – so beispielsweise Maschinen oder Sicherheitsbauteile von Maschinen – und die einer Konformitätsbewertung durch eine als Dritte auftretende Stelle unterzogen werden (vgl. Art. 6 Abs. 1).[263]

Folgende Sektoren des MRA entsprechen den durch Anhang I Abschnitt A der KI-Verordnung abgedeckten Bereichen:

- Maschinen (Kapitel 1 MRA)
- Persönliche Schutzausrüstungen (Kapitel 2 MRA)
- Spielzeuge (Kapitel 3 MRA)
- Medizinprodukte (Kapitel 4 MRA)
- Gasverbrauchseinrichtungen und Heizkessel (Kapitel 5 MRA)

[263] In der Verordnung wird unterschieden zwischen Harmonisierungsrechtsvorschriften gemäss dem neuen Rechtsrahmen («New Legislative Framework», NFL; s. dazu MARTINA ARIOLI, Risikomanagement [Fn. 222], 4 N 8 f.), die in Anhang I Abschnitt A der Verordnung aufgeführt sind, und Rechtsvorschriften nach dem alten Konzept, die in Anhang I Abschnitt B der Verordnung genannt sind. Die KI-Verordnung gilt direkt für Bereiche, die von Anhang I Abschnitt A erfasst sind, während die Europäische Kommission für die Bereiche, die unter das alte Konzept fallen, den Anforderungen an Hochrisiko-KI-Systeme Rechnung tragen muss, wenn sie delegierte Rechtsakte oder Durchführungsrechtsakte für KI-Systeme erlässt, bei denen es sich um Sicherheitsbauteile gemäss KI-Verordnung handelt. Ein grosser Teil der vom MRA erfassten Bereiche fällt unter Abschnitt A und nur zwei im MRA geregelte Bereiche sowie die Zivilluftfahrt und die Interoperabilität des Eisenbahnsystems gemäss den bilateralen Abkommen über Land- und Luftverkehr betreffen Abschnitt B.

- Druckgeräte (Kapitel 6 MRA)
- Funkanlagen und Telekommunikationsendgeräte (Kapitel 7 MRA)
- Geräte und Schutzsysteme zur bestimmungsgemässen Verwendung in explosionsgefährdeten Bereichen (ATEX) (Kapitel 8 MRA)
- Aufzüge (Kapitel 17 MRA)
- Seilbahnen (Kapitel 19 MRA)

Folgende Sektoren des MRA entsprechen den durch Anhang I Abschnitt B der KI-Verordnung abgedeckten Bereichen:

- Kraftfahrzeuge (Kapitel 12 MRA)
- Land- und forstwirtschaftliche Zugmaschinen (Kapitel 13 MRA)

Die KI-Verordnung gilt somit für die Mehrheit der durch das MRA abgedeckten Sektoren, die somit drei Jahre nach dem Inkrafttreten der Verordnung (Datum, ab dem Art. 6 Abs. 1 der KI-Verordnung und die entsprechenden Pflichten gelten, vgl. Ziff. 5.2.17) den Bestimmungen für Hochrisiko-KI-Systeme unterstehen. In den Bereichen, die unter die in Anhang I der Verordnung aufgeführten Harmonisierungsrechtsvorschriften der EU fallen, wird daher die Bewertung der Einhaltung der Anforderungen an die Hochrisiko-KI-Systeme in die Konformitätsbewertungsverfahren gemäss diesen Vorschriften integriert (vgl. Ziff. 5.2.7.1 und 5.2.7.4.1). Die Pflichten der KI-Verordnung ergänzen somit die bestehenden sektorspezifischen Pflichten.

Für Schweizer Anbieter von KI-Produkten, die sowohl unter das MRA und die Harmonisierungsrechtsvorschriften der EU gemäss Anhang I der KI-Verordnung (Hochrisiko-KI-Systeme gemäss Art. 6 Abs. 1 KI-Verordnung) fallen, erfolgt die Konformitätsbewertung gemäss dem im MRA vorgesehenen Verfahren, muss aber auch die Bewertung der Konformität mit den Anforderungen an Hochrisiko-KI-Systeme gemäss der KI-Verordnung beinhalten. In der gegenwärtigen Situation ohne schweizerische KI-Gesetzgebung und solange die Bestimmungen der KI-Verordnung für die oben genannten Produkte nicht in das MRA aufgenommen wurden, muss die Konformitätsbewertung für die Einhaltung der Anforderungen der KI-Verordnung durch eine Konformitätsbewertungsstelle der EU und gemäss EU-Recht durchgeführt werden. Um ein solches Produkt oder einen solchen Produktbestandteil in der EU in Verkehr zu bringen, müssen Schweizer Akteure zudem in der EU einen Bevollmächtigten bezeichnen, der sicherstellen muss, dass das Konformitätsbewertungsverfahren durchgeführt wurde (vgl. Art. 22 und 54 KI-Verordnung). Ausserdem müssen die Einführer (die gemäss Definition der KI-Verordnung in der EU niedergelassen sind) ihre Kontaktdaten auf der Verpackung angeben (Art. 23). Somit kommt es zu neuen technischen Handelshemmnissen.

5.3.2.3 Mögliche Erweiterung des MRA

Zwei Schritte wären nötig, um im Rahmen des MRA diese technischen Handelshemmnisse beim EU-Marktzugang gänzlich zu beseitigen. Erstens muss die Schweiz ihre Produktevorschriften mit denjenigen der EU harmonisieren, im vorliegenden Fall also mit den relevanten Produktvorschriften der KI-Verordnung. Zweitens müsste das MRA entsprechend erweitert werden.

Die EU und die Schweiz könnten folglich beschliessen, das MRA zu erweitern und die Produktvorschriften der KI-Verordnung in das MRA zu integrieren. Dies würde die gegenseitige Anerkennung künftiger Konformitätsbewertungen von Produkten, die KI-Systeme beinhalten,

ermöglichen. Diese Option ist jedoch abhängig vom Ausgang der aktuellen Verhandlungen zwischen der Schweiz und der EU.

Selbst wenn die Schweiz eine KI-Gesetzgebung schaffen würde, die sich an jene der EU anlehnt, würden die in der Schweiz durchgeführten Konformitätsbewertungsverfahren für die KI-Aspekte nicht anerkannt werden, solange das MRA nicht aktualisiert wurde.

5.3.3 Verhältnis zum Angemessenheitsbeschluss der Europäischen Kommission im Bereich Datenschutz

Die Übermittlung von Personendaten aus einem EU-Staat in einen Drittstaat unterliegt, bis auf einige Ausnahmen, der Voraussetzung, dass die Daten im Empfängerstaat auf ähnliche Weise geschützt sind. Dieses Schutzniveau kann durch einen Vertrag, Ad-hoc-Garantien (z. B. Standardvertragsklauseln) oder durch einen Angemessenheitsbeschluss gewährleistet werden (s. Art. 45 DSGVO und Art. 36 Richtlinie (EU) 2016/680[264]).[265] Dieser Beschluss bestätigt, dass der Empfängerstaat über ein Datenschutzniveau verfügt, das als gleichwertig gilt. In diesem Rahmen nimmt die Europäische Kommission unter Berücksichtigung der Anforderungen des EU-Rechts eine eingehende Prüfung der rechtlichen Situation im betreffenden Staat vor, insbesondere der einschlägigen Rechtsvorschriften und deren Umsetzung.

Die Schweiz verfügt seit dem Jahr 2000 über einen Angemessenheitsbeschluss der Europäischen Kommission.[266] Dieser wurde im Januar 2024 nach einer mehrjährigen Überprüfung zusammen mit jenen von zehn anderen Ländern[267] bestätigt. Dieser Beschluss betrifft die Übermittlung von Daten im privaten und öffentlichen Sektor, mit Ausnahme der Übermittlung im Rahmen der Strafverfolgung. Er ist sehr wichtig, denn er erlaubt, Personendaten ohne zusätzliche Garantien zu übermitteln. Er erleichtert die Wirtschaftsbeziehungen.

Die Verwendung von KI-Systemen in der EU kann zum einen die Übermittlung von Personendaten in die Schweiz und zum anderen die Bearbeitung von in die Schweiz übermittelten Personendaten von EU-Staatsangehörigen beinhalten. Das im Schweizer Recht gebotene Schutzniveau könnte daher relevant sein.

---

[264] Richtlinie (EU) 2016/680 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten im Bereich der Strafverfolgung (Fn. 234).

[265] Die Liste der Länder mit einem Angemessenheitsbeschluss kann abgerufen werden unter: https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en (abgerufen am 26. August 2024).

[266] Entscheidung der Kommission vom 26. Juli 2000 gemäss der Richtlinie 95/46/EG des Europäischen Parlaments und des Rates über die Angemessenheit des Schutzes personenbezogener Daten in der Schweiz, ABl. L 215, 25. August 2000, 0001–0003.

[267] Bericht der Kommission an das Europäische Parlament und den Rat über die erste Überprüfung der Wirkungsweise der Angemessenheitsfeststellungen gemäss Artikel 25 Absatz 6 der Richtlinie 95/46/EG vom 15. Januar 2024.

5.3.4 Verhältnis zum Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit

Die KI-Konvention ist ein verbindlicher völkerrechtlicher Vertrag, dem die Mitgliedstaaten des Europarats, die Staaten, die nicht Mitglieder des Europarats sind, und die EU beitreten können. Die KI-Verordnung ist eine EU-interne Regelung. Die KI-Konvention richtet sich an die Staaten und enthält allgemeine Regeln und Prinzipien, die bei einer Ratifizierung grundsätzlich in innerstaatliches Recht umgesetzt werden müssen. Die KI-Verordnung ist hingegen direkt anwendbar und enthält ausführlichere Vorschriften für die KI-Systeme in der EU. Sie sieht Pflichten für die Akteure im KI-Bereich vor. Die Art und die Tragweite dieser Texte sind somit verschieden.

Das Ziel der KI-Konvention, die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit zu schützen, erklärt sich aus dem Auftrag des Europarats, die Grundrechte zu schützen. Die KI-Verordnung ist in erster Linie eine Regulierung der KI-Produkte.[268] Ihr Zweck ist es, das Funktionieren des Binnenmarkts zu verbessern, indem ein einheitlicher Rechtsrahmen für KI-Systeme in der EU geschaffen wird, und gleichzeitig die Grundrechte zu schützen (vgl. Art. 1 Abs. 1 und E. 1 der Verordnung). Die Ziele sind also teilweise verschieden. Die KI-Konvention hat nicht zum Ziel, einheitliche Regeln für den Verkehr von KI-Produkten zu gewährleisten. In diesem Sinn unterscheidet sich die KI-Verordnung insbesondere durch ihre technischen Inhalte von der KI-Konvention. Es scheint daher, dass sich die beiden Texte gegenseitig ergänzen sollen.

Um dort, wo es möglich ist, einen Vergleich der KI-Konvention und der KI-Verordnung vornehmen zu können, enthält die Analyse eine Tabelle, die mögliche Korrelationen zwischen den Bestimmungen dieser beiden Texte aufzeigt (vgl. Anhang 1).

5.3.5 Weitere ausgewählte Elemente

In diesem Kapitel werden einige andere Merkmale der KI-Verordnung aus Sicht des Schweizer Rechts präsentiert, die sowohl im Zusammenhang mit der Frage der Auswirkungen der Verordnung auf die Schweizer Akteure als auch vor dem Hintergrund einer allfälligen Annäherung der schweizerischen Gesetzgebung an diese Verordnung wichtig erscheinen. Die Analyse erhebt keinen Anspruch auf Vollständigkeit:

- Wie bereits erwähnt, will die KI-Verordnung das Funktionieren des Binnenmarkts im Zusammenhang mit dem Verkehr von KI-Produkten in der EU verbessern und gleichzeitig die Grundrechte schützen (Art. 1 Abs. 1 der KI-Verordnung). Die Schweiz hat nicht den gleichen Bedarf, ihren Binnenmarkt zu vereinheitlichen.

- Die KI-Verordnung verweist häufig auf die europäischen Harmonisierungsrechtsvorschriften, die die im Binnenmarkt angebotenen Produkte regeln. Beispielsweise werden KI-Systeme als Hochrisiko-KI-Systeme eingestuft, wenn es sich um Produkte handelt, die unter die Harmonisierungsrechtsvorschriften der EU nach Anhang I fallen, oder um KI-Systeme, die als Sicherheitsbauteile dieser Produkte verwendet werden

[268] DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 5 N 7; MARTINA ARIOLI, Risikomanagement (Fn. 222), 4 N 8 f.

sollen (vgl. Art. 6 Abs. 1 der Verordnung). Zudem fügt sich die KI-Verordnung in die bestehenden Konformitätsbewertungsmechanismen ein und ergänzt diese mit Vorschriften (vgl. insb. Art. 43 der Verordnung).

Aus Sicht des Schweizer Rechts erhöht diese enge Verknüpfung mit den europäischen Harmonisierungsvorschriften die Komplexität der Verordnung.

- Die KI-Verordnung enthält zudem Mechanismen, die dem europäischen Gesetzgeber ein rasches Eingreifen ermöglichen, um den technologischen Entwicklungen Rechnung zu tragen. So sieht beispielsweise Artikel 7 Absatz 1 der Verordnung vor, dass die Europäische Kommission delegierte Rechtsakte zur Änderung von Anhang III durch Hinzufügen von Anwendungsfällen für Hochrisiko-KI-Systeme erlassen darf. Diese Befugnis hat zur Folge, dass sich der Inhalt der KI-Verordnung potenziell rasch verändern kann. Dies ist eine grosse Herausforderung für die Schweizer Akteure, die mögliche Entwicklungen genau beobachten und sich wenn möglich anpassen müssen. Im Fall einer allfälligen Annäherung der schweizerischen Gesetzgebung an die Rechtsvorschriften der EU müsste diesem Aspekt ebenfalls Rechnung getragen werden, um nötigenfalls rasch auf die Entwicklung des europäischen Rechts reagieren zu können.

- Die KI-Verordnung enthält transversale Vorschriften, die für alle KI-Systeme im privaten und öffentlichen Sektor gelten. In einem föderalistischen Staat wie der Schweiz wäre es aufgrund der Verteilung der Kompetenzen zwischen Bund und Kantonen nicht möglich, eine Bundesgesetzgebung mit einem so weiten Geltungsbereich im öffentlichen Recht zu haben. Die kantonalen Zuständigkeiten müssten gewahrt werden.

  Die KI-Verordnung sieht beispielsweise in Anhang III Nummer 3 Buchstabe c vor, dass «KI-Systeme, die bestimmungsgemäss zum Zweck der Bewertung des angemessenen Bildungsniveaus, das eine Person im Rahmen von oder innerhalb von Einrichtungen aller Ebenen der allgemeinen und beruflichen Bildung erhalten wird oder zu denen sie Zugang erhalten wird, verwendet werden sollen», als Hochrisiko-Systeme gelten. Da jedoch die Schweiz ein föderalistisches Bildungssystem hat, in dem die Kompetenzen der Kantone von besonderer Bedeutung sind, könnte eine solche Regel höchstens auf Bundesebene für die Bereiche vorgesehen werden, für die der Bund zuständig ist.

  Im Fall einer Annäherung der schweizerischen Gesetzgebung an die Rechtsvorschriften der EU im KI-Bereich müsste daher die Verteilung der Kompetenzen zwischen Bund und Kantonen gebührend berücksichtigt werden.

- Zudem müsste sich der Schweizer Gesetzgeber bewusst sein, dass die europäische Regelung die Erarbeitung von harmonisierten Normen, denen im KI-Bereich grosse Bedeutung zukommt, an die Normungsorganisationen delegiert (vgl. Ziff. 5.2.11).

- Die KI-Verordnung als solche enthält nur wenige individuelle Rechte (vgl. Art. 85 und 86 der Verordnung).[269] Sie ist jedoch nicht das einzige relevante Instrument in der EU im KI-Bereich. Andere europäische Rechtsvorschriften sind in diesem Bereich ebenfalls anwendbar.[270]

  Insbesondere im Zusammenhang mit dem Schutz der Grundrechte stellt die KI-Verordnung ein zusätzliches Element dar, das einen Beitrag leistet zur Umsetzung anderer grundlegender Rechtsvorschriften, die die Grundrechte schützen. So sind die Grundrechte in der EU und den Mitgliedstaaten beispielsweise bereits durch die geltende Gesetzgebung geschützt (z. B. im Bereich des Datenschutzes und der Nichtdiskriminierung). Die KI-Verordnung stellt sicher, dass die KI-Systeme ab der Konzeption diesen anderen Rechtsvorschriften entsprechen. Bei einem Verstoss erhalten die Behörden aufgrund der Anforderungen der KI-Verordnung zudem Zugang zu den Informationen, die nötig sind, um festzustellen, ob der Einsatz von KI rechtmässig erfolgt ist.[271] In welchem Umfang die Rechte des Einzelnen tatsächlich geschützt sind, hängt somit stark von den bestehenden Garantien in anderen materiellrechtlichen Vorschriften ab.

  Ein weiteres Beispiel ist die Haftung. In der KI-Verordnung sind beispielsweise die Fragen der zivilrechtlichen Haftung nicht geregelt. Um Ersatz für Schäden zu erhalten, die Personen durch die Verwendung von KI-Systemen erlitten haben, müssen sich geschädigte Personen auf andere materielle Rechtsvorschriften beziehen (s. z. B. Vorschlag für eine Richtlinie über die Haftung für fehlerhafte Produkte, vgl. Ziff. 6.3.2). Die KI-Verordnung schafft die Voraussetzungen, um die Umsetzung dieser materiellrechtlichen Vorschriften zu erleichtern.

  Eine allfällige Annäherung der Schweiz an die KI-Verordnung müsste dem Umstand Rechnung tragen, dass diese nicht alle Fragen im Zusammenhang mit KI abschliessend regelt.

Diese Erwägungen zeigen, dass die KI-Verordnung aus Sicht des Schweizer Rechts Besonderheiten aufweist, die in rechtlicher Hinsicht gewisse Fragen aufwerfen, sei es im Zusammenhang mit den Auswirkungen auf die Schweizer Akteure oder im Fall einer allfälligen Annäherung der schweizerischen Gesetzgebung an diese Verordnung.

---

[269] Für eine Kritik zu diesem Ansatz vgl. ANGELA MÜLLER, Der Artificial intelligence Act der EU (Fn. 223), 19 f.; MARTINA ARIOLI, Risikomanagement (Fn. 222), 16 N 54.

[270] Vgl. in diesem Sinn: DAVID ROSENTHAL, Der EU AI Act (Fn. 221), 5 N 9, der festhält, dass in der KI-Verordnung mehrmals erwähnt wird, dass diese zusätzlich zum bestehenden Recht gilt, vgl. z. B. Art. 2 Abs. 7 und E. 10.

[271] Vgl. Europäische Kommission, Künstliche Intelligenz – Fragen und Antworten, *Wie werden die Grundrechte durch die neuen Vorschriften geschützt*?, Dezember 2023, abrufbar unter: https://ec.europa.eu/commission/presscorner/detail/de/qanda_21_1683 (abgerufen am 26. August 2024).

5.4 Zwischenfazit

Die Untersuchung der KI-Verordnung hat ermöglicht, einen Überblick über ihren Inhalt zu erhalten und die wichtigsten Herausforderungen der Verordnung aus Sicht des Schweizer Rechts aufzuzeigen. Folgende Schlussfolgerungen können formuliert werden:

- Im Mittelpunkt der Verordnung stehen die Pflichten der Anbieter von Hochrisiko-KI-Systemen. Daher sind die Schweizer Akteure, die unter diese Kategorie fallen, von der KI-Verordnung am stärksten betroffen.

- Die KI-Verordnung ergänzt die harmonisierten Rechtsvorschriften der EU, indem sie darauf verweist und sie durch Vorschriften vervollständigt. Dies trägt zur Komplexität der Verordnung bei.

- Bestünde der politische Wille, sich an die KI-Verordnung anzunähern, müsste den jeweiligen Besonderheiten der schweizerischen und europäischen Rechtsordnung, beispielsweise der Verteilung der Kompetenzen zwischen Bund und Kantonen, Rechnung getragen werden.

- Die KI-Verordnung bezweckt, das Funktionieren des Binnenmarkts der EU im Zusammenhang mit den KI-Systemen zu harmonisieren und gleichzeitig die Grundrechte zu schützen. Sie enthält jedoch nur wenige individuelle Rechte. Der Schutz der individuellen Interessen im KI-Bereich ist grundsätzlich durch andere europäische Rechtsnormen, beispielsweise im Bereich des Datenschutzes und der Nichtdiskriminierung, gewährleistet. Die KI-Verordnung stärkt diese Rechte und deren Umsetzung im Rahmen von KI.

- In Bezug auf das Verhältnis zur KI-Konvention des Europarats hat die Analyse ergeben, dass die Art, die Tragweite und die Ziele der beiden Texte teilweise unterschiedlich sind. Es ist davon auszugehen, dass sie sich ergänzen.

- Würde die Schweiz eine der KI-Verordnung gleichwertige Gesetzgebung verabschieden und würden diese Gesetzgebung und die KI-Verordnung der EU in das MRA aufgenommen, so könnte die Schweiz von einer gegenseitigen Anerkennung der zukünftigen Konformitätsbewertungen von Produkten profitieren, die KI-Systeme beinhalten.

  Ohne gegenseitige Anerkennung müssen sich die betroffenen Schweizer Akteure, die Produkte auf den EU-Markt exportieren, für die Bewertung der Konformität der KI-Aspekte ihrer Produkte an die Anforderungen der KI-Verordnung halten.

Die KI-Verordnung wurde in der EU erst vor Kurzem verabschiedet. Viele Fragen im Zusammenhang mit ihrer Anwendung, sowohl aus rechtlicher als auch aus praktischer Sicht, müssen noch geklärt werden. Mit der Analyse wurden die wichtigsten Herausforderungen aufgezeigt, die sich zum jetzigen Zeitpunkt für die Schweiz abzeichnen. Die Analyse müsste vertieft werden, wenn die Politik eine Annäherung an diese Verordnung anstreben sollte.

6 Weitere ausgewählte Rechtsgebiete

6.1 Einleitung

Die KI-Konvention des Europarats und die KI-Verordnung der EU sind wichtige internationale Entwicklungen. Allerdings decken sie nicht alle für KI relevante Rechtsbereiche ab. In einigen Fällen verweisen sie lediglich auf Rechtsvorschriften aus anderen Bereichen. Dies ist beispielsweise der Fall beim Recht des geistigen Eigentums, beim Haftpflichtrecht und bei der strafrechtlichen Verantwortlichkeit.

Im Sinne eines etwas breiteren Überblicks über die Rechtslage in der Schweiz im Hinblick auf die Herausforderungen von KI werden in diesem Kapitel einige weitere ausgewählte Rechtsgebiete erörtert und der aktuelle Stand der Gesetzgebung skizziert.

Auch in diesen Bereichen sind weitere Vertiefungen erforderlich, um eine vollständige Übersicht zu erhalten.

6.2 Geistiges Eigentum[272]

KI verändert immer mehr die Art und Weise, wie heute Erfindungen oder Inhalte wie Texte, Musik oder Bilder entstehen. Das führt im Bereich des geistigen Eigentums dazu, dass etablierte Grundsätze allenfalls zu überdenken sind. Insbesondere im Urheberrecht und im Patentrecht tauchen Fragestellungen auf, welche zu Herausforderungen führen können. Es ist zu prüfen, inwieweit die veränderten Verhältnisse einer Anpassung des Urheber- und des Patentrechts bedürfen.

6.2.1 Urheberrecht und künstliche Intelligenz

6.2.1.1 Allgemeines

Das URG regelt primär den Schutz der Urheberinnen und Urheber von Werken der Literatur und Kunst (Art. 1 Abs. 1 Bst. a URG).

Artikel 2 Absatz 1 URG hält dabei fest, dass Werke «[…] unabhängig von ihrem Wert oder Zweck, geistige Schöpfungen der Literatur und Kunst [sind], die individuellen Charakter haben». Der Begriff der geistigen Schöpfung setzt eine Gedankenäusserung voraus. Urheberin oder Urheber kann nur eine natürliche Person sein (Art. 6 URG). Was nicht von einem Menschen geschaffen wurde, ist kein Werk.[273] Die geistige Schöpfung muss damit Ausdruck einer menschlichen Gedankenäusserung sein. Die Begriffe der Literatur und Kunst sind weit zu verstehen. Eine nicht abschliessende Aufzählung, was alles zu Literatur und Kunst gehören kann, findet sich in Artikel 2 Absatz 2 URG (z. B. literarische Sprachwerke, Werke der Musik, Werke der bildenden Kunst usw.). Geschützt soll aber in der Regel nur sein, was individuellen Charakter besitzt. Damit wird eine gewisse Einmaligkeit verlangt bzw. Merkmale, die eine

[272] Dieses Kapitel wurde vom IGE verfasst.

[273] BBl **1989** III 477, hier 521.

Schöpfung von anderen bestehenden oder möglichen Schöpfungen abheben.[274] Originalität im Sinne einer persönlichen Prägung durch die Urheberin oder den Urheber ist gemäss bundesgerichtlicher Rechtsprechung nicht erforderlich. Vorausgesetzt wird, dass der individuelle Charakter im Werk selbst zum Ausdruck kommt. Der individuelle Charakter hängt dabei vom Spielraum der Urheberin bzw. des Urhebers ab. Ist der Spielraum klein, wird der urheberrechtliche Schutz bereits gewährt, wenn nur ein geringer Grad an selbständiger Tätigkeit vorliegt.[275]

Die Urheberinnen und Urheber haben das ausschliessliche Recht zu bestimmen, ob, wann und wie ihre Werke verwendet werden (Art. 10 Abs. 1 URG). Unterschieden wird dabei üblicherweise zwischen Nutzungsrechten und Urheberpersönlichkeitsrechten. Die Nutzungsrechte haben in der Regel eine vermögensrechtliche Komponente; so erfolgt z. B. die Erlaubnis, ein Werk zu vervielfältigen, oft gegen Bezahlung einer Vergütung. Die Urheberpersönlichkeitsrechte weisen darüber hinaus eine ideelle Komponente auf und betonen damit die persönliche Verbundenheit zwischen Urheberin bzw. Urheber und Werk. Zu den Urheberpersönlichkeitsrechten zählt z. B. das Recht auf Anerkennung der Urheberschaft (Art. 9 URG).

Eine wichtige Eigenschaft des Urheberrechtsgesetzes ist seine sogenannte Technologieneutralität. Dies bedeutet, dass das Gesetz nicht danach unterscheidet, mit welchen technischen Verfahren ein Werk genutzt wird. Analoge und digitale Vervielfältigungen eines Werkes sind beispielsweise gleichermassen erfasst. Dies bietet den Vorteil, dass nicht jeglicher technischer Fortschritt eine Anpassung des Gesetzes notwendig macht. Dies bedeutet aber nicht, dass technologische Veränderungen nie Einfluss auf das Urheberrechtsgesetz haben. Die enorme technologische Entwicklung im digitalen Umfeld (man denke an das Internet) und dadurch entstehende Nutzungsformen führten mehrere Male zu einer Teilrevision des Urheberrechtsgesetzes.[276]

6.2.1.2 Fragestellungen und Regulierungsbedarf im Zusammenhang mit KI

Die Zunahme von Anwendungen generativer KI stellt das Urheberrecht erneut vor Herausforderungen und etablierte Grundsätze sind allenfalls zu hinterfragen. Im Zusammenhang mit generativen KI-Systemen sind folgende Bereiche urheberrechtlich genauer zu analysieren: erstens das Training von KI-Systemen, zweitens die Ergebnisse von KI-Systemen und drittens die erstellten Prompts.

---

[274] BBl **1989** III 477, hier 521.

[275] BGE **143** III 373, E. 2.1 S. 377.

[276] So zum Beispiel die Änderungen des Urheberrechtsgesetzes, welche 2008 und 2020 in Kraft traten.

Das Training von KI-Systemen bedarf einer grossen Menge[277] an Daten. Die Trainingsdatensätze setzen sich zum Beispiel aus im Internet verfügbaren Bildern und Bildbeschreibungen zusammen, welche in systematischen und automatisierten Verfahren heruntergeladen werden.[278] Die KI lernt anschliessend anhand der erstellten Trainingsdatensätze. Die für das Training von KI-Systemen verwendeten Inhalte können dabei urheberrechtlich geschützt sein. Es stellt sich deshalb die Frage, ob die Verwendung von urheberrechtlich geschützten Inhalten für das Training der KI-Systeme einer Einwilligung der Rechteinhaberinnen und Rechteinhaber (oder einer Schrankenregelung) bedarf. Die Antwort auf diese Frage ist von verschiedenen Faktoren abhängig. So zunächst davon, welches Recht auf das Training der KI anwendbar ist.[279] Mit Blick auf das vorliegend interessierende Schweizer Recht ist relevant, ob das Erstellen der Trainingsdatensets und/oder das anschliessende Training des KI-Systems eine urheberrechtlich relevante «Verwendung» von Werken darstellt. Ist die Verwendung urheberrechtlich relevant, greift sie in das Ausschliesslichkeitsrecht der Rechteinhaberinnen und Rechteinhaber ein und bedarf einer Erlaubnis. Artikel 10 Absatz 2 URG zählt verschiedene erlaubnispflichtige Werkverwendungen auf, so zum Beispiel die Vervielfältigung (Kopie). Eine relevante Werkverwendung liegt vor, wenn ein Werk nicht nur konsumiert, sondern vervielfältigt, an eine andere Person vermittelt und wahrnehmbar gemacht wird.[280] Sie ist damit oft eine Vorbereitungshandlung zum Werkgenuss[281] (dieser stellt selber keinen Eingriff in das Ausschliesslichkeitsrecht dar).

In der Lehre ist man sich nun uneinig darüber, ob nur jene Vervielfältigungen als urheberrechtlich relevant erfasst werden sollen, welche gezielt auf einen anschliessenden Werkgenuss ausgerichtet sind, oder auch solche, welche den Werkgenuss lediglich potenziell ermöglichen. Die herrschende Lehre scheint den letzteren Fall zu bejahen.[282] Folgt man diesem Verständnis, so sind Vervielfältigungen, welche für die Erstellung von Trainingsdatensätzen gemacht werden und allfällige Vervielfältigungen, welche für das Training der KI-Systeme benötigt werden, als urheberrechtlich relevant anzusehen (da sie potenziell den Werkgenuss ermöglichen). In die-

---

[277] Beispielhaft: ADITYA RAMESH/PRAFULLA DHARIWAL/ALEX NICHOL/CASEY CHU/MARK CHEN, Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022, 23 (abrufbar unter: https://arxiv.org/abs/2204.061, Stand: 6.5.2024).

[278] PAULINA JO PESCH/RAINER BÖHME, Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern, GRUR 2023/14, 997 ff., 998.

[279] Antwort auf diese Frage gibt das Internationale Privatrecht.

[280] DENIS BARRELET/WILLI EGLOFF, Das neue Urheberrecht. Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, Art. 10 N 8.

[281] MATHIS BERGER, Künstliche Intelligenz und Immaterialgüterrecht, Jusletter IT 4. Juli 2024, 4.

[282] SANDRA MARMY-BRÄNDLI/ISABELLE OEHRI, Das Training künstlicher Intelligenz, sic! 2023, 655 ff., 658 (mit weiteren Hinweisen).

sem Sinne findet sich in der Lehre die Auffassung, dass hier Eingriffe in das Vervielfältigungsrecht stattfinden.[283] Es gibt jedoch auch andere Auffassungen. Verschiedene Autorinnen und Autoren sehen in der Erstellung der Trainingsdatensets zwar eine urheberrechtlich relevante Vervielfältigung, beim Training des KI-Systems jedoch einen urheberrechtsfreien Werkgenuss.[284] Wiederum andere Autorinnen und Autoren wählen einen anderen Ansatz und stellen für die Frage, ob eine urheberrechtlich relevante Vervielfältigung vorliegt darauf ab, ob es zu einer wirtschaftlichen Verwertung des Werkes kommt (dies in Anlehnung an Überlegungen im US-amerikanischen Rechtsraum). Die Schweizer Gerichte hatten die urheberrechtliche Relevanz der beiden Sachverhalte, soweit ersichtlich, noch nicht zu beurteilen. Die Frage ist damit noch nicht geklärt. Sollten die Gerichte jedoch zum Schluss kommen, dass das Erstellen der Trainingsdatensätze und/oder das Training der KI-Systeme urheberrechtlich relevante Vervielfältigungen sind, so müssten hierfür Einwilligungen der Rechteinhaberinnen und Rechteinhaber eingeholt werden, soweit nicht eine Schrankenregelung greift.

Denkbar wären hier die Schranken des betriebsinternen Gebrauchs (Art. 19 Abs. 1 Bst. c URG), der vorübergehenden Vervielfältigungen (Art. 24*a* URG) oder zur Werkverwendung zum Zweck der wissenschaftlichen Forschung (Art. 24*d* URG). Bei jeder der genannten Schranken gibt es jedoch Elemente, welche, je nach konkretem Fall, vorliegend eine Anwendung verhindern könnten: bei der Schranke zum betriebsinternen Gebrauch beispielsweise das Kriterium des internen Informations- oder Dokumentationszwecks, bei der Schranke zur vorübergehenden Vervielfältigung das Kriterium der lediglich flüchtigen oder begleitenden Vervielfältigung und bei der Wissenschaftsschranke der Hauptzweck, der in der wissenschaftlichen Forschung liegen muss. Ob eine Schrankenregelung auf eine konkrete KI angewendet werden könnte, wäre deshalb im Einzelfall genauer zu prüfen.[285]

Unabhängig davon, wie die offenen Fragen beantwortet werden, kann sich daraus allfälliger gesetzgeberischer Handlungsbedarf ergeben: Sollte das Erstellen und/oder das Training von KI-Systemen urheberrechtlich relevante Vervielfältigungen beinhalten und, mangels einer bereits bestehenden Schrankenregelung, individuelle Einwilligungen der Rechteinhaberinnen und Rechteinhaber erfordern, so wäre sicherzustellen, dass die Entwicklung von KI-Systemen künftig in angemessener Weise möglich ist. Sollten diese Vorgänge jedoch als urheberrechtlich nicht relevant angesehen werden, so wäre wiederum zu klären, was dies für die Rechteinhaberinnen und Rechteinhaber der benötigten Werke bedeutet und welche allfälligen Massnahmen zu treffen wären.

Betreffend die Ergebnisse eines KI-Systems (Output) stellt sich einerseits die Frage, ob diese Urheberrechte verletzen können und andererseits, ob sie selber urheberrechtlichen Schutz geniessen. Mit Blick auf die Frage der Verletzung von Urheberrechten ist zu prüfen, ob das KI-Ergebnis ein Werk unverändert wiedergibt oder dann in einer Weise, in der das Original-

---

[283] SANDRA MARMY-BRÄNDLI/ISABELLE OEHRI, Das Training (Fn. 282), 657 f. (mit weiteren Hinweisen).

[284] Vgl. zur Auseinandersetzung hierzu SANDRA MARMY-BRÄNDLI/ISABELLE OEHRI, Das Training (Fn. 282), 658 f.

[285] Vgl. hierzu auch MATHIS BERGER, Künstliche Intelligenz und Immaterialgüterrecht (Fn. 281), 5 ff.

werk noch erkennbar ist. Wird ein solches KI-Ergebnis ohne Einwilligung (der Rechteinhaberin bzw. des Rechteinhabers oder einer Schrankenbestimmung) verwendet, stellt dies eine Urheberrechtsverletzung dar.

Ob das KI-Ergebnis selber urheberrechtlichen Schutz geniesst, hängt (neben der Frage der Individualität) davon ab, ob es sich dabei um eine «geistige Schöpfung» handelt.[286] Das Bundesgericht stellt an das Mass der geistigen Schöpfung keine hohen Anforderungen. Bereits beim Vorliegen eines geringen Grades von selbständiger geistiger Tätigkeit wird der urheberrechtliche Schutz gewährt.[287] Nicht urheberrechtlich geschützt sind demgegenüber Kreationen, welche beispielsweise von Tieren geschaffen werden oder reine Zufallsprodukte der Technik darstellen. Kein Hinderungsgrund für die geistige Schöpfung ist jedoch, wenn sich ein Mensch für die Schaffung eines Werkes bewusst bestimmter Zufallsprinzipien (z. B. Action Painting) oder der Technik (z. B. Computer) bedient.[288] Wo der menschliche Wille über das Ergebnis entscheidet, liegt eine geistige Schöpfung vor.[289] Es ist deshalb davon auszugehen, dass es sich beim Ergebnis des KI-Systems dann um eine geistige Schöpfung handelt, wenn die Nutzerin bzw. der Nutzer trotz des Einsatzes der KI noch auf wesentliche Elemente des Ergebnisses Einfluss nimmt und zwar selbst dann, wenn das KI-System ein gewisses Mass an Zufall in das Ergebnis einbringt. Nicht als geistige Schöpfung und damit auch urheberrechtlich nicht geschützt wäre demgegenüber ein Ergebnis, welches von einem KI-System eigenständig oder unter lediglich geringer Einflussnahme der Nutzerin oder des Nutzers entstand.[290]

Betreffend den urheberrechtlichen Schutz der von der Nutzerin bzw. vom Nutzer eingegebenen Prompts kann auf die vorangehenden Ausführungen verwiesen werden. Ein von einem Menschen geschriebener Prompt erfüllt die Kriterien der geistigen Schöpfung und der Literatur. Wesentlich für den urheberrechtlichen Schutz wird deshalb sein, ob der Prompt genügenden «individuellen Charakter» aufweist. Dies ist jeweils im Einzelfall zu beurteilen.

Generative KI-Systeme werden häufig mit urheberrechtlich geschützten Werken trainiert. Gleichzeitig nutzen immer mehr Personen KI-Systeme, um selber Inhalte wie Bilder, Texte oder Musik zu erstellen. Dies wirft urheberrechtliche Fragen auf, welche zurzeit noch nicht alle verbindlich beantwortet wurden. Sind beispielsweise die Ergebnisse von KI urheberrechtlich geschützt und ist das Training von KI urheberrechtlich relevant oder nicht? Es ist

[286] Vgl. hierfür vorangehend Ziff. 6.2.1.1.

[287] BGE **59** II 401, S. 405.

[288] DENIS BARRELET/WILLI EGLOFF, Das neue Urheberrecht. Kommentar (Fn. 280), Art. 2 N 8.

[289] BBl **1989** III 477, hier 521.

[290] In der Schweiz liegt diesbezüglich noch kein verbindlicher Entscheid vor. Im Ausland zeichnet sich jedoch eine entsprechende Tendenz im Immaterialgüterrecht ab, so z. B. in den USA im Urheberrecht (U.S. Copyright Office, Robert J. Kasunic, Zarya of the Dawn [Registration # VAu001480196]) oder in Deutschland im Patentrecht (Beschluss des Bundespatentgerichtes X ZB 5/22 vom 11. Juni 2024).

davon auszugehen, dass insbesondere die Antwort auf letztere Frage möglichen regulatorischen Bedarf aufzeigen wird. Sollte das Training von KI urheberrechtlich relevant sein, ist zu prüfen, wie die (Weiter-)Entwicklung von KI sichergestellt werden kann. Sollte das Training von KI hingegen nicht urheberrechtlich relevant sein, so ist zu prüfen, inwieweit den Interessen der Rechteinhaberinnen und Rechteinhabern von Werken Rechnung zu tragen ist.

6.2.2 KI und Patentrecht

6.2.2.1 Allgemeines

Das internationale Patentrecht ist in zahlreichen Übereinkommen geregelt.[291] Für die Schweiz ist das Europäische Patentübereinkommen[292] massgebend.[293] Auf nationaler Ebene regeln das Patentgesetz (PatG) und die Patentverordnung (PatV)[294] das Patentrecht.

Das Patentrecht schützt technische Lösungen für technische Probleme, sofern diese Lösungen neu, erfinderisch und gewerblich anwendbar sind (Art. 1 PatG). Abstrakte Ideen ohne konkrete technische Schritte, insbesondere mathematische Methoden, Algorithmen oder Lernverfahren, sind nicht patentierbar.

Das Patentrecht gilt für alle Gebiete der Technik.[295] Es ist insbesondere auf computerimplementierte Erfindungen[296] und auf KI-gestützte Erfindungen anwendbar. Die Weltorganisation

[291] Die Schweiz hat insbesondere die Pariser Übereinkunft zum Schutz des gewerblichen Eigentums (SR **0.232.04**), den Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (SR **0.232.141.1**), den Patentrechtsvertrag (SR **0.232.141.2**) und das Abkommen über handelsbezogene Aspekte der Rechte an geistigem Eigentum (Anhang 1C des Abkommens zur Errichtung der Welthandelsorganisation, SR **0.632.20**) ratifiziert.

[292] SR **0.232.142.2**.

[293] Das Europäische Patentamt (EPA) ist eine von der EU unabhängige internationale Organisation, die ein zentralisiertes Patentprüfungsverfahren für 39 Länder anbietet. Die meisten Patentanmeldungen, die die Schweiz betreffen, werden direkt beim EPA eingereicht. Sobald das EPA ein Patent erteilt, ist es auch in der Schweiz gültig.

[294] SR **232.14** und SR **232.141**.

[295] Das PatG kennt einige Ausnahmen von den Artikel 1*a*, 1*b* und 2 PatG. Dazu gehören beispielsweise der menschliche Körper und seine Bestandteile, Gensequenzen, chirurgische, therapeutische oder diagnostische Verfahren, Pflanzensorten und Tierrassen. Nicht patentierbar ist ferner eine Erfindung, deren Verwendung die Menschenwürde verletzen oder die Würde der Kreatur missachten, oder in anderer Weise gegen die öffentliche Ordnung oder die guten Sitten verstossen würde (z. B. Folterwerkzeuge).

[296] Computerprogramme sind als solche nicht patentierbar. Sie können jedoch Teil einer Erfindung sein, wenn diese eine Wirkung hat, welche über das blosse Funktionieren des Programmes hinausgeht.

für geistiges Eigentum (WIPO) verzeichnete zwischen 2016 und 2022 einen Anstieg der Patentanmeldungen im Bereich der KI um 718 Prozent.[297] Aufgrund des Anstiegs der Anmeldungen in diesem Bereich hat das Europäische Patentamt (EPA) seine Prüfungsrichtlinien für KI-gestützte Erfindungen seit 2018 präzisiert.[298]

KI wird zunehmend als Werkzeug für Erfindungen im Bereich Forschung und Entwicklung eingesetzt. Dabei ermöglicht KI eine Beschleunigung, Ausweitung und Diversifizierung von Forschung und Entwicklung.

6.2.2.2 Herausforderungen

Generative KI könnte zu einer drastischen Zunahme der Menge an wissenschaftlicher und technischer Literatur führen. Diese spiegelt den Stand der Technik wider, woran Neuheit und Erfindungshöhe einer Patentanmeldung gemessen werden. Die Erlangung eines Patents könnte dadurch erschwert werden.[299]

Der Einsatz von KI in Forschung und Entwicklung dürfte die Beurteilung des Kriteriums der erfinderischen Tätigkeit für eine Erfindung verändern. Nicht jede technische Lösung oder Verbesserung einer technischen Lösung ist patentierbar. Die in der Patentanmeldung offenbarte Lösung muss erfinderisch sein, d. h. sie darf für Experten nicht naheliegend sein. KI kann komplexe Lösungen analysieren und generieren, die die Grenzen des herkömmlichen menschlichen Verständnisses überschreiten. Dies könnte den Bereich innovativer Ideen erweitern, indem Kombinationen gefunden werden, die für menschliche Fachleute nicht offensichtlich sind, und somit eine Weiterentwicklung des Kriteriums der erfinderischen Tätigkeit erzwingen.

Die Offenlegung von Trainingsdaten einer KI-basierten Erfindung wird diskutiert. Ein Patent wird erteilt, wenn die Erfindung in der Patentschrift so beschrieben wird, dass sie von einer Fachperson nachvollzogen werden kann. Bei KI-gestützten Erfindungen spielen Trainingsdaten eine wesentliche Rolle für die Reproduzierbarkeit der Erfindung. Es wird daher diskutiert, wie und in welchem Umfang diese Trainingsdaten offengelegt werden sollen. Dabei sind Verpflichtungen oder Interessen wie Datenschutz und Schutz von Geschäftsgeheimnissen oder

---

[297] https://www.wipo.int > Understand and Learn > IP in > Frontier Technologies > AI Inventions factsheet, 1 (abgerufen am 15. Mail 2024).

[298] Richtlinien für die Prüfung im Europäischen Patentamt, Teil G, 3.3.1, «Künstliche Intelligenz und maschinelles Lernen», Version 03.2024, (abrufbar unter: https://www.epo.org/de/legal/guidelines-epc/2024/g_ii_3_3_1.html, abgerufen am 15. Mai 2024); siehe auch https://www.epo.org/de/news-events/in-focus/ict/artificial-intelligence (abgerufen am 15. Mai 2024).

[299] Siehe insbesondere All Prior Art, ein Zufallsgenerator für wissenschaftliche und technische Texte, https://allpriorart.com (abgerufen am 15. Mai 2024). Das US-amerikanische-Patentamt hat übrigens am 30. April 2024 eine öffentliche Konsultation zu diesem Thema durchgeführt, Request for Comments Regarding the Impact of the Proliferation of Artificial Intelligence on Prior Art, the Knowledge of a Person Having Ordinary Skill in the Art, and Determinations of Patentability Made in View of the Foregoing, abrufbar unter: https://www.federalregister.gov/documents/2024/04/30/2024-08969/request-for-comments-regarding-the-impact-of-the-proliferation-of-artificial-intelligence-on-prior (abgerufen am 15. Mai 2024).

Geheimhaltungspflichten zu berücksichtigen. Das Europäische Patentamt hat diesen Punkt bei der letzten Aktualisierung seiner Prüfungsrichtlinien klargestellt: «*Hängt die technische Wirkung von bestimmten Merkmalen des verwendeten Trainingsdatensatzes ab, sind die zur Reproduktion dieser technischen Wirkung erforderlichen Merkmale zu offenbaren, es sei denn, der Fachmann kann sie ohne unzumutbaren Aufwand unter Verwendung seines allgemeinen Fachwissens ermitteln. Der konkrete Trainingsdatensatz selbst muss jedoch in der Regel nicht offenbart werden.*»[300] Die letzte Herausforderung betrifft das Erfordernis, in jeder Patentanmeldung eine natürliche Person als Erfinder anzugeben. Diese Frage wird durch das *Artificial Inventor Project*[301] beleuchtet, das versucht, zwei Patente im Namen von *DABUS*, einem System künstlicher Intelligenz, anzumelden, ohne dabei eine natürliche Person als Erfinder anzugeben. Dieses Vorgehen stellt diese Anforderung des Patentrechts in Frage und wirft grundsätzliche Thematiken betreffend den Zweck des Patentrechts auf.

6.2.2.3 Regulatorischer Bedarf

Ein gesetzgeberischer Handlungsbedarf scheint nicht zu bestehen. Der spektakuläre Anstieg der Anmeldungen von KI-basierten Erfindungen zwischen 2016 und 2022 belegt dies. Die Bedingungen für den Schutz dieser Erfindungen sind klar und stabil, und es werden Patente für diese Art von Erfindungen erteilt. Das System bietet zudem eine gewisse Flexibilität. Begriffe wie «Stand der Technik», «Neuheit», «erfinderische Tätigkeit» und «Erfordernis der Offenbarung der Erfindung» können von den Ämtern für geistiges Eigentum und den Gerichten bis zu einem gewissen Grad entsprechend den technologischen Entwicklungen ausgelegt werden.

Hinsichtlich des Erfordernisses einer natürlichen Person als Erfinder erscheint die durch das *Artificial Inventor Project* ausgelöste Diskussion verfrüht. Der Mensch spielt nach wie vor eine entscheidende Rolle in Forschung und Entwicklung, und es ist nach wie vor möglich, zumindest eine natürliche Person als Erfinder zu benennen.

- Der Einsatz von KI in Forschung und Entwicklung könnte zu einer Weiterentwicklung der Konzepte von Neuheit und erfinderischer Tätigkeit führen.
- Es sollte eine klare und stabile Praxis für die Offenbarung von Trainingsdaten für KI-basierte Erfindungen entwickelt werden.
- Die weltweite Entwicklung der Rechtsprechung hinsichtlich des Erfordernisses der Nennung einer natürlichen Person als Erfinder sollte aufmerksam verfolgt werden.

---

[300] Vgl. Fn. 298.

[301] The Artificial Inventor Project, https://artificialinventor.com/ (abgerufen am 15. Mai 2024). Dies ist ein Test des Patentsystems im echten Leben. Der Anmelder behauptet, ein KI-System geschaffen zu haben, das selbstständig Erfindungen generieren kann. Weltweit wurden Patentanträge gestellt. Der Fall ging bis zum Obersten Gerichtshof der USA und zum Obersten Gerichtshof Grossbritanniens. Beide sprachen sich gegen die Zulassung eines künstlichen Erfinders aus. Das EPA wies die Anmeldung zurück. In der Schweiz wies das EPA das Eintragungsgesuch ab. Eine Beschwerde ist beim Bundesverwaltungsgericht hängig.

- Gesetzgeberischer Handlungsbedarf scheint nicht zu bestehen. Das Patentrechtssystem ist flexibel genug, um sich den technologischen Entwicklungen ohne Änderungen anzupassen. Der spektakuläre Anstieg der Patentanmeldungen für KI-basierte Erfindungen seit 2016 zeigt, dass das System gut funktioniert.

6.3 Ausservertragliche Haftung

6.3.1 Allgemeines ausservertragliches Haftpflichtrecht

6.3.1.1 Vorschlag für eine EU-Richtlinie über KI-Haftung

Der Vorschlag der Europäischen Kommission vom 28. September 2022 für eine Richtlinie zur Anpassung der Vorschriften über ausservertragliche zivilrechtliche Haftung an künstliche Intelligenz (RL-Vorschlag KI-Haftung)[302] betrifft die *verschuldensabhängige ausservertragliche Haftung* und will den Herausforderungen, die KI für die Anwendung bestehender Haftungsvorschriften darstellt, mithilfe von Offenlegungsvorschriften und widerlegbaren Vermutungen begegnen. Es soll den besonderen Merkmalen bestimmter KI-Systeme, wie z. B. Undurchsichtigkeit, autonomes Verhalten und Komplexität, die es erschweren oder verunmöglichen, ein kausales Handeln oder Unterlassen von verantwortlichen Personen zu beweisen, mit Beweiserleichterungen begegnet werden. Der Vorschlag knüpft an die Begriffsbestimmungen und Kategorien des *KI-Gesetzes*[303] an und sieht für *Hochrisiko-KI-Systeme* besondere Rechtsfolgen vor, welche auch zur Durchsetzung der Anforderungen des KI-Gesetzes beitragen sollen.[304] Mit der widerlegbaren Vermutung wurde gemäss Begründung des RL-Vorschlags das am wenigsten einschneidende Instrument gewählt.[305] Der in der Entschliessung des Europäischen Parlaments vom 20. Oktober 2020[306] mit Empfehlungen an die Kommission für eine Regelung der zivilrechtlichen Haftung beim Einsatz künstlicher Intelligenz noch vorgesehene Ansatz einer Gefährdungshaftung[307] für Betreiber von Systemen mit hohen Risiken wurde nicht weiterverfolgt. Die Anwendung der Richtlinie soll fünf Jahre nach Ablauf der

[302] Vorschlag für eine Richtlinie des Parlaments und des Rates zur Anpassung der Vorschriften über aussenvertragliche zivilrechtliche Haftung an künstliche Intelligenz, COM/2022/496 final, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52022PC0496 (abgerufen am 28. August 2024).

[303] S. oben: Ziff. 5.

[304] Vorschlag Richtlinie über KI-Haftung (Fn. 302), Begründung, 3.

[305] Vorschlag Richtlinie über KI-Haftung (Fn. 302), Begründung, 7.

[306] Abl. C 404/107, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020IP0276&from=PL (abgerufen am 28. August 2024).

[307] Vgl. GERHARD WAGNER, Haftung für Künstliche Intelligenz – Eine Gesetzesinitiative des Europäischen Parlaments, ZEuP 2021, 545 ff., 556 ff.

Umsetzungsfrist geprüft werden, wobei dann allenfalls die Notwendigkeit einer verschuldensunabhängigen Gefährdungshaftung gekoppelt mit einer Pflichtversicherung zu prüfen wäre.[308] Im Einzelnen:

- **Anwendungsbereich:** Der RL-Vorschlag enthält Vorschriften über die verschuldensabhängige ausservertragliche zivilrechtliche Haftung für durch ein KI-System verursachte Schäden. Die Richtlinie gilt nicht für die Haftung aus Vertrag. Sie gilt auch nicht für die strafrechtliche Haftung (Art. 1(1)+(2) RL-Vorschlag KI-Haftung).

- **Offenlegung von Beweismitteln über Hochrisiko-KI-Systeme:** Der RL-Vorschlag schafft für diejenigen, die Schadensersatz fordern, eine Möglichkeit, *Informationen über Hochrisiko-KI-Systeme* zu erhalten, die gemäss *KI-Gesetz* aufzuzeichnen / zu dokumentieren sind (Art. 3 RL-Vorschlag KI-Haftung).[309] Die Gerichte wären auf Antrag eines potenziellen Klägers befugt, von Anbietern und Nutzern von Hochrisiko-KI-Systemen, die im Verdacht stehen, einen Schaden verursacht zu haben, die Offenlegung von Beweismitteln anzuordnen. Der potentielle Kläger muss seinen Schadenersatz plausibel machen und belegen, dass eigene Anstrengungen nicht zum Ziel geführt haben. In einer Verhältnismässigkeitsabwägung sind weiter die Interessen aller Parteien und auch Dritter, namentlich auch Geschäftsgeheimnisse, zu beachten. Kommt ein Beklagter der Anordnung nicht nach, wird ein *Sorgfaltspflichtverstoss vermutet* (Art. 3[5] RL-Vorschlag KI-Haftung).

- **Kausalitätsvermutung:** Unter bestimmten Bedingungen soll sowohl für Hochrisiko-KI-Systeme als auch für andere KI-Systeme eine *Kausalitätsvermutung* vorgesehen werden. Diese soll den Nachweis, dass eine bestimmte Eingabe, für die die potenziell haftbare Person verantwortlich ist, ein bestimmtes Ergebnis des KI-Systems verursacht hat, erleichtern. Vermutet wird ein ursächlicher Zusammenhang zwischen dem Verschulden des Beklagten und dem vom KI-System hervorgebrachten Ergebnis oder der Tatsache, dass kein Ergebnis hervorgebracht wurde. Der Vorschlag bewirkt keine Umkehr der Beweislast, es wird eine widerlegbare Vermutung geschaffen.[310] Auch entbindet sie nicht vom Nachweis des Verschuldens, das heisst im Normalfall der Verletzung einer Sorgfaltspflicht. Es wird Bezug genommen auf im Unionsrecht oder im nationalen Recht festgelegte Sorgfaltspflichten, deren unmittelbarer Zweck darin besteht, den eingetretenen Schaden zu verhindern (Art. 4[1][a] RL-Vorschlag KI-Haftung), es besteht ein direkter Bezug zum KI-Gesetz. Die Kausalitätsvermutung gilt unter folgenden Voraussetzungen:

    - Für Anbieter von Hochrisiko-KI-Systemen, wenn nachgewiesen werden kann, dass gegen Anforderungen des KI-Gesetzes in Bezug auf Training, Transpa-

[308] Vorschlag Richtlinie über KI-Haftung (Fn. 302), Begründung, 7.

[309] Vorschlag Richtlinie über KI-Haftung (Fn. 302), Erw. 17.

[310] Vorschlag Richtlinie über KI-Haftung (Fn. 302), Begründung S. 7 (Verhältnismässigkeit).

renz, Beaufsichtigung, Genauigkeit, Robustheit und Cybersicherheit verstossen wurde bzw. erforderliche Korrekturmassnahmen nicht unverzüglich ergriffen wurden (Art. 4[2] RL-Vorschlag KI-Haftung).

- Für Nutzer von Hochrisiko-KI-Systemen, wenn nachgewiesen werden kann, dass gegen Anforderungen des KI-Gesetzes in Bezug auf Verwendung, Überwachung oder Zweckbestimmung der Eingabedaten verstossen wurde (Art. 4[3] RL- Vorschlag KI-Haftung).

- Bei Nicht-Hochrisiko-KI-Systemen gilt die Vermutung, wenn es nach Auffassung des Gerichtes für den Kläger übermässig schwierig ist, den ursächlichen Zusammenhang zwischen Schaden und Verschulden zu beweisen.

Der Richtlinienvorschlag der EU-Kommission befindet sich aktuell (Stand 31.08.2024) im Rechtsausschuss des Europäischen Parlaments.[311] Dieser hat laut Presseberichten eine Zusatzstudie zur Notwendigkeit dieser zusätzlichen Richtlinie – neben dem KI-Gesetz und der revidierten Richtlinie über die Haftung für fehlerhafte Produkte[312], welche beide im März 2024 vom Parlament verabschiedet wurden – in Auftrag gegeben.[313] Die Zukunft des Richtlinienvorschlags KI-Haftung scheint damit ungewiss.

6.3.1.2 Schweizer Recht

Allgemein ist das Schweizer Zivilrecht und namentlich auch das Haftpflichtrecht mit seinen offenen Generalklauseln in der Lage, technische Entwicklungen aufzufangen und stellt den Gerichten ein Instrumentarium zur Verfügung, um im Einzelfall zu gerechten Lösungen zu gelangen.[314] Der Bundesrat hat sich in der Vergangenheit deshalb verschiedentlich gegen Rechtsänderungen aufgrund von technischen Innovationen ausgesprochen, da diese die Gefahr mit sich bringen, innert Kürze zu veralten.[315] Eine zu stark auf technische Einzelheiten

---

311 2022/0303(COD): Adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (nicht auf Deutsch verfügbar), https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2022%2F0303(COD)&l=fr (abgerufen am 26. August 2024).

312 S. dazu sogleich Ziff. 6.3.2.

313 https://www.euractiv.com/section/digital/news/picking-up-the-ai-liability-directive-after-the-tech-policy-spree/ (abgerufen am 26. August 2024)

314 Vgl. Bericht «Herausforderungen der künstlichen Intelligenz» (Fn. 1), 36 f.

315 Vgl. zu den Grundsätzen der Politik des Bundes im Umgang mit neuen Technologien Bericht «Herausforderungen der künstlichen Intelligenz» (Fn. 1), 34 ff.; Bericht des Bundesrates «Rechtliche Grundlagen für Distributed Ledger-Technologie und Blockchain in der Schweiz» vom 14. Dezember 2018, 13 ff., abrufbar unter: https://www.newsd.admin.ch/newsd/message/attachments/55150.pdf (abgerufen am 31. Juli 2024); vgl. auch Bericht des Bundesrates «Die zivilrechtliche Verantwortlichkeit von Providern» vom 11. Dezember 2015, 99 ff. abrufbar unter:

fokussierte Gesetzgebung läuft Gefahr, Regelungslücken gerade erst zu schaffen. Dennoch ist es im Zuge der Digitalisierung angezeigt, die bestehenden Rechtsgrundlagen regelmässig auf ihre Komptabilität mit neueren technischen Entwicklungen zu prüfen sowie allenfalls bestehende Lücken zu schliessen bzw. Rechtssicherheit zu schaffen.

Die Schweiz kennt keine den diskutierten EU-Regeln vergleichbaren spezifischen haftpflichtrechtlichen Regeln zur künstlichen Intelligenz. Nach der haftpflichtrechtlichen Generalklausel von Artikel 41 Absatz 1 OR ist haftbar und schadenersatzpflichtig, wer einem andern mit Absicht oder aus Fahrlässigkeit widerrechtlich Schaden zufügt. Notwendige Voraussetzungen der Haftung sind daher Schaden, Widerrechtlichkeit, Kausalität und Verschulden. Nach Artikel 8 des Zivilgesetzbuches (ZGB)[316] muss derjenige, der aus einer Tatsache Rechte ableitet, das Vorhandensein dieser Tatsache beweisen, gesetzliche Ausnahmen vorbehalten. Nach Artikel 55 Absatz 1 der Zivilprozessordnung (ZPO)[317] trifft diese Partei die sogenannte Behauptungs- und Substantiierungslast. Sie muss dem Gericht einen Beweisantrag stellen, in dem die Beweise bestimmt oder klar bestimmbar sind (Art. 152 Abs. 1 ZPO).[318] Reine Ausforschungsanträge («fishing expeditions») sind nicht zulässig. Die Gegenpartei ist zur Mitwirkung verpflichtet und muss dabei namentlich auch genau bezeichnete Urkunden herausgeben (Art. 160 Abs. 1 Bst. b ZPO). Ein Verweigerungsrecht, um sich selbst vor zivilrechtlicher Verantwortlichkeit zu schützen, besteht nicht.[319] Verweigert eine Partei die Mitwirkung unberechtigterweise, so berücksichtigt dies das Gericht bei der Beweiswürdigung (Art. 164 ZPO). Es ergibt sich daraus jedoch nicht ohne weiteres eine Tatsachenvermutung zu Ungunsten der verweigernden Partei.[320] Unter Umständen kann zur Abklärung von Prozesschancen auch eine vorsorgliche Beweisabnahme verlangt werden (Art. 158 Abs. 1 Bst. b ZPO).[321]

In gewissen Fällen sind auch Beweiserleichterungen denkbar. Gemäss Rechtsprechung und Lehre reicht eine überwiegende Wahrscheinlichkeit[322] aus, wenn «ein strikter Beweis nicht nur im Einzelfall, sondern der Natur der Sache nach nicht möglich oder nicht zumutbar ist und

---

www.bj.admin.ch > Publikationen & Service > Berichte, Gutachten und Verfügungen > Bericht und Gutachten > Die Zivilrechtliche Verantwortlichkeit von Providern (abgerufen am 31. Juli 2024).

316 SR **210**.

317 SR **272**.

318 BSK ZPO-GUYAN, Art. 152 N 3 f.

319 Botschaft ZPO vom 28. Juni 2006, BBl **2006** 7221 7317.

320 BGE **140** III 264, E. 2.3.

321 BSK ZPO-GUYAN, Art. 158 N 5.

322 «Nach dem Beweismass der überwiegenden Wahrscheinlichkeit [...] gilt ein Beweis als erbracht, wenn für die Richtigkeit der Sachbehauptung nach objektiven Gesichtspunkten derart gewichtige Gründe sprechen, dass andere denkbare Möglichkeiten vernünftigerweise nicht massgeblich in Betracht fallen»; BGE **132** III 715, E. 3.1. m.w.H.

insofern eine ‹Beweisnot› besteht.»[323] Hierzu gehört auch der Nachweis des Kausalzusammenhangs im Haftpflichtrecht.[324]

Diese Regeln gelten auch für Verfahren, in denen die Verhandlungsmaxime im engeren Sinne gilt. Im vereinfachten Verfahren, das bei Streitigkeiten mit einem Streitwert bis zu 30 000 Franken zur Anwendung kommt (Art. 243 Abs. 1 ZPO), ist die Rolle des Gerichts aktiver, da ihm eine verstärkte Fragepflicht obliegt (Art. 247 Abs. 1 ZPO).

6.3.1.3 Würdigung

Da sich die Anforderungen des Vorschlags einer Richtlinie über KI-Haftung an den Vorgaben des KI-Gesetzes ausrichten, ist eine isolierte Einführung der Richtlinie über KI-Haftung nicht denkbar. Sollte das KI-Gesetz für die Schweizer Rechtsordnung übernommen werden, wären die begleitenden Bestimmungen des Vorschlags der Richtlinie über KI-Haftung der Schweizer Rechtsordnung aber nicht komplett fremd. Die Offenlegung einer genau bezeichenbaren, bei der Gegenpartei vorhandenen Dokumentation ist unter Umständen mit den vorhandenen zivilprozessualen Instrumenten bereits nach geltendem Recht erreichbar. Die Rechtsfolgen bei Verweigerung der Mitwirkung bzw. Offenlegung von Beweismitteln ist ebenfalls vergleichbar, die im Schweizer Recht vorgesehene Berücksichtigung bei der Beweiswürdigung geht aber weniger weit als die im Richtlinienvorschlag vorgesehene Vermutung der Sorgfaltspflichtverletzung. Die Kausalitätsvermutungen des Vorschlags einer Richtlinie über KI-Haftung lassen sich aus dem geltenden Schweizer Recht jedoch nicht herleiten, auch wenn Beweiserleichterungen in solchen Fällen vorgesehen sind. Eine Übernahme der Richtlinie über KI-Haftung könnte damit die gerichtliche Durchsetzung von zivilrechtlichen Ansprüchen eventuell erleichtern und auch dazu beitragen, die Vorgaben der KI-Konvention des Europarats (namentlich Art. 8, 9 und 14 ; vgl. Ziff. 4.3) im privaten Sektor besser umzusetzen. Es bleiben aber die zusätzlichen Analysen und Ergebnisse der Diskussionen in der EU abzuwarten.

Anzumerken bleibt, dass der Anwendungsbereich dieser allgemeinen ausservertraglichen haftpflichtrechtlichen Regel wohl stark beschränkt wäre.[325] Für zahlreiche mögliche Einsatz-

---

[323] BGE **132** III 715, E. 3.1. m.w.H.

[324] BGE **132** III 715, E. 3.2. m.w.H. «Glaubhaft gemacht ist dabei eine Tatsache schon dann, wenn für deren Vorhandensein gewisse Elemente sprechen, selbst wenn das Gericht noch mit der Möglichkeit rechnet, dass sie sich nicht verwirklicht haben könnte»; BGE **132** III 715, E. 3.1. m.w.H.

[325] S. ISABELLE WILDHABER, KI und Haftung: Lösungsansätze für die Schweiz, Jusletter IT 4. Juli 2024, Rz. 52 ff.; s. auch schon: ISABELLE WILDHABER, Eine Einführung in die ausservertragliche Haftung für Künstliche Intelligenz (KI), in: HAVE (Hrsg.), Haftpflichtprozess 2021, Zürich 2021, 1 ff., 45; BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie aus schweizerischer Sicht, SJZ 2023, 627 ff., 637.

bereiche von KI-Anwendungen bestehen nämlich bereits nach geltendem Recht verschuldensunabhängige Gefährdungshaftungen, welche potentiell geschädigte Personen besser zu schützen vermögen, namentlich:

- die Haftpflicht des Motorfahrzeughalters nach Artikel 58 SVG,
- die Haftpflicht des Halters eines Luftfahrzeugs nach Artikel 64 Luftfahrtgesetz (LFG),[326]
- die Herstellerhaftung nach dem Bundesgesetz vom 18. Juni 1993 über die Produktehaftpflicht (PrHG)[327] (s. dazu das folgende Kapitel Ziff. 6.3.2).

In wichtigen praktischen Anwendungsgebieten der künstlichen Intelligenz (selbstfahrende Autos, Drohnen) bestehen damit schon heute Haftungsregeln, die den Schutz von Geschädigten sicherstellen. Durch die Haftpflicht der Halter der Fahrzeuge – verbunden mit einer Versicherungspflicht (Art. 63 SVG; Art. 70 LFG) – können Haftungslücken ausgeschlossen werden.

6.3.2 Produktehaftpflichtrecht

Die Produktehaftpflicht sieht eine verschuldensunabhängige Haftung der Herstellerinnen und Hersteller für durch fehlerhafte Produkte verursachte Sach- und Personenschäden vor. Geregelt werden somit nur die Folgeschäden von fehlerhaften Produkten und nicht die Schäden am Produkt selbst.

6.3.2.1 Revidierte EU-Richtlinie über die Haftung für fehlerhafte Produkte

Die am 12. März 2024 vom Parlament verabschiedete Richtlinie[328] über die Haftung für fehlerhafte Produkte wird die bisherige Richtlinie 85/374/EWG[329] vom 25. Juli 1985 zur Angleichung der Rechts- und Verwaltungsvorschriften der Mitgliedstaaten über die Haftung für fehlerhafte Produkte («Produkthaftungsrichtlinie») ersetzen. Sie geht zurück auf den Vorschlag

---

[326] SR **748.0**.

[327] SR **221.112.944**.

[328] Legislative Entschliessung des Europäischen Parlaments vom 12. März 2024 zu dem Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates über die Haftung für fehlerhafte Produkte, abrufbar unter: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0132_DE.html (abgerufen am 26. August 2024).

[329] Richtlinie 85/374/EWG des Rates vom 25. Juli 1985 zur Angleichung der Rechts- und Verwaltungsvorschriften der Mitgliedstaaten über die Haftung für fehlerhafte Produkte, ABl. L 210 vom 7. August 1985, 29–33, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=celex%3A31985L0374 (abgerufen am 26. August 2024).

der Europäischen Kommission vom 28. September 2022 für eine Richtlinie[330] über die Haftung für fehlerhafte Produkte. Die revidierte Richtlinie sieht umfangreiche Neuerungen vor, um *Software einschliesslich KI-Systeme* der verschuldensunabhängigen Herstellerhaftung zu unterstellen. Obwohl die revidierte Produkthaftungsrichtlinie als Teil des «KI-Pakets» vorgeschlagen wurde, geht es im Kern nicht um KI-Anwendungen, sondern um eine allgemeine Modernisierung der Herstellerhaftung für fehlerhafte Produkte im Zuge der fortschreitenden Digitalisierung. Damit soll der Entwicklung Rechnung getragen werden, dass die Hersteller die Kontrolle oder Einflussnahme auf ihre Produkte heute mit Inverkehrbringen nicht zwangsläufig verlieren.[331]

Sowohl der *Kreis der Haftenden* (Softwarehersteller, Online-Plattformen und Betreiber von KI-Systemen) als auch der *Fehlerbegriff* (fehlerhafte Upgrades und Updates, mangelhafte Cybersicherheit) werden erweitert. Auch soll die Herstellerhaftung neben Sach- und Personenschäden auf Schäden wegen dem Verlust oder der Verfälschung von Daten ausgedehnt werden. Wie in der vorgeschlagenen Richtlinie über KI-Haftung wird Beweisschwierigkeiten mit *Offenlegungspflichten* und *Kausalitätsvermutungen* begegnet. Weiter gibt es in Bezug auf Update- und Aktualisierungspflichten Schnittstellen mit zwei weiteren zivilrechtlichen EU-Richtlinien:

- *Richtlinie (EU) 2019/771*[332] vom 20. Mai 2019 über bestimmte vertragsrechtliche Aspekte des Warenkaufs (*Warenkauf-Richtlinie [EU] 2019/771*) und
- *Richtlinie (EU) 2019/770*[333] des Europäischen Parlaments und des Rates vom 20. Mai 2019 über bestimmte vertragsrechtliche Aspekte der Bereitstellung digitaler Inhalte und digitaler Dienstleistungen (*Digitale-Inhalte-und-Dienste-Richtlinie [EU] 2019/770).*

Die Bestätigung der am 12. März 2024 durch das Parlament verabschiedeten Richtlinie durch den Europäischen Rat und die Publikation im Amtsblatt stehen zur Zeit (Stand 31.08.2024)

---

[330] Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates über die Haftung für fehlerhafte Produkte, COM/2022/495 final, 28. September 2022, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52022PC0495 (abgerufen am 26. August 2024).

[331] BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie (Fn. 325), 630.

[332] Richtlinie (EU) 2019/771 vom 20. Mai 2019 über bestimmte vertragsrechtliche Aspekte des Warenkaufs, zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie 2009/22/EG sowie zur Aufhebung der Richtlinie 1999/44/EG, ABl. L 136 vom 22. Mai 2019, 28 ff.

[333] ABl. L 136 vom 22.5.2019, 1 ff.

noch aus. Nach dem Inkrafttreten haben die Mitgliedstaaten zwei Jahre Zeit, um die Richtlinie umzusetzen.[334]

6.3.2.2 Schweizer Recht

Die Schweiz hat das PrHG im Rahmen des Swisslex-Pakets eingeführt, es entspricht weitgehend der bisherigen Produkthaftungsrichtlinie 85/374/EWG. [335] Gestützt auf die Regeln des PrHG können Herstellerinnen und Hersteller für fehlerhafte Produkte haftbar gemacht werden. In der Lehre ist seit Jahren umstritten, ob auch «reine» Software, das heisst Software, die nicht mit einem körperlichen Produkt verbunden ist, als Produkt im Sinne des PrHG gelten kann.[336] Das Bundesgericht hat sich zu dieser Frage noch nicht geäussert. Ein Grossteil der jüngeren Lehre spricht sich für eine entsprechend weite, dem Wortlaut widersprechende Auslegung von Artikel 3 Absatz 1 PrHG aus, wünscht sich jedoch eine Klarstellung durch den Gesetzgeber.[337]

6.3.2.3 Würdigung

Die Unterstellung von Software unter die Herstellerhaftung des PrHG, wie sie im Vorschlag der EU-Richtlinie vorgesehen ist, wird von der Schweizer Lehre seit längerem gefordert.[338] Nicht enthalten im geltenden PrHG sind auch die weiteren Neuerungen im Vorschlag für die revidierte EU-Richtlinie, welche der zunehmenden Digitalisierung der Produkte Rechnung tragen.[339]

In seinem Bericht «Modernisierungsbedarf des Gewährleistungsrechts beim Kauf» vom 16. Juni 2023 hat der Bundesrat in Bezug auf die *Gewährleistung des Verkäufers für Sachmängel* festgehalten, dass das geltende Recht, welches auf den Einmalaustausch von Ware gegen Geld ausgerichtet ist, der Interessenlage bei digitalen Produkten oder Produkten mit digitalen Komponenten nicht gerecht wird.[340] Er hat sich deshalb für die Einführung einer Update- oder Aktualisierungspflicht für digitale Produkte und Produkte mit digitalen Komponenten nach dem Vorbild der Digitale-Inhalte-und-Dienste-Richtlinie (EU) 2019/770 und der Warenkauf-Richtlinie (EU) 2019/771 ausgesprochen. Mit den gleichlautenden Motionen 23.4316

---

[334] Art. 22 der Richtlinie über die Haftung für fehlerhafte Produkte, abrufbar unter: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0132_DE.html (abgerufen am 26. August 2024).

[335] S. Botschaft über das Folgeprogramm nach der Ablehnung des EWR-Abkommens vom 24. Februar 1993, BBl **1993** I 805 884.

[336] S. BSK OR-FELLMANN, Art. 3 PrHG N 10 m.w.H.

[337] S. nur FELLMANN, Haftpflichtrecht im Zeichen der Digitalisierung, HAVE 2021, 105 ff.; ISABELLE WILDHABER, Eine Einführung (Fn. 325), 26; BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie (Fn. 325), 638, je mit weiteren Hinweisen; für eine Unterstellung von KI-Produzenten unter die Herstellerhaftung de lege lata namentlich auch: ARIANE MORIN, L'opposabilité de la LRFP au fournisseur de l'intelligence artificielle, in: Alexandre Ri-Cha/Damiano Canapa (Hrsg.), Aspects juridiques de l'intelligence artificielle, Bern 2024, 117 ff., 120 f.

[338] S. die Nachweise oben in Fn. 337 sowie ISABELLE WILDHABER, KI und Haftung (Fn. 325), Rz. 26 ff.

[339] S. zum Anpassungsbedarf des PrHG im Einzelnen: BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie (Fn. 325), 637 ff.

[340] Bericht des Bundesrates «Modernisierungsbedarf des Gewährleistungsrechts beim Kauf» vom 16. Juni 2023, Ziff. 4.1 ff., abrufbar unter: https://www.newsd.admin.ch/newsd/message/attachments/79587.pdf (abgerufen am 31. Juli 2024).

und 23.4345 der Rechtskommissionen beider Räte wurde der Bundesrat in der Wintersession 2023 beauftragt, eine entsprechende Vorlage auszuarbeiten.[341]

Der Bundesrat und das Parlament haben sich bislang nicht zu einer allfälligen Revision des PrHG geäussert.

6.3.3 Schutz der Persönlichkeit

Im Zusammenhang mit KI-Anwendungen sind schliesslich auch die Mittel des allgemeinen zivilrechtlichen Persönlichkeitsschutzes zu thematisieren. Diese sind beim praktisch wichtigen Phänomen der *Deepfakes*[342] besonders relevant. Eine in ihrer Persönlichkeit verletzte Person kann zu ihrem Schutz gegen jeden, der an der Verletzung mitwirkt, das Zivilgericht anrufen (Art. 28 Abs. 1 ZGB). Auf diesem Weg kann die betroffene Person namentlich verlangen, dass eine bestehende Persönlichkeitsverletzung beseitigt oder eine drohende Persönlichkeitsverletzung verboten wird (Art. 28*a* Abs. 1 Ziff. 1 und 2 ZGB). Diese Regelung ist technologieneutral.[343] Es handelt sich bei der Persönlichkeit (inkl. Recht am eigenen Bild und der eigenen Stimme) um ein absolut geschütztes Rechtsgut. Artikel 28 ZGB enthält ein «stillschweigendes, allgemeines Verbot unbefugter Beeinträchtigungen fremder Persönlichkeitssphären».[344] Persönlichkeitsverletzungen mittels KI-Anwendung werden von den geltenden Regeln deshalb ohne Weiteres erfasst. Von Bedeutung ist, dass Bildmanipulationen auch schon vor der Verbreitung von KI-Anwendungen existierten. So hatte sich das Bundesgericht auch in der Vergangenheit bereits mit Fotomontagen und Bildmanipulationen zu befassen, wobei dort nicht in Frage gestellt wurde, dass auch manipulierte Bilder die Persönlichkeitsrechte der Betroffenen tangieren können.[345] Dies wurde auch in einem aktuellen erstinstanzlichen Urteil zu einem mittels KI erstellen Video gestützt: Ein Politiker verwendete zu Wahlkampfzwecken ein mit einer KI-Anwendung erstelltes, gefälschtes Video einer politischen Konkurrentin. Das Basler Zivilgericht stufte dies als Persönlichkeitsverletzung ein.[346] Gestützt auf Artikel 28*a* Absatz 1 Ziffern 1 und 2 ZGB kann sowohl das Löschen als auch das Sperren von rechtswidrigen Inhalten im Internet angeordnet werden. Auch Klagen auf Schadenersatz, Genugtuung sowie Gewinnherausgabe sind denkbar (Art. 28*a* Abs. 3 ZGB).

6.3.4 Fazit

Allgemein ist das Schweizer Zivilrecht und namentlich auch das Haftpflichtrecht mit seinen offenen Generalklauseln in der Lage, technische Entwicklungen aufzufangen und stellt den

[341] 23.4316 Mo. RK-S «Modernisierung des Gewährleistungsrechts» und 23.4345 Mo. RK-N «Modernisierung des Gewährleistungsrechts».

[342] Vgl. 23.3563 Mo. Mahaim «Deepfakes regulieren».

[343] S. zur Technologieneutralität der Schweizer Gesetzgebung und der Erfassung von Deepfakes im Strafrecht auch unten: Ziff. 6.6.

[344] HEINZ HAUSHEER/REGINA E. AEBI-MÜLLER, Das Personenrecht des Schweizerischen Zivilgesetzbuches, 5. Aufl., Bern 2020, Rz. 490 ff., 544.

[345] Vgl. Urteil des BGer 5A_553/2012 vom 14. April 2014; Urteil des BGer 5A_376/2013 vom 29. Oktober 2013.

[346] Medialex newsletter 01/24, abrufbar unter: https://medialex.ch/2024/02/07/newsletter-01-24/ (abgerufen am 31. Juli 2024).

Gerichten ein Instrumentarium zur Verfügung, um im Einzelfall zu gerechten Lösungen zu gelangen. Durch die bestehenden Gefährdungshaftungen und Versicherungspflichten im Strassenverkehr und der Luftfahrt können zudem Haftungslücken in zentralen Anwendungsbereichen ausgeschlossen werden. Auch die gerichtliche Durchsetzung der Ansprüche ist auf Grundlage der bestehenden Regeln grundsätzlich möglich. Eine Übernahme der diskutierten EU-Richtlinie über KI-Haftung könnte aber die gerichtliche Durchsetzung von zivilrechtlichen Ansprüchen eventuell erleichtern und auch dazu beitragen, die Vorgaben der KI-Konvention des Europarats (namentlich Art. 8, 9 und 14, siehe Kapitel 4.3) im privaten Sektor besser umzusetzen. Eine isolierte Einführung der Richtlinie – ohne gleichzeitige Übernahme des KI-Gesetzes – ist jedoch nicht denkbar. Auch bleiben die zusätzlichen Analysen und Ergebnisse der Diskussionen in der EU zur Richtlinie über KI-Haftung – deren Zukunft ungewiss ist – abzuwarten.

Aufgrund der technischen Entwicklungen der Produkte – nicht nur, aber auch im Zusammenhang mit KI – zeichnet sich ein allgemeiner Modernisierungsbedarf in Bezug auf das Produktehaftpflichtgesetz ab.

6.4 Allgemeines Vertragsrecht

6.4.1 UNCITRAL Model Law on Automated Contracting

Die UN Commission on International Trade Law (UNCITRAL) hat im Juli 2024 das UNCITRAL Model Law on Automated Contracting verabschiedet.[347] Dieses wurde zuvor in der «Working Group IV: Electronic Commerce» der UNCITRAL – unter aktiver Mitwirkung der Schweiz – ausgearbeitet.[348] Die Bestimmungen des Modellgesetzes sehen vor, dass die Gültigkeit von Verträgen nicht allein aufgrund des Umstandes verneint werden darf, dass an dessen Abschluss ein automatisiertes System beteiligt war. Weitere Regeln betreffen die Zuordnung von durch automatisierte Systeme generierten Nachrichten. Eine Übernahme des Modellgesetzes für die Schweiz, bzw. ob dieses durch das geltende Recht bereits erfüllt wird, wird zu gegebener Zeit geprüft werden.

6.4.2 Schweizer Recht

6.4.2.1 Zurechnung von Willenserklärungen und vertragliche Haftung

Vertragliche Bindung setzt einen tatsächlichen oder *normativen* Konsens voraus, einen ausdrücklich oder *vertrauenstheoretisch* erklärten Rechtsfolgewillen. Entscheidend ist, ob ein

[347] S. die Medienmitteilung unter: https://unis.unvienna.org/unis/pressrels/2024/unisl362.html (abgerufen am 26. August 2024).

[348] Sämtliche Arbeitsdokumente sind abrufbar unter: https://uncitral.un.org/en/working_groups/4/electronic_commerce (abgerufen am 26. August 2024).

Verhalten an den Tag gelegt wurde, aus dem die Gegenpartei in guten Treuen auf das Vorhandensein eines solchen Willens schliessen durfte.[349] Schon heute werden im Online-Bereich Verträge durch Einsatz von Software mit einem hohen Automationsgrad abgeschlossen, ohne dass sich darauf Probleme bei der Zuordnung von Willenserklärungen ergäben. Vereinzelt wird in der Literatur diskutiert, ob sich in Zukunft bei zunehmender Autonomie von KI-Anwendungen Regelungslücken ergeben könnten, wenn sich beispielsweise Willenserklärungen nicht mehr klar dem Verantwortungsbereich einer Vertragspartei zuordnen lassen.[350]

Wird ein Anspruch aus Vertragsverletzung geltend gemacht, besteht insofern eine Erleichterung für geschädigte Personen, als sie das Verschulden der vertragsverletzenden Person nicht nachweisen müssen. Vielmehr muss die Person, welche den Vertrag nicht oder nicht gehörig erfüllt hat, nachweisen, dass ihr kein Verschulden zur Last fällt (Art. 97 Abs. 1 OR). Auch hier wird in der Lehre aber bereits diskutiert, ob man sich inskünftig bei der Fehlfunktion einer KI-Anwendung (zu) leicht durch den Nachweis, dass keine eigene Sorgfaltspflichtverletzung begangen wurde, exkulpieren könnte. Zur Diskussion gestellt wird insofern eine Angleichung an die Hilfspersonenhaftung nach Artikel 101 Absatz 1 OR, welche eine Zuordnung von Fehlern autonomer Anwendungen an die Partei, welche die Anwendung einsetzt, erlauben würde.[351] Es ist gegenwärtig jedoch nicht davon auszugehen, dass hier tatsächlich eine Regelungslücke besteht, lässt sich im Einzelfall doch beispielsweise auch der Einsatz einer KI-Anwendung als Sorgfaltspflichtverletzung qualifizieren.[352] Es darf auch nicht ausser Acht gelassen werden, dass KI-Anwendungen – im Gegensatz zu Personen, die zur Erfüllung einer vertraglichen Pflicht eingesetzt werden – keine eigene Rechtspersönlichkeit haben. Letztlich handelt es sich dabei um technische Hilfsmittel, die von einer Vertragspartei eingesetzt werden und deren Risikosphäre zuzuordnen sind. Eine vorschnelle Analogie zu Stellvertreter- oder Hilfspersonenregeln könnte Haftungslücken entstehen lassen und scheint Stand heute nicht sachgerecht.

6.4.2.2 Smart Contracts

Der Bundesrat hat sich in seinem Bericht «Rechtliche Grundlagen für Distributed Ledger-Technologie und Blockchain in der Schweiz – Eine Auslegeordnung mit Fokus auf dem Finanzsektor» vom 14. Dezember 2018 zu *Smart Contracts* geäussert.

Ein Smart Contract ist ein Computerprotokoll, meist basierend auf einem dezentralen Blockchainsystem, das die automatisierte Vertragserfüllung zwischen zwei oder mehr Parteien mit vorgängig codierten Angaben ermöglicht.[353] Laut dem Erfinder des Konzepts ist die einfachste Form eines Smart Contract der Verkaufsautomat, der die Ware freigibt, sobald der

---

[349] BGE **116** II 695 E. 2.

[350] Vgl. MICHAEL MARTIN KIANIČKA, Die Agentenerklärung. Elektronische Willenserklärungen und künstliche Intelligenz als Anwendungsfalls der Rechtsscheinhaftung, Zürich 2012; MALTE GRÜTZMACHER/JÖRN HECKMANN, Autonome Systeme und KI – vom vollautomatisierten zum autonomen Vertragsschluss? Die Grenzen der Willenserklärung, Computer und Recht 2019, 553 ff.

[351] CHRISTAPOR YACOUBIAN, Digitale Systeme als «Erfüllungsgehilfen» - Relevanz der fehlenden Rechtsfähigkeit, AJP 2023, 412 ff.; CHRISTAPOR YACOUBIAN, Der Analogieschluss zu Art. 101 Abs. 1 OR und die Grenzen zulässiger Rechtsfortbildung, AJP 2024, 195 ff.; MELINDA F. LOHMANN/THERESA PRESSLER, Algorithmische Vertragserfüllung (Teil 1) – Eine zukunftsgewandte Betrachtung der Rechtsfiguren der Erfüllungsgehilfin und der Substitutin unter Analyse des Urteils BGer 4A_305/2021 vom 2. November 2021, SJZ 2023, 879 ff., 884 ff.

[352] Vgl. CORINNE-WIDMER-LÜCHINGER, Apps, Algorithmen und Roboter in der Medizin: Haftungsrechtliche Herausforderungen, HAVE 2019, 3 ff.

[353] Definition in Anknüpfung an die relativ einheitliche Lehre, siehe u. a. LEE BACON/GEORGE BAZINAS, « Smart Contracts »: The next big Battleground?, Jusletter IT 18. Mai 2017, 2; MARKUS KAULARTZ/JÖRG HECKMANN, Smart contracts – Anwendungen der Blockchain-Technologie, Computer und Recht 2016, 618 ff., 618; STEPHAN D. MEYER/BENEDIKT SCHLUPPI, « Smart Contracts » und deren Einordnung in das schweizerische Vertragsrecht, recht 2017, 204 ff., 207; ROLF H. WEBER, Smart Contracts: Vertrags- und verfügungsrechtlicher Regelungsbedarf?, sic! 2018, 291 ff. 291 f.

Preis bezahlt worden ist.[354] Trotz seiner Bezeichnung ist ein Smart Contract, so ist sich die Lehre weitgehend einig, kein Vertrag im Sinne des Obligationenrechts, sondern eine Vertragserfüllungs-«Technologie».[355] Probleme können sich durch die Starrheit der vorprogrammierten Vertragsabwicklung geben, welche die im Schweizer Recht vorgesehenen Anpassungsmöglichkeiten von Verträgen im Einzelfall – beispielsweise durch die *clausula rebus sic stantibus* – verunmöglichen. Nach heutigem Stand empfiehlt die Lehre Parteien beim Abschluss eines Smart Contract geeignete Mechanismen für möglicherweise wechselnde Lebensumstände sowie zur Streitbeilegung vorzusehen.[356] Die im Blockchainkontext verbreitete Anonymität der Parteien stellt ein grosses faktisches Hindernis für die gerichtliche Durchsetzung von allfälligen Ansprüchen dar. Dieses Problem kann gesetzgeberisch jedoch nicht gelöst werden. Der Bundesrat ist in seinem Bericht «Rechtliche Grundlagen für Distributed Ledger-Technologie und Blockchain in der Schweiz» zum Schluss gelangt, dass es im Bereich Smart Contract sicher weitere Entwicklungen geben wird, die aber erst angelaufen sind und keine Regulierung aufdrängen.

6.4.3 Würdigung

Die Diskussionen zum Einsatz von KI-Anwendungen in Vertragsbeziehungen stehen noch am Anfang, auch haben sich in der Praxis bislang soweit ersichtlich keine Probleme ergeben. Die Entwicklungen sind auf jeden Fall weiterzuverfolgen, ein Eingreifen des Gesetzgebers schiene jedoch zum jetzigen Zeitpunkt nicht angezeigt.

6.5 Arbeitsrecht

6.5.1 Einleitung

In diesem Kapitel werden arbeitsrechtliche Fragen vertieft. Der Einsatz von KI in der Personalverwaltung hat stark zugenommen und ist heute in vielen Unternehmen üblich. Dies wirft neue und spezifische rechtliche Fragen auf, insbesondere im Hinblick auf die zahlreichen Vorschriften zum Schutz der Arbeitnehmerinnen und Arbeitnehmer, die für diesen Bereich kennzeichnend sind. Diese Entwicklungen haben bereits ihren Niederschlag in spezifischen Regelungen oder Gesetzesinitiativen auf EU-Ebene oder in der Schweiz gefunden.

---

[354] NICK SZABO, Formalizing and securing relationships on public networks, First Monday Internet Journal, 1997, 1.

[355] ANDREAS FURRER, Die Einbettung von Smart Contracts in das schweizerische Privatrecht, Anwalts Revue 2018, 103 ff., 109; GABRIEL OLIVIER BENJAMIN JACCARD, Smart contracts and the role of Law, Jusletter 23. November 2017, Rz. 8–9; STEPHAN D. MEYER/BENEDIKT SCHLUPPI, Smart Contracts (Fn. 353), 208; ROLF H. WEBER, Leistungsstörungen und Rechtsdurchsetzung bei Smart Contracts: Eine Auslegeordnung möglicher Problemstellungen, Jusletter 4. Dezember 2017, Rz. 2. Zu den Kontroversen: HANS RUDOLF TRÜEB, Smart Contracts, in: Pascal Grolimund u. a. (Hrsg.), Festschrift für Anton K. Schnyder, Zürich 2018, 723 ff., 725.

[356] Siehe nur ROLF H. WEBER, Leistungsstörungen (Fn. 355), Rz. 33 ff.

6.5.2 Auf europäischer Ebene

6.5.2.1 Richtlinie zur Verbesserung der Arbeitsbedingungen in der Plattformarbeit

Dem BJ sind keine Gesetzgebungsvorhaben auf EU-Ebene bekannt, die sich allgemein mit dem Einsatz von KI im Arbeitsverhältnis befassen. Hingegen ist die EU dabei, eine Richtlinie zur Verbesserung der Arbeitsbedingungen in der Plattformarbeit zu verabschieden, die auch Regeln für den Einsatz von Algorithmen in diesem Zusammenhang enthält.[357] Die Verhandlungsführer des Parlaments und des Rates haben am 8. Februar 2024 eine politische Einigung über den Text erzielt.[358] Der Rat der EU hat diese Einigung am 11. März 2024[359] und das Europäische Parlament am 24. April 2024[360] bestätigt. Für die endgültige Verabschiedung des Textes fehlt nur noch die förmliche Annahme durch den Ministerrat und die Unterzeichnung durch die Mitgesetzgeber.[361]

Ziel der Richtlinie ist die Verbesserung der Arbeitsbedingungen und des Schutzes von personenbezogenen Daten von Plattformbeschäftigten, insbesondere durch Förderung von Transparenz, Fairness und Rechenschaftspflicht beim algorithmischen Management im Kontext der Plattformarbeit (Art. 1 Abs. 1 Bst. b). Sie enthält ferner Vorschriften zur Verbesserung des Schutzes natürlicher Personen bei der Verarbeitung ihrer personenbezogenen Daten, indem sie Massnahmen für das algorithmische Management vorsieht (Art. 1 Abs. 2). Die Richtlinie gilt für Plattformtätigkeiten, die in der Europäischen Union erbracht werden, unabhängig vom Niederlassungsort der Plattform und unabhängig von dem ansonsten anwendbaren Recht (Art. 1 Abs. 3). Plattformen mit Sitz in der Schweiz könnten daher zur Anwendung dieser Richtlinie verpflichtet sein, wenn die Plattformbeschäftigten in der EU Leistungen für die Schweiz erbringen. Die Arbeit auf einer digitalen Plattform umfasst nach der Definition der Richtlinie Dienstleistungen, die auf elektronischem Weg, wie über eine Website oder eine mobile Anwendung angeboten werden (Art. 2 Abs. 1 Nr. 1 Bst. a). Unternehmen, die ihre Arbeit

---

[357] Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates zur Verbesserung der Arbeitsbedingungen in der Plattformarbeit, 9. Dezember 2021, KOM(2021) 762 endgültig/final, abrufbar unter: https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A52021PC0762 (abgerufen am 28. August 2024).

[358] Vgl. https://www.europarl.europa.eu/news/en/press-room/20240205IPR17417/provisional-deal-on-first-eu-wide-rules-for-platform-workers (abgerufen am 28. August 2024) und https://data.consilium.europa.eu/doc/document/ST-7212-2024-ADD-1/de/pdf (abgerufen am 28. August 2024).

[359] Vgl. https://www.consilium.europa.eu/de/press/press-releases/2024/03/11/platform-workers-council-confirms-agreement-on-new-rules-to-improve-their-working-conditions/ (abgerufen am 28. August 2024).

[360] Vgl. https://www.europarl.europa.eu/news/de/press-room/20240419IPR20584/le-parlement-adopte-la-directive-sur-le-travail-des-plateformes (abgerufen am 28. August 2024).

[361] Das Parlament hat am 4. Juli 2024 eine Berichtigung angenommen (deutsche Fassung vom 9. Juli 2024 abrufbar unter: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0330-FNL-COR01_DE.pdf), mit redaktionellen oder formalen Änderungen an dem am 24. April 2024 angenommenen Text. Die Unterschiede zu dem am 24. April angenommenen Text sind in den Fussnoten angegeben.

auf herkömmliche Weise organisieren, sind daher von der Richtlinie nicht erfasst. Zu den kumulativen Definitionsmerkmalen einer digitalen Plattform gehört auch der Einsatz von Instrumenten zur Überwachung oder automatisierten Entscheidungsfindung (Art. 2 Abs. 1 Nr. 1 Bst. d). Auch Instrumente zur Entscheidungsunterstützung gehören dazu (Art. 2 Abs. 1 Nr. 9[362]).

Kapitel III (Art. 7 ff.) befasst sich mit dem algorithmischen Management und sieht unter anderem die folgenden Rechte und Pflichten vor:

- Artikel 7 Absatz 1 verbietet die Erhebung oder Verarbeitung von Daten bezüglich des Privatlebens der Arbeitnehmerin oder des Arbeitnehmers mittels Systemen zur Überwachung oder automatisierten Entscheidungsfindung, wie z. B. private Gespräche, Vorhersagen über die Art und Weise der Ausübung ihrer oder seiner Grundrechte, oder Daten über sensible persönliche Merkmale (z. B. Rasse oder ethnische Herkunft, Gesundheitszustand) oder die Verarbeitung biometrischer Daten oder von Daten über den emotionalen oder psychischen Zustand.

- Artikel 8 hält fest, dass die Verarbeitung personenbezogener Daten durch ein automatisiertes Überwachungs- oder Entscheidungssystem voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen im Sinne von Art. 35 Abs. 1 DSGVO zur Folge hat, und sieht die Pflicht zur Durchführung einer Folgenabschätzung vor. Die Folgenabschätzung ist den Arbeitnehmervertretern auszuhändigen (Art. 8 Abs. 2).

- Artikel 9 sieht eine Informationspflicht über die Nutzung automatisierter Überwachungs- oder Entscheidungssysteme vor, auch wenn die Entscheidungen keine wesentlichen Auswirkungen auf Personen haben, die auf einer Plattform arbeiten.

- Artikel 10 sieht eine Pflicht zur Überwachung und Evaluierung von Systemen für automatisierte Entscheidungen vor. Jede Entscheidung in Bezug auf Aussetzung oder Beendigung des Vertragsverhältnisses ist von Menschen zu treffen (Art. 10 Abs. 5).

- Artikel 11 sieht die Pflicht zur Erläuterung einer von einem automatisierten System getroffenen Entscheidung sowie das Recht auf Überprüfung und gegebenenfalls Berichtigung der Entscheidung vor (Art. 11 Abs. 2 und 3).

- Artikel 12 enthält Verpflichtungen zur Bewertung und Vermeidung von Gesundheits- und Sicherheitsrisiken.

- Artikel 13 sieht eine Pflicht zur Unterrichtung und Anhörung der Arbeitnehmervertretung bei Entscheidungen über die Einführung oder wesentliche Änderung automatisierter Systeme vor. Die Arbeitnehmervertretung ist darüber hinaus Adressat der Informationspflicht nach Art. 9, wird in die Überwachung und Bewertung von KI-Systemen einbezogen (Art. 10) und kann Entscheidungen solcher Systeme überprüfen lassen (Art. 11).

[362] Art. 2 Abs. 1 Bst. j gemäss Berichtigung.

- Schliesslich sieht Artikel 21 die Pflicht zur Vorlage aller relevanten Beweismittel vor, und die Artikel 22 und 23 verlangen Schutz vor Vergeltungsmassnahmen oder Entlassung wegen Geltendmachung von Rechten aus der Richtlinie.

- Die Bestimmungen der Richtlinie sind zwingend, d. h. ungünstigere Regelungen sind weder in den nationalen Gesetzen der Mitgliedstaaten noch in Gesamtarbeitsverträgen zulässig (Art. 26 Abs. 2). Günstigere Regelungen sind hingegen möglich. Die Richtlinie weicht in einigen Punkten bewusst von der DSGVO ab. Insbesondere schliesst sie die Einwilligung als Rechtfertigungsgrund für eine Datenverarbeitung aus (vgl. E. 40).

6.5.2.2 EU-Verordnung zu KI und Arbeitsrecht

Die KI-Verordnung (vgl. Ziff. 5) enthält trotz ihres allgemeinen Charakters Regelungen, die speziell für die Arbeitswelt relevant sind. Sie stuft die Anwendungen für die Arbeitswelt nach vorgegebenen Risikostufen ein.

So werden Systeme zur Erkennung von Emotionen als verbotene Anwendungen eingestuft, wenn sie am Arbeitsplatz eingesetzt werden, es sei denn, die Anwendung erfolgt aus medizinischen Gründen oder zu Sicherheitszwecken (Art. 5 Abs. 1 Bst. f). Eine Reihe von Anwendungen wird als hochriskant eingestuft (Art. 6 Abs. 2 und Anhang III Nr. 4): Systeme zur Anwerbung oder Auswahl natürlicher Personen, einschliesslich der gezielten Veröffentlichung von Stellenangeboten, der Analyse und Filterung von Bewerbungen und der Bewertung von Bewerbern; Systeme zum Treffen von Entscheidungen, welche die Arbeitsbedingungen, die Beförderung oder die Beendigung des Arbeitsverhältnisses beeinflussen, oder die dazu bestimmt sind, Aufgaben auf der Grundlage des individuellen Verhaltens, der Persönlichkeitsmerkmale oder der persönlichen Eigenschaften zuzuweisen, sowie Überwachungs- und Bewertungssysteme. Die mit der Einstufung in die Hochrisikokategorie verbundenen Pflichten sowie die Ausnahmen von dieser Einstufung wurden bereits im Abschnitt über die KI-Verordnung erläutert (vgl. Ziff. 5.2.7).

Es stellt sich die Frage, ob der Arbeitgeber Adressat der Pflichten aus der KI-Verordnung sein kann, insbesondere der Pflichten gemäss Kapitel 3 der Verordnung über die Hochrisikokategorie. Der Arbeitgeber wird grundsätzlich die Person sein, die das KI-System im Sinne von Art. 3 Nr. 4 KI-Verordnung einsetzt (Betreiber). So sieht Art. 26 Abs. 7 der Verordnung u. a. vor, dass der Arbeitgeber die Arbeitnehmervertreter und die betroffenen Arbeitnehmenden darüber informiert, dass sie der Verwendung eines Hochrisiko-KI-Systems unterliegen werden (vgl. zu weiteren Pflichten zu Lasten des Betreibers Ziff. 5.2.7.3.3). Gestützt auf Art. 25 Abs. 1 KI-Verordnung kann jedoch auch der Arbeitgeber als Anbieter betrachtet werden, da jede Person, die wesentliche Änderungen an einem Hochrisiko-KI-System vornimmt oder dessen Zweckbestimmung ändert, als Anbieter mit den entsprechenden Pflichten gilt.

Auch diskutiert wurde die Wechselwirkung zwischen diesen allgemeinen Regelungen und den auf EU-Ebene vorgesehenen arbeitsrechtlichen Bestimmungen.[363] Zunächst ist klar, dass

[363] Vgl. insbesondere AUDE CEFALIELLO/MIRIAM KULLMANN, Offering false security: How the draft artificial intelligence act undermines fundamental workers rights, European Labour Law Journal 2022, 542 ff.

die Verordnung einen horizontalen und damit subsidiären Charakter hat und alle arbeitsrechtlichen Verpflichtungen weiterhin gelten. Darüber hinaus kann es sein, dass bei der Konzeption des Systems, in das KI integriert wird, Massnahmen erforderlich sind, um es mit den bestehenden arbeitsrechtlichen Vorschriften in Einklang zu bringen. Diese Massnahmen obliegen jedoch dem Arbeitgeber, der nicht notwendigerweise Anbieter und Entwickler des Systems ist. Der Arbeitgeber, der sich für die Einführung eines KI-Systems entscheidet, muss sicherstellen, dass er seinen arbeitsrechtlichen Verpflichtungen nachkommt, was bedeutet, dass er über die Informationen und Mittel verfügen muss, um die Funktionsweise des Systems gegebenenfalls anzupassen.

6.5.3 Schweizerisches Arbeitsrecht: Aktueller Stand und Diskussion

6.5.3.1 Motion 23.4492 Gysi «Künstliche Intelligenz am Arbeitsplatz. Mitwirkungsrechte der Arbeitnehmenden stärken»

Die am 22. Dezember 2023 eingereichte Motion 23.4492 verlangt, «die Mitwirkungsrechte der Arbeitnehmenden beim Einsatz von Künstlicher Intelligenz (KI) und algorithmischen Systemen am Arbeitsplatz auf gesetzlicher Ebene zu stärken, wenn diese Systeme für Empfehlungen, Prognosen, Entscheidungen usw. benutzt werden, die Arbeitnehmende betreffen oder Arbeitnehmendendaten verwenden». Hierfür «sollen das Mitspracherecht ausgeweitet, Informationsrechte gestärkt, kollektive Klagerechte geschaffen sowie Sanktionsmöglichkeiten geprüft werden». Die Motion basiert auf einer Stellungnahme der Gewerkschaft syndicom und der NGO Algorithmwatch, die wiederum auf einer von diesen Organisationen in Auftrag gegebenen Studie beruht.[364]

6.5.3.2 Einsatz von KI in der schweizerischen Arbeitswelt

Eine 2018 durchgeführte und 2020 wiederholte Umfrage bei Schweizer Grossunternehmen mit fünf Fallstudien bietet einen Überblick über die Verbreitung des Einsatzes von KI in Schweizer Unternehmen.[365] Rund zwei Drittel der Befragten gaben an, solche Instrumente zu nutzen, wobei zwischen 2018 und 2020 ein Anstieg zu verzeichnen ist.[366]

Die am häufigsten genannten Einsatzbereiche waren: Rekrutierung, Mitarbeiterbindung und -entwicklung, Leistungsmanagement, Arbeitsplatzgestaltung, Compliance sowie Auswahl- und Entlassungsentscheidungen. Die Studie zeigt auch die am häufigsten genannten Verwendungszwecke wie Mitarbeiterbindung und -entwicklung (63 %), Leistungsmanagement (47 %)

---

[364] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden beim Einsatz von ADM-Systemen am Arbeitsplatz, November 2023, abrufbar unter: https://syndicom.ch/fileadmin/user_upload/Web/Website/Dossiers/KI/2023_Rechtsgutachten_final.pdf (abgerufen am 28. August 2024).

[365] Diese Studie wurde durchgeführt im Rahmen des Nationalen Forschungsprogramms «Big Data» (PNR 75), vgl. https://www.nfp75.ch/de (abgerufen am 28. August 2024).

[366] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 5.

und Rekrutierung (39 %). Weit verbreitet sind beispielsweise die Vorabprüfung von Bewerbungen bei der Einstellung oder Instrumente zur Überwachung von Computereingaben oder Internetnutzung. In diesen Fällen stellt sich die Frage, welche Rolle KI spielt oder ob diese Werkzeuge überhaupt unter die Definition von KI fallen.[367] Die vorliegende Studie konzentriert sich auf Systeme zur Unterstützung der Entscheidungsfindung (Teilautomatisierung) oder auf vollautomatisierte Systeme.

6.5.3.3 Herausforderungen durch KI im schweizerischen Arbeitsrecht

Die folgenden Herausforderungen wurden im Zusammenhang mit KI am Arbeitsplatz in der oben genannten Studie oder in anderen Publikationen identifiziert:

- Datenschutz[368]: Diese Systeme können potenziell in grossem Umfang und kontinuierlich Personendaten von Arbeitnehmenden sammeln. Die Sammlung kann undifferenziert erfolgen und die Zwecke können vielfältig oder unklar sein.

- Gesundheitsschutz: Instrumente zur Leistungsmessung oder -überwachung können die psychische Gesundheit und das Wohlbefinden von Beschäftigten beeinträchtigen, wenn diese sich ständig bewertet fühlen.[369] Andererseits können Überlastungssituationen entstehen, wenn ein KI-Tool die Aufgabenerfüllung anhand von realitätsfernen Ausführungsparametern überwacht.

- Verzerrungen bei der Bewertung, starre Entscheidungsfindung: Die gemessenen oder in das System eingegebenen Parameter berücksichtigen nicht unbedingt alle Besonderheiten einer Situation (Hindernisse auf dem üblichen Weg, die es unmöglich machen, Lasten in der vorgesehenen Zeit zu transportieren; ein Stau, der es nicht erlaubt, eine Strecke in einer Standardzeit zurückzulegen). Der Einsatz von KI führt auch eine zahlenbasierte Leistungslogik ein, die wichtige, aber nicht quantifizierbare Elemente wie persönliche Qualitäten und soziale Kompetenzen oder nicht messbare Schwierigkeiten ausser Acht lässt.[370] Eine

---

[367] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 10.

[368] ISABELLE WILDHABER, Die Roboter kommen - Konsequenzen für Arbeit und Arbeitsrecht, RDS 2016, 315 ff., 346 f.; ISABELLE WILDHABER, Répercussions de la robotique et de l'intelligence artificielle sur le lieu de travail, in: Jean-Philippe Dunand/Pascal Mahon/Aurélien Witzig (Hrsg.), La révolution 4.0 au travail - Une approche multidisciplinaire, Genf/Zürich/Basel 2019, 201 ff., 228 f.; WOLFGANG DÄUBLER, Digitalisierung und Arbeitsrecht, Frankfurt am Main 2022, §9, N 5 ff.

[369] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 18: 22 % der Studienteilnehmer gaben an, dass das Verhalten der Arbeitnehmenden und wie sie zusammenarbeiten überwacht werde. Siehe auch AUDE CEFALIELLO/MIRIAM KULLMANN, Offering false security (Fn. 363), 559.

[370] Vgl. dazu JEAN-CHRISTOPHE SCHWAAB, Les nouvelles tendances en matière d'évaluation du personnel et le droit du travail, Revue de droit du travail et d'assurance-chômage, 2019, 103 ff., 114 f.

Software, die die Tastatureingaben registriert, könnte zu unbegründeten negativen Bewertungen führen, wenn die Person an einem schwierigen Thema arbeitet und häufig innehält, um nachzudenken. Auch Diskussionen unter Kollegen, ihre Qualität und ihr positiver Beitrag lassen sich mit KI-gestützten Instrumenten nur schwer erfassen. Diese Grenzen der automatisierten Entscheidungsfindung am Arbeitsplatz haben dazu geführt, dass über verschiedene Massnahmen nachgedacht wurde, wie z. B. die Notwendigkeit der Entscheidungsfindung oder -kontrolle durch einen Menschen, die Entscheidungsautonomie des Menschen gegenüber den Ergebnissen der Maschine oder das Recht, die von einem KI-System getroffene oder auf dessen Bewertungen beruhende Entscheidung erläutern oder von einem Menschen überprüfen zu lassen.[371] Wie in Ziff. 6.5.2 dargelegt, wurden diese Rechtsbehelfe bereits in die von der EU erlassenen Rechtsvorschriften aufgenommen. Dies ist teilweise auch in der Schweiz mit dem neuen DSG der Fall (vgl. Ziff. 6.5.3.4).

- Diskriminierende Verzerrungen: Die Frage des Diskriminierungspotenzials ist im Zusammenhang mit der Integration von KI in Entscheidungsprozesse am Arbeitsplatz, insbesondere bei der Personalrekrutierung, viel diskutiert worden.[372] Dieses Thema wird in der vorliegenden Analyse gesondert behandelt (vgl. Ziff. 4.3.2.5), aber trotzdem auch im vorliegenden Zusammenhang erwähnt. Verzerrungen ergeben sich insbesondere daraus, dass die vorhandenen Datensätze und Entscheidungsmengen, die von Instrumenten mit integrierter KI zum «Lernen» genutzt werden, die in der Gesellschaft bestehenden diskriminierenden Verzerrungen enthalten können.

- Transparenz und Begründung von Entscheidungen: In der Lehre wird darauf hingewiesen, dass das Phänomen der Erteilung von Anweisungen durch automatisierte Systeme in der Arbeitswelt Realität geworden ist.[373] Diese Anweisungen müssen jedoch den rechtlichen Rahmen des Arbeitsrechts einhalten, was bedeutet, dass man wissen muss, auf welcher Grundlage die Entscheidung getroffen wurde, und daher auch, wie die KI-Systeme programmiert sind. Arbeitnehmende, die eine Anweisung anfechten wollen, wissen dies jedoch nicht. Dies gilt auch für jede andere Entscheidung, die von einem KI-System getroffen wird oder auf einem KI-System beruht. Auch dieses Thema wurde weiter oben allgemein behandelt (vgl. Ziff. 4.3.2.3).

---

[371] AIDA PONCE DEL CASTILLO, Le travail à l'ère de l'IA: pourquoi la réglementation est nécessaire pour protéger les travailleurs, etui. 2020, 13; JEREMIAS ADAMS-PRASSL/ HALEFOM ABRAHA/AISLINN KELLY-LYTH et al., Regulating algorithmic management: A blueprint, European Labour Law Journal 2023, 124 ff., 139–140, 143.

[372] Unter anderen ISABELLE WILDHABER, Répercussions de la robotique (Fn. 368), 213 ff. Vgl. auch https://www.rts.ch/info/sciences-tech/13464935-la-place-de-lintelligence-artificielle-dans-le-recrutement.html , https://www.humanrights.ch/fr/nouvelles/discrimination-algorithmique-protection , https://algorithmwatch.ch/de/findhr/ (abgerufen am 28. August 2024).

[373] ISABELLE WILDHABER, Répercussions de la robotique (Fn. 368), 210 f.

6.5.3.4 Anwendbare Bestimmungen im Schweizer Recht

Die Regeln, die zur Lösung dieser Probleme herangezogen werden können, sind unterschiedlicher Natur:

- Zunächst sind die Datenschutzbestimmungen zu nennen. Das DSG gilt für Arbeitsverhältnisse in der Privatwirtschaft. Artikel 328*b* OR regelt diese Frage zusätzlich zu den allgemeinen Bestimmungen speziell für den Arbeitsvertrag. Artikel 328*b* erster Satz OR nennt zwei mögliche Zwecke für die Datenerhebung durch den Arbeitgeber: die Abklärung der Eignung des Arbeitnehmers für die Arbeitsleistung und die Durchführung des Arbeitsvertrages. Diese Bestimmung gehört zum relativ zwingenden Recht. Der zweite Satz verweist auf die Anwendbarkeit des DSG.

  Die Frage, ob der Arbeitgeber Daten zu anderen als den in Artikel 328*b* OR vorgesehenen Zwecken beschaffen darf, war lange Zeit umstritten, insbesondere im Zusammenhang mit der Einwilligung des Arbeitnehmenden, die im DSG einen Rechtfertigungsgrund darstellt (Art. 31 Abs. 1). Während sich ein Grossteil der Lehre einig war, dass die Einwilligung eine Datenbearbeitung zu anderen Zwecken nicht rechtfertigen kann,[374] hat das Bundesgericht diese Möglichkeit kürzlich bejaht:[375] Die beiden in Artikel 328*b* OR genannten Fälle begründen eine Vermutung der Rechtmässigkeit der Datenbearbeitung, wobei sich die Bearbeitung zu anderen Zwecken auf die allgemeinen Rechtfertigungsgründe des DSG stützen muss. Trotz dieser offenen Haltung des Bundesgerichts bestehen in der Lehre starke Vorbehalte gegen die Möglichkeit einer freiwilligen Einwilligung im Sinne von Artikel 6 Absatz 6 DSG im Arbeitsrecht oder in jeder Situation eines strukturellen Ungleichgewichts zwischen den Parteien.[376]

  Die durch KI aufgeworfenen Probleme lassen sich mit den allgemeinen Regeln (Art. 328*b* OR und das DSG) abdecken. Zunächst setzen die Vorschriften über den Zweck der Datenbearbeitung gewisse Grenzen. So ist der Arbeitgeber, der ein KI-System einsetzt, an die in Artikel 328*b* OR festgelegten Zwecke gebunden. Er muss die Beschaffung von Daten zu anderen Zwecken rechtfertigen. Der Zweck muss zudem gemäss Artikel 6 Absatz 3 DSG bestimmt und für die betroffene Person erkennbar sein. Auch die übrigen Grundsätze des DSG müssen eingehalten werden, insbesondere die Verhältnismässigkeit (Art. 6 Abs. 2 DSG). Die Einhaltung dieser Grundsätze führt bei

---

[374] Siehe insbesondere den Bericht des Bundesrates vom 16. November 2016 in Erfüllung des Postulats 12.3166 Meier-Schatz, Rechtliche Folgen der zunehmenden Flexibilisierung des Arbeitsplatzes, Ziff. 7.7.1 und Referenzen, abrufbar unter: https://www.newsd.admin.ch/newsd/message/attachments/80239.pdf (abgerufen am 28. August 2024).

[375] Urteil des BGer, 4A_518/2020, E. 4.2.4.

[376] CÉLIAN HIRSCH, Droit du travail et intelligence artificielle: défis des décisions automatisées pour les employeurs, in: Valérie Défago/Jean-Philippe Dunand/Pascal Mahon/Samantha Posse/David Raedler (Hrsg.), La protection des données dans les relations de travail à la lumière de la nouvelle loi fédérale sur la protection des données, Genf/Zürich 2024, 95 ff., 112 und die dortigen Referenzen; ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 16.

KI-Systemen zu einem Spannungsfeld, namentlich aufgrund der mit der Entwicklung der Systeme einhergehenden Beschaffung von grossen Datenmengen, der Vielfalt der verfolgten Zwecke oder der Verwendung zu anderen als den ursprünglich vorgesehenen Zwecken. So zeigt die oben genannte Studie, dass die Einführung von KI-Systemen von einigen Arbeitgebern gestoppt wurde, weil diese Systeme die oben genannten Grundsätze nicht einhielten.[377] Darüber hinaus sind verschiedene Bestimmungen des DSG zu erwähnen, die für die Arbeitswelt besonders relevant sind. Erstens kann es sich bei den gesammelten Daten um besonders schützenswerte Personendaten im Sinne von Artikel 5 Buchstabe c DSG handeln, wie z. B. die Angabe einer Gewerkschaftsmitgliedschaft oder Gesundheitsdaten, und/oder sie können als Grundlage für ein Profiling mit hohem Risiko dienen. Für die Beschaffung solcher Daten sowie für das Profiling mit hohem Risiko sind besondere Voraussetzungen zu erfüllen, unter anderem gegebenenfalls die ausdrückliche Einwilligung der betroffenen Person (Art. 6 Abs. 7 Bst. a DSG) und die Pflicht zur Durchführung einer Datenschutz-Folgenabschätzung, wenn besonders schützenswerte Personendaten in grossem Umfang bearbeitet werden (Art. 22 Abs. 1 und 2 Bst. a DSG).

Zu erwähnen ist auch Artikel 21 Absätze 1 und 2 DSG, der sich speziell auf KI-Systeme bezieht und spezifische Informationsrechte und -pflichten im Falle einer automatisierten Einzelentscheidung vorsieht (vgl. Ziff. 4.3.2.3). Es sei daran erinnert, dass der Arbeitgeber grundsätzlich von seinen Pflichten befreit ist, wenn er die ausdrückliche Einwilligung des Arbeitnehmenden einholt (Art. 21 Abs. 3 Bst. b DSG), unter Vorbehalt der oben erwähnten Problematik der Einwilligung im Arbeitsrecht. Der Begriff der automatisierten Einzelentscheidung ist relativ eng gefasst, da er keine Systeme umfasst, die eine Bewertung liefern, die als Grundlage für eine von einem Menschen getroffene Entscheidung dient.[378]

- Die Vorschriften über den Gesundheitsschutz können in mehrfacher Hinsicht zum Tragen kommen. Der Gesundheitsschutz ist vor allem in Artikel 328 OR und in Artikel 6 ArG geregelt. Konkretisiert wird die Pflicht zum Gesundheitsschutz in den Verordnungen zum ArG.

  Von besonderer Bedeutung ist in diesem Zusammenhang Artikel 26 ArGV 3, der in Absatz 1 Überwachungs- oder Kontrollsysteme, die dazu bestimmt sind, das Verhalten der Arbeitnehmenden am Arbeitsplatz zu überwachen, verbietet. Darunter sind alle technischen Systeme zu verstehen, mit denen Daten über das Verhalten der Arbeitnehmenden aufgezeichnet werden können, wie Kameras, Geräte zum Abhören von Telefongesprächen oder zur Überwachung der Tätigkeit am Computer, Ortungssysteme oder

---

[377] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 8.

[378] CÉLIAN HIRSCH, Droit du travail et intelligence artificielle (Fn. 376), 104 f., obwohl die Rechtsprechung der EU zu einer breiten Auslegung zu tendieren scheint; ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 15.

Informatikwerkzeuge mit künstlicher Intelligenz.[379] Solche Überwachungssysteme sind jedoch zulässig, wenn sie zu rechtmässigen Zwecken eingesetzt werden, z. B. zur Gewährleistung der Sicherheit des Betriebsgeländes oder zur Qualitäts- oder Leistungskontrolle, sofern sie dem Grundsatz der Verhältnismässigkeit entsprechen und die betroffenen Arbeitnehmenden darüber informiert wurden.[380]

In diesem Zusammenhang sind noch weitere Pflichten relevant: Artikel 6 Absatz 2 ArG bestimmt, dass der Arbeitgeber die Arbeitnehmenden vor Überbeanspruchung zu schützen hat, und Artikel 2 Absatz 1 ArGV 3 legt fest, dass sich der Schutz auf die physische und psychische Gesundheit bezieht und insbesondere die Vermeidung von übermässig starker Beanspruchung umfasst. Dem Arbeitgeber obliegt eine Informations- und Anleitungspflicht (Art. 5 Abs. 1 ArGV 3) sowie die Pflicht zur Information und Beratung der Arbeitnehmenden oder ihrer Vertreter in Fragen des Gesundheitsschutzes (Art. 48 Abs. 1 Bst. a ArG).

- Das Mitwirkungsgesetz[381] regelt die Information und die Mitsprache der Arbeitnehmenden im Betrieb. Es handelt sich um ein Rahmengesetz, das durch eine Reihe von Einzelgesetzen konkretisiert wird, von denen einige in Artikel 10 aufgeführt sind. Das Gesetz sieht in Betrieben mit mindestens 50 Arbeitnehmenden einen Anspruch auf Vertretung vor (Art. 3). Die Arbeitnehmervertretung wiederum hat Anspruch auf Information über alle Angelegenheiten, deren Kenntnis Voraussetzung für eine ordnungsgemässe Erfüllung ihrer Aufgaben ist (Art. 9 Abs. 1). Über dieses allgemeine Informationsrecht hinaus sind die Mitwirkungsrechte in Spezialgesetzen geregelt. Es ist jedoch kein allgemeines Konsultationsrecht vorgesehen. Gemäss Artikel 10 sind in vier Bereichen besondere Mitwirkungsrechte vorgesehen: Arbeitssicherheit und Arbeitnehmerschutz, Übergang von Betrieben, Massenentlassungen, Anschluss an eine Einrichtung der beruflichen Vorsorge.

- Relevant sind auch die Grenzen des Weisungsrechts: Das Weisungsrecht nach Artikel 321*d* OR muss mit dem zwingenden Recht, insbesondere mit Artikel 328 OR (Persönlichkeitsschutz), dem Vertrag und dem Grundsatz von Treu und Glauben vereinbar sein.[382] Jegliche von einer KI erteilte Weisung muss daher in diesem Rahmen liegen, andernfalls die Arbeitnehmerin oder der Arbeitnehmer deren Befolgung verweigern darf. Insbesondere kann sich der Arbeitgeber nicht auf einen automatisierten Entscheidungsmechanismus berufen, um sich zu rechtfertigen, oder die Verantwortung auf den Systementwickler abwälzen.[383] Für den Arbeitgeber bedeutet dies, dass er in der Lage sein

---

[379] SECO, Wegleitung zur ArGV 3 (Fn. 126), Art. 26, 326-2.

[380] BGE **130** II 425, E. 4.4; SECO, Wegleitung zu ArgGV 3 (Fn. 126), Art. 26, 326-1.

[381] SR **822.14**

[382] BGE **132** III 115, E. 5.2.

[383] ISABELLE WILDHABER, Die Roboter kommen (Fn. 368), 330 f.

muss, die Rechtmässigkeit der getroffenen Entscheidung nachzuweisen, was voraussetzt, dass er weiss, wie das KI-System funktioniert und programmiert ist, und dass er in der Lage ist, dies dem Arbeitnehmenden zu erläutern.

- Schliesslich ist auch der allgemeine Persönlichkeitsschutz relevant (Art. 328 OR): Jede Entscheidung über die zu verrichtende Arbeit, die auf einer von der KI generierten Bewertung beruht, kann die Persönlichkeit des Arbeitnehmers verletzen, wenn sie nicht auf der objektiven Realität der zu verrichtenden Arbeit beruht. Dies kann auch bei einer Zuweisung, die die Situation des Arbeitnehmers verschlechtert, oder bei einer Sanktion der Fall sein.

6.5.3.5 Würdigung

Die bevorstehende Verabschiedung der Richtlinie über Plattformarbeit wird zu einer unterschiedlichen Regelung im Arbeitsrecht der EU und im schweizerischen Recht führen. Das Schweizer Recht siehe keine Spezialregelungen für das algorithmische Management von Arbeit vor, weder speziell für digitale Plattformen, die Arbeitsleistungen anbieten, noch allgemein für alle Arbeitsverhältnisse.

Die Stellungnahme der Gewerkschaft syndicom und von Algorithmwatch,[384] auf die sich die oben zitierte Studie von WILDHABER/EBERT stützt, zeigt Lücken auf. Diese Positionen werden in der Motion 23.4492 Gysi aufgegriffen. Ihre Forderung zielt hauptsächlich auf eine Stärkung der Mitbestimmung der Arbeitnehmenden.

Einige Forderungen zielen auf eine Stärkung der Mitwirkungsrechte im Allgemeinen ab, ohne speziell auf die aufkommende KI einzugehen. Andere Forderungen beziehen sich speziell auf den Einsatz von KI. Dies ist der Fall bei möglichen Regelungen zur Information und Mitsprache der Arbeitnehmenden bei der Einführung und Nutzung von KI-Systemen durch den Arbeitgeber. Derzeit sind solche allgemeinen Informationsrechte in Artikel 9 Absatz 1 des Mitwirkungsgesetzes an das Vorliegen von Spezialgesetzen geknüpft. Bei der Einführung eines KI-Systems sind nur die Vorschriften über den Gesundheitsschutz relevant (Art. 48 Abs. 1 Bst. a ArbG; vgl. Ziff. 6.5.3.4). Mitwirkungsrechte ergeben sich aus dem geltenden Recht somit allenfalls indirekt bei einer möglichen Gesundheitsbeeinträchtigung. Hinzu kommt, dass Artikel 9 Absatz 1 des Mitwirkungsgesetzes zwar ein Informationsrecht vorsieht, dieses aber sehr allgemein formuliert ist. Die Datenschutzbestimmungen begründen zwar ein Informationsrecht der Arbeitnehmenden, aber nur ein individuelles, und kein kollektives, während KI-Systeme aber Fragen aufwerfen, die alle Arbeitnehmenden betreffen.[385] Diese kollektive Dimension betrifft zwar nicht nur Arbeitsverhältnisse, das Arbeitsrecht enthält aber Bestimmungen zur kollektiven Interessenvertretung der Arbeitnehmenden.

Schliesslich zeigt die Analyse, dass die allgemeinen arbeitsrechtlichen Bestimmungen die Einrichtung und den Einsatz von KI-Systemen durch den Arbeitgeber auf verschiedenen Ebe-

---

[384] Download verfügbar unter https://syndicom.ch/fr/themes/dossier/intelligenceartificielleia/algorithmes-au-travail (abgerufen am 28. August 2024).

[385] In diesem Sinn: ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (Fn. 364), 16 f.

nen regeln. So muss sich der Arbeitgeber nacheinander fragen, ob sein System allenfalls aufgrund von Artikel 26 ArGV 3 verboten ist, ob es den Anforderungen des Datenschutzes, insbesondere in Bezug auf die Zweckbindung oder die Verhältnismässigkeit, genügt, oder ob es spezifischeren Verpflichtungen im Zusammenhang mit der Bearbeitung von besonders schützenswerten Daten oder der Erstellung von Profilen mit hohem Risiko untersteht. Darüber hinaus muss er die in Artikel 21 DSG festgelegten Verpflichtungen in Bezug auf automatisierte Einzelentscheidungen einhalten. Schliesslich muss er sicherstellen, dass er die von oder mit Hilfe von KI-Systemen getroffenen Entscheide begründen kann, z. B. um die Rechtmässigkeit von Weisungen an Arbeitnehmende zu überprüfen oder um eine Kündigung zu begründen, wenn ein Arbeitnehmender dies verlangt (Art. 335 Abs. 2 OR). Die derzeitige Rechtslage hat den Vorteil, dass sie relativ viele Grenzen setzt, aber den Nachteil, dass jede einschlägige Norm nur einen Teilaspekt des Problems abdeckt und auf einem anderen Begriff beruht, wie z. B. «Überwachungs- oder Kontrollsystem zur Verhaltensüberwachung», «automatisierte Einzelentscheidung» oder «Profiling mit hohem Risiko». Diese Vielfalt macht die Rechtslage kompliziert und wird dazu führen, dass KI-Systeme unterschiedlichen Regeln unterliegen, je nachdem, ob sie unter eine, mehrere oder keine der in den verschiedenen Vorschriften definierten Kategorien fallen. Darüber hinaus besteht Unsicherheit in Bezug auf die Rechtfertigung der Datenbearbeitung durch Einwilligung gemäss DSG in arbeitsvertraglichen Beziehungen.

EU-Recht: Im Vergleich zur neuen EU-Richtlinie über Plattformarbeit besteht in der Schweiz eine Gesetzeslücke. Ebenso fehlt im Schweizer Recht ein Äquivalent zu den spezifischen Regeln der KI-Verordnung in Bezug auf das Arbeitsverhältnis.

Zwar gewährleisten die allgemeinen arbeits- und datenschutzrechtlichen Bestimmungen hier einen gewissen Arbeitnehmerschutz, jedoch findet das in der EU eingeführte Normensystem, das KI-Systeme von der Konzeption bis zum Einsatz über die gesamte Lebensdauer begleitend regelt, im geltenden Schweizer Recht keine Entsprechung. Auch die spezifischen Bestimmungen über die Mitwirkung der Arbeitnehmenden haben im schweizerischen Recht keine Entsprechung, obwohl die bestehenden Vorschriften über den Gesundheitsschutz und die Sicherheit am Arbeitsplatz einen Teil der Situationen abdecken.

Parlamentarische Vorstösse: Die Motion 23.4492 Gysi und die ihr zugrunde liegenden Dokumente weisen ebenfalls auf Lücken hin. Ein Teil dieser Lücken ist nicht KI-spezifisch. Ein gesondertes Recht auf Information und Mitwirkung bei der Einführung und Anwendung von KI-Systemen könnte aber angesichts des sehr allgemein gehaltenen Mitwirkungsgesetzes sinnvoll sein.

Die Analyse des regulatorischen Bedarfs im Arbeitsrecht muss mit der allgemeinen Analyse des Regelungsbedarfs im Bereich der KI koordiniert werden. Mit der Umsetzung des oben identifizierten übergreifenden Regelungsbedarfs, insbesondere im Fall einer möglichen Ratifizierung der KI-Konvention (vgl. Ziff. 4.6), könnte bereits eine Reihe von Herausforderungen in der Arbeitswelt angegangen werden. Zu denken ist insbesondere an die Grundsätze der Transparenz und Aufsicht (Art. 8), der Gleichstellung und Nichtdiskriminierung (Art. 10) und des Datenschutzes (Art. 11). Der Anwendungsbereich der Konvention auf privatrechtliche Beziehungen stellt jedoch eine Einschränkung dar. Die Konvention ist

nämlich auf Situationen beschränkt, in denen Grundrechte, Demokratie oder Rechtsstaatlichkeit betroffen sind (vgl. Ziff. 4.2.3.1). Eine zweite Einschränkung betrifft die Mitwirkung der Arbeitnehmenden und die damit zusammenhängenden arbeitsrechtlichen Regelungen.

Sollte sich die Schweiz für eine Annäherung an die KI-Verordnung entscheiden und beispielsweise bestimmte problematische Anwendungen im privaten Sektor verbieten, hätte dies auch Auswirkungen auf das Arbeitsrecht. Sollte die Schweiz diesen Weg hingegen nicht gehen, stellt sich die Frage, ob eine Übernahme der spezifischen arbeitsrechtlichen Regelungen sinnvoll wäre.

Ein spezifischer arbeitsrechtlicher Regelungsbedarf scheint daher derzeit nur punktuell zu bestehen. Die Entwicklung des allgemeinen rechtlichen Rahmens sollte jedoch beobachtet werden, um gegebenenfalls mit spezifischen Massnahmen eingreifen zu können.

6.6 Strafrecht

6.6.1 Grundsätzliche Anwendbarkeit

Grundsätzlich ist das schweizerische Strafgesetzbuch (StGB) technologieneutral ausgestaltet und bleibt unabhängig von der technologischen Vorgehensweise oder des spezifischen Instrumentariums der Täterschaft anwendbar.[386] Die materielle Reichweite des Schutzes ist daher grundsätzlich geeignet und darauf ausgerichtet, um den Einsatz unterschiedlicher Technologien durch einen Täter oder einen Täterkreis zu erfassen. Bedient sich die Täterschaft KI-Systemen wie zum Beispiel der Deepfake-Technologie, um ein Delikt gegen die Ehre oder den Privatbereich zu begehen, sind die entsprechenden Straftatbestände (Art. 173 ff. StGB, insbesondere auch der Identitätsmissbrauch gem. Art. 179^{decies} StGB) anwendbar. Verfolgt die Täterschaft durch den Einsatz von KI-Systemen weitergehende Absichten, zum Beispiel das Erlangen eines Vermögensvorteils mittels Täuschung, sind die entsprechenden Straftatbestände aus dem Bereich des Vermögensstrafrechts auch beim Einsatz einer solchen neuen, sich in Entwicklung befindlichen Technologie anwendbar (z. B. Betrug gem. Art. 146 StGB oder die Warenfälschung gem. Art. 155 StGB). Solange die Täterschaft KI-Systeme einsetzt und dabei tatbestandsmässig handelt, bleibt die Verantwortung grundsätzlich beim Menschen oder Täterkreis, der diese Technologie nutzt. Sind die Technologien in bestimmten Konstellationen besonders raffiniert oder nimmt die Qualität der Täuschung zu, wie im Beispiel der erwähnten Deepfakes, kann der Täterschaft im Rahmen der Strafzumessung im Einzelfall durchaus eine besonders arglistige Vorgehensweise oder Täuschungsabsicht vorgeworfen und eine Erhöhung der auszufällenden Strafe in Betracht gezogen werden.

Auch die strafbaren Handlungen gegen die sexuelle Integrität sind grundsätzlich technologieneutral ausgestaltet. Für Deepfakes ist namentlich Artikel 197 StGB (Pornografie) relevant.

[386] Siehe dazu die Stellungnahme des Bundesrats vom 16. August 2023 zur Mo. 23.3563, Deepfakes regulieren, abrufbar unter: https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20233563 (abgerufen am 29. August 2024).

Grundsätzlich sind im Kontext von Artikel 197 Absätze 4 und 5 StGB unter anderem die Herstellung, der Besitz und die Verbreitung von pornographischen Deepfakes, die als «harte Pornografie» gelten, strafbar.[387] Unabhängig von der angewandten Technik gelten sämtliche Handlungen als tatbestandsmässig, die Pornografie hervorbringen, d. h. nicht nur Neuanfertigungen, sondern auch Vervielfältigungen, Reproduktionen oder Bearbeitungen von anderweitig erstellten pornografischen Darstellungen.[388] Das Strafgesetzbuch misst der sexuellen Integrität von Kindern und Jugendlichen in diesem Kontext eine besondere Bedeutung zu und pönalisiert auch Gegenstände und Vorführungen, die *nicht tatsächliche* sexuelle Handlungen mit Minderjährigen zum Inhalt haben (also zum Beispiel mittels Datenverarbeitung generierte Bilder oder Darstellungen; Art. 197 Abs. 4 und 5 StGB). Das heisst, dass auch mittels Herstellung eines KI-Systems frei erfundene pornografische Darstellungen mit Minderjährigen i. d. R. als tatbestandsmässig zu qualifizieren sind. Zudem bilden die Artikel 187 und/oder 197 StGB einen Anknüpfungspunkt für die Strafbarkeit von Straftaten gegen Minderjährige im Ausland (Art. 5 StGB).

Des Weiteren sind bei tatbestandsmässigem Verhalten beispielsweise auch die öffentliche Aufforderung zu Verbrechen oder zur Gewalttätigkeit (Art. 259 StGB) oder eine Diskriminierung und Aufruf zu Hass (Art. 261bis StGB) anwendbar, unabhängig davon, unter Zuhilfenahme welcher Technologie eine Person oder eine Gruppe von Personen herabgesetzt oder diskriminiert wird oder mittels welcher spezifischen Vorgehensweise zu einem Vergehen mit Gewalttätigkeiten aufgefordert wird.

6.6.2 Strafrechtliche Verantwortlichkeit

6.6.2.1 Allgemeine Bemerkungen

Das Schweizer Strafrecht und damit auch die strafrechtliche Verantwortlichkeit fussen grundsätzlich auf dem sogenannten Schuldprinzip.[389] Die Schuld ist als eines der zentralsten Elemente des Schweizer Strafrechts Voraussetzung für eine Bestrafung (Art. 19 Abs. 1 StGB) sowie Grundlage für die Strafzumessung (Art. 47 Abs. 1 StGB). Demzufolge gibt es grundsätzlich keine Strafe ohne Schuld («*nulla poena sine culpa*»). Entsprechend knüpft die strafrechtliche Verantwortlichkeit grundsätzlich an der Schuldfähigkeit des Menschen an und erfasst in diesem Sinne nicht eine Technologie, eine Maschine oder ein System an sich (wie z. B. einen Roboter) als Rechtssubjekt. Die Frage, ob eine autonome Technologie als solche ein schuldfähiges und damit strafrechtliches Rechtssubjekt darstellen kann oder soll und ob

---

[387] Ausführlicher dazu siehe BRIGITTE TAG/MARTIN WYSS, Die strafrechtliche Einordnung von pornografischen Deepfakes, Jusletter 29. April 2024, 1 ff.

[388] ULRICH WEDER, in: Andreas Donatsch (Hrsg.), StGB/JStG, Mit weiteren Erlassen und Kommentar zu den Strafbestimmungen des SVG, BetmG, AIG und OBG, Orell Füssli Kommentar, Zürich 2022, Art. 197 N 18, m.H.; PK StGB-TRECHSEL/BERTOSSA, Art. 197 N 15; BRIGITTE TAG/MARTIN WYSS, Die strafrechtliche Einordnung (Fn. 387), 11.

[389] BGE **123** IV 1, E. 2.

die grundsätzlichen Prinzipien des Strafrechts in diese Richtung verändert und angepasst werden sollen, wurde zwar in der Lehre bereits aufgeworfen, wird jedoch überwiegend abgewiesen.[390] Ob ein solcher Regulierungsansatz im Strafrecht mit einem Mehrwert verbunden wäre, ist durchaus kritisch zu betrachten und muss grundsätzlich in Frage gestellt werden. KI-Systeme sind grundsätzlich nämlich nicht in der Lage, schuldfähig zu handeln. Während es einem menschlichen Täter zum Tatzeitpunkt möglich ist, sich aufgrund seines Willens für oder gegen ein Unrecht zu entscheiden oder ein solches in Kauf zu nehmen, «entscheidet» ein KI-System vielmehr aufgrund von Algorithmen und Datensätzen.[391]

Da ein KI-System folglich nicht selbst strafrechtlich belangt werden kann, stellt sich die Frage nach der Zurechnung und den Übergängen der Verantwortlichkeit im Zusammenhang mit durch KI-Systemen ausgelösten Schädigungen vielmehr zwischen der Ebene des Nutzers und des Herstellers (und ggf. des Halters) von automatisierten (bis hin zu autonomen) KI-Systemen. Dabei würde das Herstellerunternehmen als solches im Kontext des Unternehmensstrafrecht dann relevant, wenn die Schädigung in geschäftlicher Verrichtung im Rahmen des Unternehmenszwecks ausgelöst wurde und keinem Mitarbeiter individuell zugerechnet werden kann (Art. 102 StGB).

Grundsätzlich sind durch Hersteller in Bezug auf KI-Systeme dieselben Sorgfaltspflichten einzuhalten wie bei herkömmlichen technischen Produkten. Je nach Art des KI-Systems, dessen Einsatzbereichs und der daraus folgenden Gefährdung von strafrechtlich geschützten Rechtsgütern können solche Sorgfaltspflichten u.U. allerdings variieren.

Eine Herausforderung beim Einsatz von KI dürfte dabei auch die Beurteilung der strafrechtlichen Verantwortlichkeiten sein, wenn eine Maschine weitgehend automatisiert (bis hin zu autonom) die Kontrolle in einer Situation übernimmt und autonom algorithmenbasierte «Entscheidungen» oder Handlungen vornimmt, welche bisher ausschliesslich durch Menschen ausgeführt wurden (z. B. im Medizinalbereich, im Strassenverkehr oder auch im Rahmen eines modernen Waffeneinsatzes).

Je komplexer ein KI-basiertes System ist, umso schwerer dürfte im Schadensfall feststellbar und durch Gerichte eruierbar sein, wo der Fehler zu lokalisieren ist und wer diesen Fehler schuldhaft verursacht hat (sog. Blackbox-Problematik). Anders als bei herkömmlichen Pro-

---

[390] Zur Möglichkeit einer originären strafrechtlichen Verantwortlichkeit von Technik m. w. V. siehe MONIKA SIMMLER/NORA MARKWALDER, Roboter in der Verantwortung? – Zur Neuauflage der Debatte um den funktionalen Schuldbegriff, Zeitschrift für die gesamte Strafrechtswissenschaft 01/2017, 20 ff., 22; NORA MARKWALDER/MONIKA SIMMLER, Roboterstrafrecht, Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz, Aktuelle Juristische Praxis 02/2017, 171 ff., 172.

[391] ANNA LOHMANN, Strafrecht im Zeitalter von Künstlicher Intelligenz, Der Einfluss von autonomen Systemen und KI auf die tradierten strafrechtlichen Verantwortungsstrukturen, Baden-Baden 2021.

dukten gibt es bei KI-Produkten nämlich verschiedene Entwicklungs- und Automatisierungsstufen, die teilweise eine weitgehende «Selbständigkeit» des Produkts ermöglichen. Darin liegt ein Stück weit der Clou solcher Produkte. Das heisst, dass gewisse Technologien ihr datengetriebenes maschinelles Lernen und die Datenauswertung auf Algorithmen basieren, die sich *autonom* fortentwickeln und anpassen können.[392] Die diesbezüglichen Vorgänge in einem KI-basierten (bzw. «smarten») Produkt sind daher nicht immer *ex ante* vorhersehbar und zuweilen nicht einmal *ex post* transparent und nachvollziehbar.[393] Die konkrete Zuordnung von vorhersehbaren Gefahren und Ursachen und damit auch der strafrechtlichen Verantwortung dürfte daher im Einzelnen oft sehr komplex sein und von einer Vielzahl von spezifischen Umständen eines Falles abhängen. So zum Beispiel vom konkreten Verhalten der Nutzerin oder des Nutzers und der Voraussehbarkeit und Vermeidbarkeit von Systemmängeln. Daher bleiben in Bezug auf diese Elemente schwierige Fragen bestehen, die nicht durchwegs pauschal zu beantworten sind. Angesichts der nach wie vor massgeblichen Entwicklungen im Zusammenhang mit KI-Systemen werden sich Abgrenzungsfragen wohl abhängig von der konkreten technologischen Entwicklung, ihres Einsatzes und konkreten Gefährdungsbereichs näher beantworten und herauskristallisieren.

6.6.2.2 Beispiel: Einsatz von KI-Systemen beim automatisierten Fahren

Beim automatisierten Fahren beispielsweise übernehmen Fahrzeuge weitgehend die Steuerung, was im Falle fehlerhafter Objekterkennung und Kollisionsvermeidung durch die Maschine zu Unfällen führen kann. Solche Unfälle, bei denen nicht nur die Fahrerin oder der Fahrer, sondern auch KI-Systeme im Einsatz sind, stellen eine Herausforderung für die strafrechtliche Verantwortungszurechnung dar. Grundsätzlich muss die oder der Fahrzeuglenkende das Fahrzeug gemäss Artikel 31 SVG jederzeit beherrschen. Dieser Grundsatz kann sich im Zusammenhang mit KI-Systemen allerdings als wesentlich komplexer erweisen und wird im Hinblick auf automatisierte Fahrzeuge zumindest teilweise relativiert.[394] Die Zurechenbarkeit eines Schadens hängt jeweils von den Einflussmöglichkeiten der fahrzeuglenkenden Person ab. Ein Verschulden ist folglich zu bejahen, wenn diese Person es trotz Pflicht, Möglichkeit und Zumutbarkeit unterlässt, das Fehlverhalten eines KI-basierten Systems zu unterbrechen oder zu verhindern bzw. dessen Schaden zu mindern oder rückgängig zu machen. Im Bereich des automatisierten Fahrens besteht insgesamt eine Tendenz, dass sich die Strafbarkeit mit fortschreitender Fahrzeugautomatisierung zunehmend (auch) zum Hersteller hin verlagert (sog. «accountability shift»).[395] Das heisst die strafrechtliche Verantwortung

---

[392] MARCO SCHREYER/ANITA GIERBL/T. FLEMMING RUUD/DAMIAN BORTH, Stichprobenauswahl durch Anwendung von Künstlicher Intelligenz, Expert Focus 2/22, 10 ff., 13 ff.; OMLOR SEBASTIAN, Methodik 4.0 für ein KI-Deliktsrecht, in: Zeitschrift zum Innovations- und Technikrecht 04/2020, 221 ff., 221.

[393] LEA BACHMANN, Prozedurale Entlastung von Herstellern «smarter» Produkte im Strafrecht?, Schweizerische Zeitschrift für Strafrecht 140/2022, 77 ff., 78.

[394] So werden bei automatisierten Fahrzeugen der Stufen 3 und 4 die Lenkerinnen und Lenker teilweise von den Aufmerksamkeits- und Beherrschungspflichten befreit. Die oder der Fahrzeuglenkende von Fahrzeugen der Stufe 3 darf sich zwar bei aktiviertem System anderen Dingen widmen, bleibt aber verpflichtet, die Kontrolle über das Fahrzeug wieder zu übernehmen, sobald das Automatisierungssystem dazu auffordert. Siehe dazu BBl **2021** 3026, S. 12.

[395] NADINE ZURKINDEN, Vertrauen in Fahrzeugautomatisierung als strafmindernder Umstand? Anmerkungen zur Urteilsbegründung des Regionalgerichts Emmental-Oberaargau vom 30. Mai 2018, PEN 17 16 DIP, Jusletter 3. Dezember 2018, 1 ff.; CHRISTOF RIEDO/STEFAN MAEDER, Die Benutzung automatisierter Motorfahrzeuge aus strafrechtlicher Sicht, in: Thomas Probst/Franz Werro (Hrsg.), Strassenverkehrsrechts-Tagung vom 21.–22. Juni 2016, 85 ff., 94.

kann, je nach Fall, sowohl beim Nutzer als auch beim Hersteller (und gemäss dem SVG je nachdem auch beim Halter[396]) liegen.

Der Nutzer eines automatisierten Fahrzeugs kann strafrechtlich zur Verantwortung gezogen werden, wenn er seine gebotenen Sorgfaltspflichten verletzt; das heisst, wenn er trotz Kenntnis oder fahrlässig ein automatisiertes Fahrzeug verwendet oder nicht angemessen überwacht und er in der Pflicht war, einen Fehler zu erkennen und zu verhindern.[397]

Der Hersteller von automatisierten Fahrzeugen trägt demgegenüber (grundsätzlich unabhängig vom Automatisierungsgrad) die Last der Produkthaftung[398] bzw. gewisse Sorgfaltspflichten im Hinblick auf das erlaubte Risiko, das mitunter durch technische Reglemente und Typengenehmigungen definiert wird. Grundsätzlich geht man in diesem Zusammenhang davon aus, dass der Hersteller das Risiko der Gefährdung durch sorgfältige Programmierung und Instruktion weitgehend beherrschen kann.[399] Demzufolge kann der Hersteller für Schäden haftbar gemacht werden, die durch Fehler oder Mängel des Produkts entstehen und ihm ein Verschulden an der Fehlfunktion nachgewiesen werden kann. Dabei geht es nicht nur um Sachschäden, sondern auch um mögliche Schädigungen bzw. Verletzungen von Leib und Leben wie z. B. Körperverletzung oder Tod. Wenn eine Fehlfunktion bekannt war und nicht behoben wurde oder ein Hersteller zum Beispiel um die Schwierigkeiten eines systemgesteuerten Fahrzeugs wusste, stehende Objekte zu erkennen, und er den potentiellen Kunden nicht umfassend instruiert hat, und dadurch ein Unfall generiert wird, kann der Hersteller strafrechtlich grundsätzlich im Zusammenhang mit einer Sorgfaltspflichtverletzung zur Verantwortung gezogen werden.[400] Bei KI-Systemen, die, wie es häufig der Fall sein wird, nach dem Inverkehrbringen aufgrund der breiten Nutzung trainiert werden und die Nutzer durch die Auswahl des Lernverfahrens, der Trainingsdaten und der Dauer des Lernprozesses auf die KI einwirken,[401] kann sich entsprechend die Kontroll- und Einflussmöglichkeit der Hersteller verringern.[402] In Bezug auf automatisierte Fahrzeuge ist allerdings darauf hinzuweisen, dass die Software des KI-Systems jeweils Bestandteil der Typengenehmigung ist und diese die durch das Training erlangten Erkenntnisse nicht selber verwerten darf. Vielmehr muss der Hersteller auf der Basis der neuen Erkenntnisse die Software beispielsweise durch Updates weiterentwickeln und unter Umständen eine neue Typengenehmigung erlangen. Ist die Einflussnahme auf oder das Zusammenwirken der Nutzerin oder des Nutzers mit dem KI-System

---

396 Der Halter hat auch bei automatisierten Fahrzeugen eine originäre Halterverantwortlichkeit, d. h. eine strafrechtliche Verantwortlichkeit für die Überwachung der Gefahrenquellen und das Motorfahrzeug (Art. 93, Abs. 2, Bst. b; 95, Abs. 1, Bst. e und 96, Abs. 3, SVG). Der Halter ist damit grundsätzlich u.a. (mit-)verantwortlich für die Betriebssicherheit des Fahrzeugs.

397 Ausführlicher dazu sowie auch zur Frage bis zu welchem Automatisierungsgrad ein Nutzer noch als Fahrzeugführer i.S.d. SVG verstanden werden kann: CHRISTOF RIEDO/STEFAN MAEDER, Die Benutzung automatisierter Motorfahrzeuge (Fn. 395), 91 f.

398 Relevant in diesem Kontext sind das Bundesgesetz über die Produktesicherheit (PrSG, SR 930.11) und das Produktehaftpflichtgesetz (PrHG, SR 221.112.944).

399 Siehe Überblick bei MELINDA F. LOHMANN, Ein zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz, Warum die Schweiz ihr Produkthaftungsrecht aktualisieren muss, Haftung und Versicherung 04/2021, 120.

400 NADINE ZURKINDEN, Vertrauen in Fahrzeugautomatisierung (Fn. 395), 14 f.

401 HERBERT ZECH, Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung?, Gutachten für den 73. Deutschen Juristentag, München 2020, A 35 ff.; MELINDA F. LOHMANN, Roboter als Wundertüten – eine zivilrechtliche Haftungsanalyse, Aktuelle Juristische Praxis 2017, Sonderheft Roboterrecht, 152 ff., 158.

402 HERBERT ZECH, Entscheidungen (Fn. 401), A 89; MELINDA F. LOHMANN, Roboter (Fn. 401), 158.

allerdings schadensursächlich, können sich diese dennoch auf die strafrechtliche Verantwortlichkeit der Nutzerin oder des Nutzers auswirken.[403]

6.6.3 Herausforderungen im Rahmen der Rechtsdurchsetzung

Die obigen Ausführungen zu den strafrechtlichen Bestimmungen zeigen, dass das geltende materielle Recht aufgrund seiner Technologieneutralität in seiner Grundausrichtung ein grundsätzlich adäquates Instrumentarium bietet, um Strafbarkeiten rund um den Einsatz von KI-Systemen zu erfassen. Das Strafrecht bleibt unabhängig von der technologischen Vorgehensweise oder des spezifischen Instrumentariums einer Täterschaft anwendbar und operabel. Des Weiteren bietet das geltende Strafrecht auch einen Rahmen, der grundsätzliche Sorgfaltspflichten und Verantwortlichkeiten (für Hersteller, Halter oder Nutzer) statuiert. Allerdings bleibt, wie in anderen Rechtsbereichen auch, darauf hinzuweisen, dass es sich bei Straftaten, die unter Zuhilfenahme von KI-Systemen begangen werden oder anderweitig unter Mitwirkung von KI zu strafrechtlich relevanten Schädigungen führen, zuweilen als schwierig erweisen dürfte, die Täterschaft dahinter zu eruieren oder – bei Vorliegen entsprechender Sorgfaltspflichten – die Voraussehbarkeit und Vermeidbarkeit von Systemmängeln nachzuweisen. Die Identifizierung der Täter, die Suche nach Beweisen sowie die klare Eruierung, die Lokalisierung und der Nachweis von Fehlerquellen und Sorgfaltspflichtverletzungen bei automatisierten Produkten wird zweifelsohne auch im strafrechtlichen Kontext eine besondere Herausforderung darstellen. Bei KI-Systemen stellen sich im Hinblick auf die Transparenz und die Nachvollziehbarkeit ihrer automatisierten Entscheidungs- und Lernschritte gerade in der Praxis erhebliche Herausforderungen. Die Schwierigkeiten werden also nicht unbedingt im materiellen Recht selbst begründet, sondern oftmals in der Rechtsdurchsetzung, wenn entweder die Täterschaft, die Technologien und die darauf auffindbaren Beweismittel oder auch die Hersteller sich im Ausland befinden oder *post facto* die genaue Schadensursache nicht konkret ermittelt werden kann (Blackbox-Problematik). Die Schweizer Rechtsdurchsetzung kann ferner auch faktisch an ihre Grenzen kommen, wenn Täter oder Entwickler im Ausland weniger strengen Rechtsvorschriften unterworfen sind oder die komplexen Funktionsweisen und Kausalzusammenhänge von KI-Systemen durch die Staatsanwaltschaft oder Richter schwer zu eruieren, nachzuvollziehen und zu begründen sind. Das heisst, dass einerseits die Blackbox-Problematik von automatisierten Systemen und andererseits die weitgehende Anonymität von Tätern im Ausland (z. B. bei Deepfakes) in der Praxis durchaus Schwierigkeiten mit sich bringen können, die nicht zwingend durch weitergehende materielle Normen angegangen werden können.

Ist die Täterschaft oder ein Hersteller wie beispielhaft angesprochen im Ausland, stellen sich auf der Ebene der Beweisermittlung häufig Schwierigkeiten bezüglich einer funktionierenden, raschen internationalen Zusammenarbeit, und weniger im nationalen Recht selbst. Die Schweiz hat sich folglich an der Ausarbeitung eines Übereinkommens der Vereinten Nationen

[403] NADJA BRAUN BINDER/THOMAS BURRI/MELINDA FLORINA LOHMANN/MONIKA SIMMLER/FLORENT THOUVENIN/KERSTIN NOËLLE VOKINGER, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, Jusletter 28. Juni 2021, 21.

über die Cyberkriminalität beteiligt und ist Vertragspartei der entsprechenden Europaratskonvention vom 23. November 2001[404]. In diesen Abkommen geht es im Hinblick auf eine verbesserte Rechtsdurchsetzung grundsätzlich um die Sicherung und die internationale Zusammenarbeit im Rahmen grenzüberschreitender digitaler Kriminalität. Obschon das erwähnte Übereinkommen über Cyberkriminalität des Europarats, das für die Schweiz am 1. Januar 2012 in Kraft getreten ist, diesbezüglich bereits einige dahingehende internationale Instrumente bietet, sind weitere Entwicklungen noch in Bewegung. So stellen auch die e-Evidence-Regeln der EU einen wichtigen Schritt im Bereich des grenzüberschreitenden Zugriffs auf Daten als Beweismittel im Rahmen von Strafverfahren in Europa dar.[405] Ähnliches verfolgt das System des US CLOUD Act.[406] Die Schweiz setzt sich mit den Entwicklungen in der EU und den USA auseinander und wird zu gegebener Zeit unter Berücksichtigung der politischen und rechtlichen Möglichkeiten entscheiden, wie sie damit umgehen will.

6.6.4 Zusammenfassung und Ausblick

Die obigen Ausführungen zeigen, dass das materielle Strafrecht im Hinblick auf die strafrechtlichen Verantwortlichkeiten und Sorgfaltspflichten beim Einsatz von KI einen grundsätzlich adäquaten und im Einzelfall durchaus anwendbaren Rechtsrahmen bietet. Dabei knüpft das geltende Recht sowohl für Vorsatz- als auch für Fahrlässigkeitsdelikte an das weiterhin und auch im Hinblick auf den Umgang mit KI-Systemen relevante Schuldprinzip an.[407] Das heisst, es kommt im Einzelnen jeweils auf das strafrechtlich relevante, vorwerfbare Verschulden der oder des potenziell Verantwortlichen an. Insbesondere bei Vorsatzdelikten, die unter Zuhilfenahme von KI-Technologien wie z. B. von Deepfakes auf strafrechtlich relevante Schädigungen oder Vermögensvorteile abzielen, stellt i. d. R. weniger die materielle Rechtslage, sondern vielmehr die Rechtsdurchsetzung eine Herausforderung dar, da Täter ihre digitalen Spuren oftmals verschleiern und sich im Ausland oder hinter Deepfakes «verstecken» können.

Für die fahrlässige Strafbarkeit im Zusammenhang mit einem KI-System spielt es hingegen primär eine Rolle, ob deren Einsatz, die Herstellung oder die Inverkehrbringung per se als Sorgfaltspflichtwidrigkeit oder als sogenanntes erlaubtes Risiko qualifiziert wird sowie ob und welche Sorgfaltspflichten und Fehlerquellen *ex ante* erkennbar und vermeidbar waren.[408] Damit ist eine verschuldensunabhängige Kausalhaftung im Schadensfall im Strafrecht grundsätzlich ausgeschlossen.[409] Eine Person kann folglich für eine Sorgfaltspflichtverletzung nur

---

[404] SR **0.311.43**.

[405] Siehe den Bericht zur e-Evidence-Vorlage der EU, Gutachten des Bundesamts für Justiz vom 24. Oktober 2023 (abrufbar unter: www.bj.admin.ch > Publikationen & Service > Berichte, Gutachten und Verfügungen > Berichte und Gutachten > Bericht zur e-Evidence-Vorlage der EU, abgerufen am 29. August 2024).

[406] Siehe den Bericht zum US CLOUD Act, Gutachten des Bundesamts für Justiz vom 17. September 2021 (abrufbar unter: www.bj.admin.ch > Publikationen & Service > Berichte, Gutachten und Verfügungen > Berichte und Gutachten > Bericht zum US CLOUD Act, abgerufen am 29. August 2024) sowie das darauf basierende «Data Access Agreement» zwischen den USA und dem Vereinigten Königreich, das im Oktober 2022 in Kraft trat.

[407] MONIKA SIMMLER, Strafrechtliche Verantwortung im Zeitalter autonomer Technik: Vom Individual- zum Unternehmensstrafrecht?, in: Daniel Fink u. a. (Hrsg.), Strafjustiz zwischen künstlicher Intelligenz und prädiktiven Algorithmen, Basel 2021.

[408] BRAUN BINDER u. a., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (Fn. 403), 23. Zum erlaubten Risiko beim Einsatz von KI-basierten Systemen bei selbstfahrenden Fahrzeugen: NADINE ZURKINDEN, Strafrecht und selbstfahrende Autos – ein Beitrag zum erlaubten Risiko, recht 2016, 144 ff.

[409] In bestimmten Bereichen, wie z. B. bei der zivilrechtlichen Haftung im Zusammenhang mit automatisierten Fahrzeugen, können beim Einsatz von künstlicher Intelligenz hingegen gewisse Konstellationen einer verschuldensunabhängigen Nutzer- oder Halterhaftung entstehen (siehe

strafrechtlich verantwortlich gemacht werden, wenn diese für sie erkennbar und vermeidbar gewesen wäre. Wo die jeweiligen Grenzen (insb. zwischen der Nutzer-, der Hersteller- und der Halterverantwortlichkeiten) konkret verlaufen, ist im Einzelnen allerdings nicht durchwegs klar[410] bzw. wird es schwierig und unter Umständen auch nicht sinnvoll sein, diese Abgrenzung im Hinblick auf zunehmend selbständige Technologien und Prozesse pauschal festzulegen. Demzufolge bleiben in Bezug auf die konkrete Auslegung und Anwendung dieser Sorgfaltsplichten in Zukunft durchaus Herausforderungen bestehen. In dieser Hinsicht wird insbesondere auch die Rechtsprechung wichtig bleiben.

Die jüngeren Entwicklungen im Bereich des automatisierten Fahrens sprechen grundsätzlich dafür oder deuten zumindest darauf hin, dass sich strafrechtliche Verantwortlichkeiten gerade mit Blick auf die zunehmende Fahrzeugautomatisierung stärker auf den Hersteller oder Personen, die im Hersteller-Unternehmen arbeiten, verschieben.[411] Hier werden naheliegenderweise gerade in Bezug auf weitgehend automatisierte bis hin zu autonomen Systemen wohl weiterhin praktische Abgrenzungsfragen im Hinblick auf die Risiko- und Einflusssphären von Herstellern und Nutzern[412] (und je nach Bereich auch Haltern) sowie in Bezug auf die konkreten Sorgfaltspflichten im Kontext der Unternehmensstrafbarkeit oder der fahrlässigen Mittäterschaft bestehen bleiben.[413] In dieser Hinsicht kann eine Klärung der Sorgfaltspflichten bei der Herstellungssicherheit, der Zulassung und beim Einsatz von KI beispielsweise via konkreter Zulassungs- oder Sicherheitsanforderungen eine bessere Voraussehbarkeit und damit auch Rechtssicherheit schaffen. Damit gemeint sind insbesondere Sorgfaltspflichten für den Einsatz und das Inverkehrbringen von KI-Systemen durch Private oder in Unternehmen sowie im Zusammenhang mit der Produktesicherheit.

Im Zusammenhang mit dem Einsatz von KI-Systemen im Strassenverkehr hat der Bundesrat an seiner Sitzung vom 18. Oktober 2023 die Vernehmlassung über zwei neue Verordnungen eröffnet, mit denen er das automatisierte Fahren regeln will. Die Vernehmlassung dauerte bis

---

Art. 58 SVG). Diese sind jedoch von strafrechtlichen Verantwortlichkeiten zu unterscheiden. Dazu MELINDA F. LOHMANN, Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts, Baden-Baden 2016, 211 ff.

410 BRAUN BINDER u. a., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (Fn. 403), 23. Zu ähnlichen Fragen, die sich in Bezug auf (autonome) soziale Roboter stellen, siehe MONIKA SIMMLER/OLIVIA ZINGG, Rechtliche Aspekte sozialer Roboter, Gutachten im Auftrag der TA-SWISS, 2011, 19 ff. und 48 ff.

411 Siehe dazu NADINE ZURKINDEN, Vertrauen in Fahrzeugautomatisierung (Fn. 395), 1 ff. mit Verweis auf die Urteilsbegründung des Regionalgerichts Emmental-Oberaargau vom 30. Mai 2018, PEN 17 16 DIP. Zu dieser Verlagerungstendenz auch: MONIKA SIMMLER, Strafrechtliche Verantwortung (Fn. 407).

412 BRAUN BINDER u. a., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (Fn. 403), 21.

413 Zu den entsprechenden Herausforderungen MONIKA SIMMLER/OLIVIA ZINGG, Rechtliche Aspekte sozialer Roboter (Fn. 410), 55 ff.; MONIKA SIMMLER, Strafrechtliche Verantwortung (Fn. 407); BRAUN BINDER u. a., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (Fn. 403), 23.

zum 2. Februar 2024. Einerseits geht es um die im vorliegenden Kontext relevante Verordnung über das automatisierte Fahren (AFV) und andererseits um die Verordnung über die Finanzhilfen zur Förderung neuartiger Lösungen für den Verkehr auf öffentlichen Strassen (ÖStFV).[414] Dadurch sollen verschiedene Sorgfaltspflichten für den Hersteller, den Händler, den Fahrzeughalter und Nutzer statuiert und die Verletzung besonders wichtiger Pflichten mit Busse bestraft werden. Unter dem Hinweis auf die absehbare dynamische Entwicklung im Bereich des automatisierten Fahrens werden folglich die Auswirkungen dieser Regulierungen vorab abzuwarten und ihr Mehrwert und Zusammenspiel mit internationalen Entwicklungen in diesem Bereich zu evaluieren sein.

Zusammenfassend lässt sich sagen, dass das Schweizer Strafrecht grundsätzlich mit seinem technologieneutralen Ansatz ein generell geeignetes Instrument bietet, um den Einsatz von KI-Systemen insbesondere bei vorsätzlichen Straftaten zu erfassen.

Ein Grossteil der Herausforderungen liegt folglich primär bei (praktischen) Abgrenzungsfragen in Bezug auf die Verantwortlichkeiten, auf der Ebene der Rechtsanwendung sowie generell bei der Rechtsdurchsetzung. Dabei zieht die Konkretisierung von Sorgfaltspflichten im Umgang mit KI-Systemen, wie sie in der erwähnten Verordnung zum automatisierten Fahren (AFV) vorgesehen sind, in eine richtige Richtung, um zumindest in diesem Bereich mehr Klarheit zu schaffen. Ein darüberhinausgehendes Eingreifen des Gesetzgebers schiene daher zum jetzigen Zeitpunkt verfrüht und müsste jeweils auch im Kontext eines entsprechenden Anwendungsbereichs evaluiert werden. Die in der AFV anvisierten Sorgfaltspflichten würden zudem bereits in ihrer aktuellen Form dazu beitragen, die Grundsätze der Transparenz und Rechenschaftspflicht (Art. 8 und 9) umzusetzen, die sich aus der Konvention des Europarats über KI ergeben (siehe die Kapitel 4.3.2.3 und 4.3.2.4), falls die Schweiz dieses Abkommen ratifiziert. Eine erhöhte Transparenz könnte in der Folge wiederum – jeweils im Rahmen des Zumutbaren und Möglichen – zumindest teilweise auch der Blackbox-Problematik entgegenwirken.

Angesichts der sich in Bewegung befindlichen Entwicklungen auf schweizerischer, europäischer und internationaler Ebene werden die Analyse und daraus resultierende Folgerungen im Hinblick auf einen konkreten, über das Bestehende und Anvisierte hinausgehende Regulierungsbedarf weiterhin wichtig sein. So hat beispielsweise der Europarat im Anschluss an seine 78. Plenarsitzung des European Committee on Crime Problems (CDPC) im November 2020 auch eine Regulierung von KI-Systemen im Kontext des Strafrechts anvisiert.[415] Es ist folglich nicht auszuschliessen, dass neben der KI-Konvention auch themenspezifischere Diskussionen und Entwicklungen folgen könnten.

[414] Medienmitteilung vom 18. Oktober 2023, Der Bundesrat will automatisiertes Fahren ermöglichen, abrufbar unter: www.admin.ch > Dokumentation > Medienmitteilungen > Der Bundesrat will automatisiertes Fahren ermöglichen (abgerufen am 29. August 2024).

[415] Siehe Artificial Intelligence and Criminal Law - European Committee on Crime Problems (https://www.coe.int/en/web/cdpc/artificial-intelligence-and-criminal-law, abgerufen am 29. August 2024).

7 Schlussfolgerungen

Die Analyse hat bestätigt, dass das Schweizer Recht bereits Bestimmungen enthält, die auch auf KI-Systeme anwendbar sind. KI entwickelt sich also nicht in einem rechtlichen Vakuum.

An dieser Stelle ist darauf hinzuweisen, dass das BJ nicht das gesamte bestehende Bundesrecht umfassend analysiert hat. Die Analyse enthält daher meist nur Zwischenergebnisse, die noch vertieft werden müssen.

Wie in der Methodik erwähnt und zur Vermeidung der Wiederholung der Analyse im Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» von 2019[416] wurde der regulatorische Handlungsbedarf hauptsächlich unter der Annahme einer Ratifizierung der KI-Konvention des Europarats durch die Schweiz beurteilt. Die Analyse stellt zudem rechtliche Überlegungen zur KI-Verordnung im Falle einer Annäherung der Schweizer Gesetzgebung an das EU-Recht. Dies hängt jedoch von politischen Entscheidungen ab. Im Rahmen der Analyse wurden auch weitere ausgewählte Rechtsgebiete untersucht, in denen Entwicklungen und Herausforderungen bezüglich KI bestehen.

Hinsichtlich KI-Konvention scheint das Schweizer Recht im Falle einer Ratifizierung eine teilweise Umsetzung ermöglichen zu können. Es müssten jedoch regulatorische Eingriffe erwogen werden insbesondere zur Stärkung der Transparenz, des Rahmens für das Risiko- und Folgenmanagement sowie der Aufsichtsmechanismen (vgl. Ziff. 4.6). In Bezug auf den Privatsektor hat die Analyse gezeigt, dass sich der Anwendungsbereich der KI-Konvention auf die Fälle beschränkt, in denen eine direkte oder indirekte horizontale Wirkung der Grundrechte zwischen Privaten besteht oder in Zukunft erkannt wird (vgl. Ziff. 4.2.3.1).

Eine Annäherung des schweizerischen Rechts an die KI-Verordnung würde ein umfassendes Eingreifen des Gesetzgebers erfordern. Die KI-Verordnung enthält sehr spezifische Vorschriften für die verschiedenen KI-Systeme, je nach den von ihnen ausgehenden Risiken, sowie zahlreiche Verpflichtungen für die verschiedenen Akteure. Derzeit sind noch viele Fragen offen, insbesondere politische Fragen wie die, ob die Einführung gleichwertiger Rechtsvorschriften einen effizienteren Zugang zum EU-Markt ermöglichen würde, z. B. durch die gegenseitige Anerkennung von KI-Rechtsvorschriften im Rahmen einer Ausweitung des MRA auf den Bereich der KI (vgl. Ziff. 5.4). Weitergehende Analysen, die z. B. die Konturen einer möglichen schweizerischen Gesetzgebung in Anlehnung an die KI-Verordnung präzisieren könnten, wären ebenfalls notwendig, wenn der politische Wille dazu vorhanden ist, in diese Richtung zu gehen. Sie würden sich auch auf die EU-Richtlinien im Bereich des Privatrechts beziehen (vgl. Ziff. 6.3 ff.).

In den untersuchten ausgewählten Rechtsgebieten (vgl. Ziff. 6) hat die Analyse gezeigt, dass sich zwar gewisse Fragen stellen, die sich jedoch mit den geltenden Vorschriften grundsätzlich beantworten lassen. Ein besserer Schutz könnte auch durch die Verabschiedung allgemeiner Normen zur Umsetzung der KI-Konvention erreicht werden, z. B. durch die Stärkung der Transparenz von KI-Systemen.

---

[416] Bericht «Herausforderungen der künstlichen Intelligenz» (Fn. 1).

Um die rechtliche Analyse zu vertiefen und den Handlungsbedarf zu präzisieren, müssen zunächst die politischen Antworten auf die beiden oben genannten Arbeitshypothesen, d. h. die Ratifizierung der KI-Konvention durch die Schweiz und eine allfällige Annäherung an die europäische KI-Verordnung, vorliegen.

Abkürzungsverzeichnis

ABl.	Amtsblatt der Europäischen Union
Abs.	Absatz
AFV	Verordnung über das automatisierte Fahren
AHVG	Bundesgesetz über die Alters- und Hinterlassenenversicherung, SR **831.10**
ArG	Bundesgesetz über die Arbeit in Industrie, Gewerbe und Handel, SR **822.11**
ArGV 3	Verordnung 3 zum Arbeitsgesetz, SR **822.113**
Art.	Artikel
ASTRA	Bundesamt für Strassen
BAKOM	Bundesamt für Kommunikation
BBl	Bundesblatt
BehiG	Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen, SR **151.3**
BFS	Bundesamt für Statistik
BGE	Amtliche Sammlung der Entscheidungen des Schweizerischen Bundesgerichts
BGer	Bundesgericht
BGÖ	Bundesgesetz über das Öffentlichkeitsprinzip der Verwaltung, SR **152.3**
BJ	Bundesamt für Justiz
BPR	Bundesgesetz über die politischen Rechte, SR **161.1**
BSK [Gesetz]- AUTOR	Basler Kommentar
Bst.	Buchstabe
BV	Bundesverfassung der Schweizerischen Eidgenossenschaft, SR **101**

CAI	Committee on Artificial Intelligence / Ausschuss für künstliche Intelligenz
CEN	Europäisches Komitee für Normung
CENELEC	Europäisches Komitee für elektrotechnische Normung
CNAI	Kompetenznetzwerk für künstliche Intelligenz
CO	Loi fédérale complétant le Code civil suisse (Livre cinquième : Code des obligations), RS **220**
CP	Code pénal suisse, RS **311.0**
CR [Gesetz]-AUTOR	Commentaire romand
Cst.	Constitution fédérale de la Confédération suisse, RS **101**
DSG	Bundesgesetz über den Datenschutz, SR **235.1**
DSV	Verordnung über den Datenschutz, SR **235.11**
DSGVO	Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)
DV	Direktion für Völkerrecht
E.	Erwägung
EDA	Eidgenössisches Departement für auswärtige Angelegenheiten
EDI	Eidgenössisches Departement des Innern
EDÖB	Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter
EGMR	Europäischer Gerichtshof für Menschenrechte

EJPD	Eidgenössisches Justiz- und Polizeidepartement
EMBAG	Bundesgesetz über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben, SR **172.019**
EMRK	Europäische Konvention zum Schutze der Menschenrechte und Grundfreiheiten, SR **0.101**
et al.	*et alii*
ETSI	Europäisches Institut für Telekommunikationsnormen
EU	Europäische Union
EuGH	Gerichtshof der Europäischen Union
EUV	Vertrag über die Europäische Union
f.	folgende
ff.	fortfolgende
Fn.	Fussnote (innerhalb dieses Dokuments)
GIG	Bundesgesetz über die Gleichstellung von Frau und Mann, SR **151.1**
GPAI-Modell	general-purpose AI model / KI-Modell mit allgemeinem Verwendungszweck
Hrsg.	Herausgeber
IGE	Eidgenössisches Institut für Geistiges Eigentum
KI	künstliche Intelligenz
KI-Konvention	Rahmenübereinkommen des Europarats über künstliche Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit
KI-Verordnung	Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU)

	Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)
Konvention 108+	Aktualisiertes Übereinkommen des Europarats zum Schutz des Menschen bei der automatischen Verarbeitung personenbezogener Daten
LFG	Bundesgesetz über die Luftfahrt, SR **748.0**
LPD	Loi fédérale sur la protection des données, RS **235.1**
Mo.	Motion
MRA	Abkommen zwischen der Schweizerischen Eidgenossenschaft und der Europäischen Gemeinschaft über die gegenseitige Anerkennung von Konformitätsbewertungen, SR **0.946.526.81**
N / Nr	Nummer
NDB	Nachrichtendienst des Bundes
OECD	Organisation for Economic Cooperation and Development / Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
OR	Bundesgesetz betreffend die Ergänzung des Schweizerischen Zivilgesetzbuches (Fünfter Teil: Obligationenrecht), SR **220**
ÖStFV	Verordnung über die Finanzhilfen zur Förderung neuartiger Lösungen für den Verkehr auf öffentlichen Strassen
PatG	Bundesgesetz über die Erfindungspatente, SR **232.14**
PrHG	Bundesgesetz über die Produktehaftpflicht, SR **221.112.944**
Ref.	Referenz

RTVG	Bundesgesetz über Radio und Fernsehen, SR **784.40**
RTVV	Radio und Fernsehverordnung, SR **784.401**
SECO	Staatssekretariat für Wirtschaft
SGK [Gesetz]-AUTOR	St. Galler Kommentar
SiK-N	Sicherheitspolitische Kommission des Nationalrates
SKMR	Schweizerisches Kompetenzzentrum für Menschenrechte
SR	Systematische Rechtssammlung
StGB	Schweizerisches Strafgesetzbuch, SR **311.0**
SVG	Strassenverkehrsgesetz, SR **741.01**
TA-SWISS	Stiftung für Technologiefolgen-Abschätzung
UNCITRAL	United Nations Commission on International Trade Law / Kommission der Vereinten Nationen für internationales Handelsrecht
UNO-Pakt I	Internationaler Pakt über wirtschaftliche, soziale und kulturelle Rechte, SR **0.103.1**
UNO-Pakt II	Internationaler Pakt über bürgerliche und politische Rechte, SR **0.103.2**
URG	Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, SR **231.1**
usw.	und so weiter
UVEK	Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation
VDSZ	Verordnung über Datenschutzzertifizierungen, SR **235.13**
vgl.	vergleiche

VIG	Bundesgesetz über das Vernehmlassungsverfahren, SR **172.061**
VwVG	Bundesgesetz über das Verwaltungsverfahren, SR **172.021**
ZG	Zollgesetz, SR **631.0**
ZGB	Schweizerisches Zivilgesetzbuch, SR **210**
Ziff.	Ziffer
ZPO	Schweizerische Zivilprozessordnung, SR **272**
z. B.	zum Beispiel

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Justiz- und Polizeidepartement EJPD

Bundesamt für Justiz BJ
Direktionsbereich Öffentliches Recht
Fachbereich Rechtsetzungsprojekte I

Anhang 1

Die nachstehende Tabelle soll, soweit möglich, einen Vergleich zwischen der KI-Konvention und der KI-Verordnung ermöglichen. Die Tabelle erhebt keinen Anspruch auf Vollständigkeit. Die Artikel der KI-Konvention sind in der linken Spalte aufgeführt, während die rechte Spalte die Bestimmungen der KI-Verordnung, die einen Zusammenhang mit den betreffenden Artikeln der KI-Konvention haben, enthält. Von der KI-Verordnung werden nur die relevantesten Bestimmungen aufgeführt.

KI-Konvention des Europarats	**KI-Verordnung der Europäischen Union**
Art. 1 Abs. 1 Ziel und Zweck	Art. 1 Abs. 1 Gegenstand
Art. 1 Abs. 2 Abgestufte und differenzierte Massnahmen	Risikobasierter Ansatz (Art. 5, Art. 6 ff., Art. 50, Art. 51 ff.)
Art. 2 Begriffsbestimmung «System künstlicher Intelligenz»	Art. 3 Nr. 1 Die Definition ist im Wesentlichen die gleiche.
Art. 3 Anwendungsbereich	Art. 2 Anwendung im öffentlichen und privaten Sektor. Wie die KI-Konvention gilt auch die KI-Verordnung nicht für KI-Systeme, die ausschliesslich für militärische Zwecke oder Zwecke der nationalen Sicherheit oder Verteidigung eingesetzt werden. Ausschluss von Forschung und Entwicklung. Die KI-Verordnung sieht noch andere Ausnahmen vor.
Art. 4 Schutz der Menschenrechte	Art. 1 Abs. 1 Eines der Ziele der KI-Verordnung ist es, das Funktionieren des EU-Binnenmarktes in Bezug auf KI-Produkte zu verbessern und gleichzeitig den Schutz der Grundrechte zu gewährleisten. Sehr wenige individuelle Rechte, aber Ziel des Grundrechtsschutzes durch Festlegung von Produktsicherheitsregeln. Art. 77 Befugnisse der für den Schutz der Grundrechte zuständigen Behörden

Art. 5 Integrität demokratischer Prozesse und Achtung der Rechtsstaatlichkeit	Art. 1 Abs. 1 Eines der Ziele der KI-Verordnung ist es, das Funktionieren des EU-Binnenmarktes in Bezug auf KI-Produkte zu verbessern und gleichzeitig den Schutz der Grundrechte zu gewährleisten. Sehr wenige individuelle Rechte, aber Ziel des Grundrechtsschutzes durch Festlegung von Produktsicherheitsregeln.
Art. 6 Allgemeiner Ansatz	Nicht relevant
Art. 7 Menschenwürde und individuelle Autonomie	Art. 5 Verbotene Praktiken im KI-Bereich
Art. 8 Transparenz und Aufsicht	Art. 8 ff. Anforderungen an Hochrisiko-KI-Systeme, siehe insbesondere Art. 11 Technische Dokumentation, Art. 12 Aufzeichnungspflichten, Art. 13 Transparenz und Bereitstellung von Informationen für die Betreiber Art. 26 Pflichten der Betreiber (z. B. Informationspflicht). Art. 50 Transparenzpflichten Art. 49 und 71 Registrierung und Datenbank Art. 86 Recht auf Erläuterung der Entscheidungsfindung im Einzelfall
Art. 9 Rechenschaftspflicht und Verantwortung	Pflichtenkatalog für die verschiedenen Akteure in der Lieferkette von KI-Systemen (Anbieter, Betreiber usw.). Art. 17 Abs. 1, Bst. m Rechenschaftsrahmen für die interne Verantwortlichkeit Art. 99 ff. Sanktionen
Art. 10 Gleichstellung und Nichtdiskriminierung	Art. 10 Daten und Daten-Governance, insb. Art. 10 Abs. 2 Bst. f Art. 15 Abs. 4 Genauigkeit, Robustheit und Cybersicherheit
Art. 11 Privatsphäre und Schutz personenbezogener Daten	Art. 10 Daten und Daten-Governance Art. 15 Abs. 4 Genauigkeit, Robustheit und Cybersicherheit
Art. 12 Zuverlässigkeit	Art. 14 Menschliche Aufsicht Art. 15 Genauigkeit, Robustheit und Cybersicherheit

	Art. 40 ff. Harmonisierte Normen Art. 56 ff. Praxisleitfäden Art. 95 Verhaltenskodizes
Art. 13 Sichere Innovation	Art. 57 ff Massnahmen zur Innovationsförderung
Art. 14 Rechtsmittel	Art. 85 Beschwerde Art. 86 Recht auf Erläuterung der Entscheidungsfindung im Einzelfall
Art. 15 Verfahrensgarantien	Art. 26 Pflichten der Betreiber (z. B. Informationspflicht) Art. 50 Transparenzpflichten Art. 86 Recht auf Erläuterung der Entscheidungsfindung im Einzelfall
Art. 16 Rahmen für das Risiko- und Folgenmanagement	Art. 9 Risikomanagementsystem Art. 11 Technische Dokumentation Art. 12 Aufzeichnungspflichten Art. 16 Qualitätsmanagementsystem Art. 27 Grundrechte-Folgenabschätzung für Hochrisiko-KI-Systeme Art. 5 Verbotene Praktiken
Art. 17 Nichtdiskriminierung	Nicht relevant, betrifft die Umsetzung der KI-Konvention
Art. 18 Rechte von Menschen mit Behinderungen und von Kindern	Nicht relevant, betrifft die Umsetzung der KI-Konvention
Art. 19 Öffentliche Konsultation	Art. 67 Beratungsforum Beteiligung aller relevanten Interessenträger an der Ausarbeitung harmonisierter Normen (Art. 40 ff.), Praxisleitfäden (Art. 56), Verhaltenskodizes (Art. 95) usw.
Art. 20 Digitale Kompetenzen und Fähigkeiten	Art. 4 KI-Kompetenz Art. 13 Bereitstellung von Informationen für die Betreiber
Art. 21 Schutz der bestehenden Menschenrechte	Die KI-Verordnung stellt an mehreren Stellen klar, dass sie zusätzlich zum bestehenden Rechtsrahmen gilt und diesen nicht einschränken soll. Dies gilt insbesondere für den Datenschutz (vgl. E. 10 der Verordnung).

Art. 22 Umfassenderer Schutz	Nicht relevant
Art. 23 Konferenz der Vertragsparteien	Nicht relevant
Art. 24 Berichterstattungspflicht	Nicht relevant
Art. 25 Internationale Zusammenarbeit	Art. 57 Abs. 13 Grenzüberschreitende Zusammenarbeit Reallabore
Art. 26 Wirksame Aufsichtsmechanismen	Art. 28 ff. Notifizierende Behörden und notifizierte Stelle Art. 40 ff. Normen und Konformitätsbewertung Art. 64 ff. Governance Art. 74 ff. Marktüberwachung Art. 99 ff. Sanktionen
Art. 27 ff. Schlussbestimmungen	Nicht relevant

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de justice et police DFJP

Office fédéral de la justice OFJ
Domaine de direction Droit public
Unité Projets législatifs I

31 août 2024

Analyse juridique de base dans le cadre de l'état des lieux sur les approches de régulation en matière d'intelligence artificielle

BJ-D-F0D63401/159

Condensé

La présente analyse juridique est réalisée par le DFJP (OFJ) dans le cadre du mandat du Conseil fédéral du 22 novembre 2023 de réaliser un état des lieux sur les approches de régulation en matière d'intelligence artificielle (IA). Ce dernier doit notamment identifier les besoins de réglementation et des approches de régulation, qui tiennent compte de la convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (convention sur l'IA) et du règlement de l'Union européenne proposant des règles harmonisées concernant l'intelligence artificielle (règlement sur l'IA). Il ressort de l'analyse juridique que le droit suisse est pleinement applicable dans le domaine de l'IA, mais que des besoins d'adaptation seraient nécessaires dans l'hypothèse de la ratification de la convention sur l'IA par la Suisse, et de manière plus générale pour appréhender les défis de l'IA. Un rapprochement avec le règlement sur l'IA nécessiterait une intervention plus large du législateur.

Contexte

La présente analyse juridique doit servir d'appui à l'état des lieux sur les approches de régulation en matière d'IA mandaté par le Conseil fédéral le 22 novembre 2023. Le mandat du Conseil fédéral part du postulat qu'il y a lieu de se pencher à nouveau sur le besoin de légiférer en Suisse en matière d'IA. En effet, depuis 2019 et le rapport du groupe de travail interdépartemental « Intelligence artificielle », qui concluait que la législation en vigueur était suffisante, la situation a bien changé. L'utilisation des systèmes d'IA s'est en particulier étendue dans des domaines de plus en plus variés du secteur privé comme du secteur public. Au plan international, des textes ont également émergé. Tel est le cas en particulier de la convention du Conseil de l'Europe sur l'IA et du règlement de l'UE sur l'IA.

L'analyse juridique a pour but d'identifier de possibles lacunes du cadre juridique suisse compte tenu des défis posés par l'IA. Elle vise à déterminer le besoin de légiférer compte tenu du droit actuel en fonction des options politiques qui pourraient être prises par le Conseil fédéral. Elle a été réalisée par l'OFJ, avec des contributions de la DDIP (pour les ch. 3 et 5.2.11), de l'OFCOM (pour les chapitres 5.2 et 5.3.2) et de l'IPI (pour le ch. 6.2).

Contenu

L'analyse juridique se base tout d'abord sur la convention du Conseil de l'Europe sur l'IA, qui lierait la Suisse en cas de ratification. En tant que premier traité international sur l'IA, cette convention recense les principaux enjeux juridiques posés par l'IA en matière de protection des droits de l'homme, de la démocratie et de l'État de droit, et son examen permet ainsi d'identifier d'éventuelles lacunes du droit suisse eu égard à ces problématiques reconnues au niveau international. L'analyse porte sur le droit international et fédéral, à l'exclusion du droit cantonal et communal. L'analyse juridique présente dans un second temps quelle serait la situation en cas de rapprochement du droit suisse avec le règlement de l'UE sur l'IA. Elle examine également quelques autres domaines du droit, qui ne sont pas appréhendés de manière exhaustive ni par la convention sur l'IA, ni par le règlement sur l'IA, comme le droit de la propriété intellectuelle, et le droit de la responsabilité civile et pénale.

Les conclusions de l'analyse juridique confirment que le droit suisse contient déjà des normes qui règlent l'IA. Ces normes permettraient de transposer en partie la convention sur l'IA. Des

compléments semblent toutefois nécessaires, notamment pour améliorer la transparence des systèmes d’IA, analyser leur impact sur les droits fondamentaux, et pour assurer des mécanismes de contrôle. En cas de volonté de rapprochement avec la règlementation de l’UE sur l’IA, des règles plus détaillées devraient être adoptées. En effet, le règlement sur l’IA constitue principalement une règlementation sur la sécurité des produits, prévoyant des obligations spécifiques à la charge des différents opérateurs dans le domaine de l’IA. Pour ce qui est des autres domaines du droit qui ont été examinés, l’analyse a démontré que certaines questions se posent, mais qu’en principe les règles en vigueur apportent des réponses. En outre, l’adoption de normes générales pour se conformer à la convention sur l’IA, par exemple renforçant la transparence des systèmes d’IA, aurait souvent pour effet d’améliorer la protection dans ces domaines, par exemple en matière de responsabilité civile et pénale.

L’analyse juridique se focalise sur les grands enjeux et présente le plus souvent des conclusions qu’il conviendra encore d’approfondir eu égard aux choix politiques qui seront effectués, notamment s’agissant d’une éventuelle ratification de la convention du Conseil de l’Europe sur l’IA, ou d’un éventuel rapprochement avec la législation de l’UE.

L’analyse juridique est réalisée en tenant compte des développements intervenus jusqu’au 31 août 2024.

Table des matières

Condensé **2**
Table des matières **4**
1 Introduction **8**
2 But et méthodologie de l'analyse juridique de base **8**
3 Instruments au niveau international **11**
3.1 Traités internationaux 11
3.2 Instruments non-contraignants dans le domaine de l'IA 12
4 Convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit du Conseil de l'Europe **13**
4.1 Remarques introductives 13
4.1.1 Contexte 13
4.1.2 Applicabilité et justiciabilité 13
4.1.3 Mise en œuvre en droit interne et fédéralisme 15
4.2 Chapitre I : Dispositions générales 16
4.2.1 Article 1 – Objet et but 16
4.2.2 Définitions 17
4.2.2.1 Article 2 – Définition des systèmes d'intelligence artificielle 17
4.2.2.2 Cycle de vie 20
4.2.3 Article 3 – Champ d'application 21
4.2.3.1 Secteur public et secteur privé 21
4.2.3.2 Sécurité nationale 26
4.2.3.3 Recherche et développement 27
4.2.3.4 Défense nationale 28
4.2.3.5 En résumé 28
4.3 Obligations et principes 29
4.3.1 Chapitre II : Obligations générales 29
4.3.1.1 Article 4 – Protection des droits de l'homme 29
4.3.1.2 Article 5 – Intégrité des processus démocratiques et respect de l'État de droit 30
4.3.2 Chapitre III : Principes relatifs aux activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle 37
4.3.2.1 Article 6 – Approche générale 37
4.3.2.2 Article 7 – Dignité humaine et autonomie personnelle 38
4.3.2.3 Article 8 – Transparence et contrôle 40
4.3.2.4 Article 9 – Obligation de rendre des comptes et responsabilité 44
4.3.2.5 Article 10 – Égalité et non-discrimination 46
4.3.2.6 Article 11 – Respect de la vie privée et protection des données à caractère personnel 55
4.3.2.7 Article 12 – Fiabilité 59
4.3.2.8 Article 13 – Innovation sûre 61
4.3.3 Chapitre IV : Recours 65
4.3.3.1 Généralités 65
4.3.3.2 Article 14 – Recours 65

4.3.3.3 Article 15 – Garanties procédurales ... 71
4.3.4 Chapitre V : Évaluation et atténuation des risques et des impacts négatifs ... 74
4.3.5 Chapitre VI : Mise en œuvre de la convention ... 78
4.3.5.1 Généralités ... 78
4.3.5.2 Article 17 – Non-discrimination ... 78
4.3.5.3 Article 18 – Droits des personnes handicapées et des enfants ... 78
4.3.5.4 Article 19 – Consultation publique ... 79
4.3.5.5 Article 20 – Maîtrise du numérique et compétences numériques ... 80
4.3.5.6 Article 21 – Sauvegarde des droits de l'homme reconnus ... 81
4.3.5.7 Article 22 – Protection plus étendue ... 81
4.4 Chapitre VII : Mécanisme de suivi et coopération ... 81
4.4.1 Généralités ... 81
4.4.2 Article 23 – Conférence des Parties ... 81
4.4.3 Article 24 – Obligation de rapport ... 82
4.4.4 Article 25 – Coopération internationale ... 82
4.4.5 Article 26 – Mécanismes de contrôle effectifs ... 83
4.5 Chapitre VIII : Clauses finales ... 84
4.6 Conclusions intermédiaires ... 84
5 Règlement de l'Union européenne établissant des règles harmonisées concernant l'intelligence artificielle ... 87
5.1 Structure du chapitre et méthode ... 87
5.2 Contenu du règlement ... 87
5.2.1 Contexte ... 87
5.2.2 Objectifs de la réglementation ... 88
5.2.3 Approche fondée sur les risques ... 89
5.2.4 Définitions ... 90
5.2.4.1 Système d'intelligence artificielle ... 90
5.2.4.2 Modèle d'intelligence artificielle à usage général ... 90
5.2.5 Champ d'application ... 91
5.2.5.1 Étendue du champ d'application ... 91
5.2.5.2 Exceptions ... 92
5.2.6 Pratiques interdites ... 94
5.2.7 Systèmes d'intelligence artificielle à haut risque ... 96
5.2.7.1 Classification ... 96
5.2.7.2 Exigences applicables aux systèmes d'IA à haut risque ... 98
5.2.7.3 Obligations incombant aux fournisseurs et à d'autres parties ... 100
5.2.7.4 Évaluation de la conformité ... 104
5.2.8 Obligations de transparence pour les fournisseurs et les déployeurs de certains systèmes d'IA ... 106
5.2.9 Autres systèmes d'intelligence artificielle ... 108
5.2.10 Modèles d'intelligence artificielle à usage général ... 108
5.2.10.1 Remarques générales ... 108
5.2.10.2 Obligations pour tous les modèles d'intelligence artificielle à usage général ... 109

5.2.10.3 Obligations pour les modèles d'intelligence artificielle à usage général présentant un risque systémique ... 110
5.2.11 Les normes harmonisées et leur rôle dans le règlement sur l'intelligence artificielle ... 111
5.2.11.1 Les normes harmonisées ... 111
5.2.11.2 Le système de normalisation de l'UE ... 112
5.2.11.3 La présomption de conformité ... 113
5.2.11.4 Les travaux de normalisation dans le domaine de l'intelligence artificielle ... 113
5.2.12 Mesures de soutien à l'innovation ... 114
5.2.13 Gouvernance ... 115
5.2.14 Surveillance et contrôle de l'application ... 117
5.2.15 Droits individuels ... 118
5.2.16 Sanctions ... 119
5.2.17 Entrée en vigueur et application ... 119
5.3 Appréciation ... 120
5.3.1 Effets juridiques sur les opérateurs suisses ... 120
5.3.1.1 Opérateurs concernés ... 120
5.3.1.2 Effets ... 121
5.3.2 Relations avec l'Accord entre la Suisse et l'UE relatif à la reconnaissance mutuelle en matière d'évaluation de la conformité ... 122
5.3.2.1 Fonctionnement de l'accord ... 122
5.3.2.2 Impact du règlement sur l'IA sur les opérateurs suisses concernés par l'accord ... 123
5.3.2.3 Extension possible de l'ARM ... 124
5.3.3 Relations avec la décision d'adéquation de la Commission européenne en matière de protection des données ... 125
5.3.4 Relation avec la convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit ... 126
5.3.5 Autres éléments choisis ... 126
5.4 Conclusions intermédiaires ... 129
6 Autres domaines du droit spécifiques ... 130
6.1 Introduction ... 130
6.2 Propriété intellectuelle ... 130
6.2.1 Droit d'auteur et intelligence artificielle ... 130
6.2.1.1 Généralités ... 130
6.2.1.2 Problématiques et nécessité de légiférer du fait de l'intelligence artificielle 131
6.2.2 Intelligence artificielle et droit des brevets ... 135
6.2.2.1 Généralités ... 135
6.2.2.2 Défis ... 135
6.2.2.3 Besoin de légiférer ... 136
6.3 Responsabilité extracontractuelle ... 137
6.3.1 Droit de la responsabilité civile extracontractuelle ... 137
6.3.1.1 Proposition de directive européenne sur la responsabilité en matière d'intelligence artificielle ... 137

6.3.1.2 Droit suisse ... 140
6.3.1.3 Appréciation ... 141
6.3.2 Droit de la responsabilité du fait des produits ... 142
6.3.2.1 Directive européenne actualisée sur la responsabilité du fait des produits défectueux ... 143
6.3.2.2 Droit suisse ... 144
6.3.2.1 Appréciation ... 144
6.3.3 Protection de la personnalité ... 145
6.3.4 Conclusion ... 145
6.4 Droit général des contrats ... 146
6.4.1 Loi type de la CNUDCI sur les contrats automatisés ... 146
6.4.2 Droit suisse ... 146
6.4.2.1 Imputation de la manifestation de volonté et responsabilité contractuelle ... 146
6.4.2.2 Smart contracts ... 147
6.4.3 Appréciation ... 148
6.5 Droit du travail ... 148
6.5.1 Introduction ... 148
6.5.2 Au niveau européen ... 149
6.5.2.1 Directive visant à améliorer les conditions de travail des travailleurs des plateformes ... 149
6.5.2.2 Règlement de l'UE sur l'intelligence artificielle et droit du travail ... 151
6.5.3 Situation et discussions en droit suisse du travail ... 152
6.5.3.1 Motion 23.4492 Gysi « Intelligence artificielle. Renforcer les droits de participation des travailleurs » ... 152
6.5.3.2 Utilisation de l'intelligence artificielle dans le monde du travail en Suisse ... 152
6.5.3.3 Défis posés par l'intelligence artificielle en droit du travail en Suisse ... 153
6.5.3.4 Règles applicables en droit suisse ... 154
6.5.3.5 Appréciation ... 157
6.6 Droit pénal ... 160
6.6.1 Applicabilité de principe ... 160
6.6.2 Responsabilité pénale ... 161
6.6.2.1 Remarques générales ... 161
6.6.2.2 Exemple : utilisation de systèmes d'intelligence artificielle dans la conduite automatisée ... 162
6.6.3 Difficultés d'application du droit ... 164
6.6.4 Résumé et perspectives ... 165
7 Conclusions ... 169
Table des abréviations ... 171
Annexe 1 ... 177

1 Introduction

Par décision du 22 novembre 2023, le Conseil fédéral a chargé le DETEC (OFCOM) et le DFAE (Division Europe) de réaliser d'ici fin 2024 un état des lieux sur les approches de régulation en matière d'intelligence artificielle (IA) en Suisse. Un nœud central a été constitué afin de diriger les travaux, composé de représentants du DETEC (OFCOM), DFJP (OFJ) et DFAE (Division Europe, DDIP). Il a été décidé que le DFJP (OFJ) dirige le projet d'analyse juridique de base.

Actuellement, la Suisse ne connaît pas de législation transversale sur l'IA. Cela étant, comme relevé dans le rapport de 2019 du groupe de travail interdépartemental « Intelligence artificielle », l'IA ne se développe pas dans un vide juridique.[1] Le cadre légal en vigueur lui est pleinement applicable. Il s'agit notamment de la Constitution, de la LPD, de la LEg, des règles de responsabilité civile et pénale, etc.

Depuis le rapport de 2019, l'IA a connu des développements technologiques importants et le cadre réglementaire international a aussi évolué. La présente analyse juridique a pour but de faire un point de la situation et d'examiner quels sont les éventuels besoins d'adaptation en droit suisse. La structure et la méthodologie sont présentées ci-après (cf. ch. 2).

L'analyse juridique sert de base – aux côtés de l'analyse des réglementations en matière d'IA dans différents pays et régions du monde, et de celle sur les activités de régulation en lien avec l'IA dans les secteurs, établies par l'OFCOM – pour l'élaboration de l'état des lieux commandé par le Conseil fédéral. Suite à cet état des lieux, le Conseil fédéral pourra se déterminer sur l'octroi d'un mandat concret pour un éventuel projet de réglementation de l'IA en 2025 et régler les compétences au sein de l'administration.

2 But et méthodologie de l'analyse juridique de base

Le but de la présente analyse juridique de base est d'examiner en quoi les systèmes d'IA posent de nouveaux défis et d'évaluer si l'ordre juridique suisse en vigueur présente des lacunes qu'il convient de combler pour faire face à ceux-ci. Compte tenu du court délai à disposition pour la réalisation de cette analyse, celle-ci se concentre sur les principaux enjeux juridiques. Par ailleurs, les besoins de légiférer n'ont pas toujours pu être identifiés avec certitude et devront être approfondis le cas échéant. Enfin, l'analyse porte sur le droit international et fédéral, à l'exclusion du droit cantonal et communal (sur ce point, cf. ch. 4.1.3). Elle a été réalisée par l'OFJ, en collaboration avec la DDIP (pour les ch. 3 et 5.2.11), l'OFCOM (pour les chapitres 5.2 et 5.3.2) et l'IPI (pour le ch. 6.2).

Pour réaliser ce travail, il a été choisi de prendre comme point de départ la convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et

[1] Défis de l'intelligence artificielle, Rapport du groupe de travail interdépartemental « Intelligence artificielle » au Conseil fédéral, 2019, 96, disponible sous www.sbfi.admin.ch > Politique FRI > Politique fédérale pour la formation, la recherche et l'innovation 2025-2028 > Thèmes transversaux dans le domaine FRI > Numérisation dans le domaine FRI > Intelligence artificielle.

l'État de droit (ci-après : convention sur l'IA)[2]. Le mandat du Conseil fédéral de novembre 2023 prévoit en effet que les éventuelles approches de régulation doivent tenir compte de cette convention. Cette dernière a été négociée au sein du Comité sur l'intelligence artificielle du Conseil de l'Europe (CAI) et a été formellement adoptée le 17 mai 2024 par le Comité des Ministres du Conseil de l'Europe (cf. ch. 4). La Suisse est membre du Conseil de l'Europe et les travaux du CAI ont été présidés par un Suisse. En outre, une délégation suisse a aussi participé activement aux travaux.

En cas de décision politique de ratifier cette convention, le texte serait contraignant pour la Suisse. Dans la mesure où il est formulé en termes très généraux, il devrait être transposé (cf. ch. 4.1.2). L'analyse vise à esquisser les éventuels besoins d'adaptation du droit suisse en cas de ratification. Il est donc prévu de présenter le contenu des obligations découlant de la convention, puis d'examiner l'état actuel du droit suisse. L'analyse indique le cas échéant si le niveau de protection offert par ce dernier est suffisant. Compte tenu notamment du délai à disposition pour la rédaction de la présente analyse, cet examen n'est pas exhaustif et dans bien des cas des approfondissements demeurent nécessaires afin de clarifier toutes les questions juridiques qui se posent.

La convention sur l'IA, en tant que premier traité international contraignant dédié à l'IA, pose un cadre juridique général et appréhende les principaux enjeux juridiques dans ce domaine en matière de protection des droits de l'homme, de la démocratie et de l'État de droit. Elle aborde notamment les enjeux de protection des droits fondamentaux, de transparence, de protection des données, d'égalité et de non-discrimination, de responsabilité et de procédure. L'examen de ce texte permet ainsi également d'identifier les éventuelles lacunes du droit suisse d'un point de vue général, compte tenu de ces problématiques.

Dans un deuxième temps, l'analyse se penche sur le règlement de l'UE établissant des règles harmonisées concernant l'intelligence artificielle[3] (ci-après : règlement sur l'IA) (cf. ch. 5). Ce texte n'est pas contraignant pour la Suisse. Il n'y a donc en l'état pas d'obligation juridique de le reprendre en droit interne. Cependant, au vu de la position géographique de la Suisse et des liens économiques avec l'UE, on ne peut s'économiser l'examen de ce règlement. Par ailleurs, il s'applique aux fournisseurs suisses qui, dans l'UE, mettent sur le marché ou mettent en service des systèmes d'IA ou qui mettent sur le marché des modèles d'IA à usage général. Il s'applique également aux fournisseurs et déployeurs suisses de systèmes d'IA dès que les résultats générés sont utilisés dans l'UE. L'examen de cette règlementation découle en outre du mandat du Conseil fédéral du 22 novembre 2023.

---

[2] Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, STCE 225, disponible sous www.coe.int > Droits humains > Intelligence artificielle et droits humains > Convention-cadre (consulté le 26 août 2024).

[3] Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 ; JO L, 2024/1689, 12 juillet 2024, disponible sous https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AL_202401689 (consulté le 26 août 2024).

Ensuite, l’analyse juridique présente une appréciation du règlement sur l’IA en abordant des questions pertinentes du point de vue du droit suisse. L’analyse expose également le rapport et les synergies entre le règlement sur l’IA et la convention sur l’IA. En particulier, il convient d’examiner l’approche suivie par le règlement sur l’IA et de voir en quoi ce dernier va plus loin ou diffère de la convention sur l’IA.

Finalement, l’analyse appréhende l’état du droit suisse par rapport aux enjeux de l’IA dans des domaines du droit qui ne sont pas exhaustivement appréhendés par la convention sur l’IA ou le règlement sur l’IA, mais pour lesquels l’IA pose des défis (cf. ch. 6). Il s’agit en particulier d’aspects relevant du droit de la propriété intellectuelle, et de la responsabilité civile et pénale.

À noter que les questions sectorielles, comme les projets de régulation dans les domaines de l’énergie et de la santé, sont traitées dans le cadre d’une analyse séparée sous la direction de l’OFCOM, sous réserve de certains aspects qui illustrent des enjeux transversaux, à savoir notamment la protection contre les discriminations, la protection de la sphère privée ou les questions de responsabilité.

L’analyse juridique est réalisée en tenant compte des développements intervenus jusqu’au 31 août 2024.

3 Instruments au niveau international[4]

Sur le plan international, outre la convention sur l'IA et le règlement sur l'IA déjà mentionnés ci-dessus et qui feront l'objet d'un examen séparé, plusieurs autres instruments, juridiquement contraignants (cf. ch. 3.1) ou non (cf. ch. 3.2) s'appliquent au domaine de l'IA.

3.1 Traités internationaux

L'IA n'évolue pas dans un vide juridique au plan international. Au contraire, le cadre juridique existant s'applique également en matière d'IA. Ainsi, de très nombreux traités, coutumes et principes généraux du droit international lient la Suisse dans ce domaine.

À titre d'exemple, on peut mentionner la CEDH ou la Convention 108+[5] du Conseil de l'Europe. Comme beaucoup d'autres traités internationaux, ces conventions sont applicables à l'IA, même si elles ne s'y réfèrent pas de manière explicite. Le caractère général et flexible de ces normes permet en effet de s'adapter aux développements technologiques.

Ainsi, dans son arrêt Glukhin c. Russie du 4 juillet 2023[6], la CourEDH a par exemple considéré que le recours à la technologie de reconnaissance faciale était susceptible de violer les art. 8 (droit au respect de la vie privée) et 10 CEDH (liberté d'expression). Cet exemple montre que le droit existant permet déjà d'appréhender certaines situations dans lesquelles un système d'IA est utilisé de manière contraire au droit (pour d'autres considérations sur cet arrêt, cf. ch. 4.3.1.2).

La CourEDH reconnaît par ailleurs dans certains cas l'existence d'obligations positives, à la charge des États, de créer les conditions permettant la protection et l'exercice des droits garantis par la CEDH. Ainsi l'interprétation du droit international existant pourrait déjà, à certaines conditions, donner lieu à une obligation positive d'adopter des normes de protection dans le domaine de l'IA.[7] En effet, les normes internationales de protection des droits de l'homme n'imposent pas uniquement une obligation de s'abstenir de restreindre les droits des individus ; elles exigent également de l'État qu'il garantisse la protection des individus contre les actes d'autrui et qu'il prenne des mesures positives pour sauvegarder les droits de l'homme. Dans ces conditions, une obligation positive de légiférer dans le domaine de l'IA pourrait être déduite du droit international existant.

---

[4] Ce chapitre a été rédigé sur la base de contributions de la DDIP.

[5] La Convention 108 a été modernisée par le protocole d'amendement à la Convention pour la protection des personnes à l'égard du traitement des données à caractère personnel, adopté par le Comité des Ministres lors de sa 128e session à Elseneur, le 18 mai 2018. Le protocole a été ratifié par la Suisse le 7 septembre 2023 mais il n'est pas encore entré en vigueur.

[6] CourEDH, *Glukhin c. Russie*, 11519/20 (4 juillet 2023).

[7] WOLFGANG HOFFMANN-RIEM, Artificial Intelligence as a Challenge for Law and Regulation, in : Thomas Wischmeyer/Timo Rademacher (éds.), Regulating Artificial Intelligence, Cham 2020, 1 ss, 8.

3.2 Instruments non-contraignants dans le domaine de l'IA

En sus du cadre juridique évoqué ci-dessus, plusieurs instruments internationaux non-contraignants (*soft law*) pertinents pour la Suisse existent dans le domaine de l'IA.

Il s'agit notamment des textes suivants : la Recommandation sur l'intelligence artificielle adoptée par l'OCDE le 22 mai 2019, révisée le 8 novembre 2023 pour mettre à jour la définition d'un « système d'IA » ; la Déclaration du Comité des Ministres du Conseil de l'Europe sur les capacités de manipulation des processus algorithmiques, Decl(13/02/2019)1 ; la Recommandation du Comité des Ministres du Conseil de l'Europe aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, CM/Rec(2020)1 ; la Recommandation sur l'éthique de l'intelligence artificielle de l'UNESCO, adoptée le 23 novembre 2021 ; la Déclaration des chefs d'État et de gouvernement adoptée lors du 4ème Sommet du Conseil de l'Europe à Reykjavik les 16 et 17 mai 2023 ; la Déclaration de Bletchley par les pays participants au Sommet sur la sécurité de l'intelligence artificielle les 1 et 2 novembre 2023 ; la résolution de l'Assemblée générale de l'ONU du 21 mars 2024 sur l'intelligence artificielle.

D'autres instruments non-contraignants sont par ailleurs en cours d'élaboration. Une Recommandation sur l'impact des systèmes d'IA, leur potentiel de promotion de l'égalité, y compris l'égalité de genre et les risque qu'ils peuvent entraîner en matière de non-discrimination sera notamment rédigée par un comité du Conseil de l'Europe d'ici fin 2025. Des lignes directrices sur les implications de l'IA générative pour la liberté d'expression seront finalisées au sein d'un autre comité du Conseil de l'Europe également en 2025. La Suisse participe activement à ces deux comités.

Bien que non-contraignants, ces textes sont importants dans la mesure où ils reflètent un certain consensus au niveau international dans le domaine de l'IA. Depuis plusieurs années, la Suisse s'engage d'ailleurs activement dans ces différents processus internationaux sur l'IA. En outre, ces instruments servent de référence pour la Suisse. Ainsi, ils ont par exemple été mentionnés à plusieurs reprises dans le rapport de 2019 du groupe de travail interdépartemental intitulé « Défis de l'intelligence artificielle »[8].

Il convient pour finir de mentionner encore les Principes directeurs internationaux et le Code de conduite international du processus Hiroshima pour les organisations qui développent des systèmes d'IA avancés du 30 octobre 2023 ainsi que la Déclaration des dirigeants du G7 sur le processus d'IA d'Hiroshima du 30 octobre 2023. La Suisse n'a pas participé à ces initiatives.

[8] Rapport défis de l'intelligence artificielle (n. 1).

4 Convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit du Conseil de l'Europe

4.1 Remarques introductives

4.1.1 Contexte

Le CAI a été chargé en juin 2022 par le Comité des Ministres du Conseil de l'Europe d'élaborer un instrument juridique contraignant à caractère transversal sur l'IA, fondé sur les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit.

Le CAI est composé de représentants de tous les États membres du Conseil de l'Europe, ainsi que de l'UE. Plusieurs États observateurs participent aussi aux travaux, à savoir l'Argentine, l'Australie, le Canada, le Costa Rica, le Saint-Siège, Israël, le Japon, le Mexique, le Pérou, les États-Unis d'Amérique et l'Uruguay. Des représentants de la société civile prennent également part à ce processus.

Durant les négociations de la convention, le comité était présidé par un Suisse. La Suisse était par ailleurs représentée au sein du CAI par une délégation composée de l'OFCOM, de la DDIP et de l'OFJ.

Les travaux de négociation de la convention sur l'IA se sont terminés en mars 2024. L'adoption du texte par le Comité des Ministres a eu lieu en mai 2024. La convention sera ouverte à la signature en septembre 2024. La tâche principale du CAI est terminée, mais le comité poursuit encore ses travaux en lien avec des aspects connexes, notamment l'élaboration d'une méthodologie facultative servant d'instrument de gestion des risques et des impacts des systèmes d'IA.

4.1.2 Applicabilité et justiciabilité

Il se pose en premier lieu la question de savoir si la convention sur l'IA est d'applicabilité directe (ou justiciable ou *self-executing),* c'est-à-dire qu'elle fonde directement des droits et obligations pour les personnes physiques et morales, ou si elle s'adresse uniquement aux États, qui doivent la transposer pour lui donner effet.

Sont directement applicables les normes qui sont suffisamment concrètes et précises pour que des personnes physiques ou morales en retirent des droits et des obligations sur lesquels elles pourront fonder une action devant les autorités judiciaires et administratives. Corollairement, les autorités d'application du droit et les tribunaux pourront appliquer directement ces normes internationales.[9]

Le TF estime que le caractère directement applicable d'une disposition de droit international implique que celle-ci soit, considérée dans son contexte et à la lumière tant de l'objet que du

[9] La relation entre droit international et droit interne, Rapport du Conseil fédéral du 5 mars 2010 en réponse au postulat 07.3764 de la Commission des affaires juridiques du Conseil des Etats du 16 octobre 2007 et au postulat 08.3765 de la Commission des institutions politiques du Conseil national du 20 novembre 2008, FF **2010** 2067, 2089.

but du traité, inconditionnelle et suffisamment précise pour produire un effet direct et s'appliquer comme telle à un cas d'espèce ou constituer le fondement d'une décision concrète.[10] Le caractère directement applicable d'une norme est en fin de compte affaire d'interprétation de la disposition concernée par les autorités étatiques.

Le TF a produit une abondante jurisprudence sur la façon de déterminer l'application directe. Un fort besoin de protection de l'individu justifie plutôt l'application directe de la norme. Il en va inversement lorsque les répercussions sur l'État dans son ensemble sont importantes, que les circonstances sont complexes et difficiles à apprécier dans l'examen judiciaire du cas d'espèce ou que les conséquences financières sont lourdes. Le TF est également réservé lorsque l'application directe requiert une appréciation politique ou l'examen de questions de principe.[11]

Ne sont pas directement applicables (c'est-à-dire sont non justiciables ou *non-self executing*) les normes de nature programmatoire, soit celles qui s'adressent aux États. Elles doivent être concrétisées par les autorités étatiques concernées avant de fonder des droits et obligations pour les particuliers.[12]

L'examen des normes de la convention sur l'IA selon les critères ci-dessus, amène à conclure que d'une manière générale la convention s'adresse aux États et qu'elle n'est pas directement applicable. Cela découle aussi du texte lui-même, dont l'art. 1, par. 2, prévoit que « Chaque Partie adopte ou maintient les mesures législatives, administratives ou autres appropriées pour donner effet aux dispositions de la présente Convention ». Il n'est cependant pas exclu qu'un tribunal parvienne à une autre conclusion dans un cas d'espèce. L'évolution de la jurisprudence pourrait aussi jouer un rôle.

Un tribunal pourrait tout de même examiner cette convention dans le cadre d'un recours déposé par un particulier, qui se plaindrait de l'absence de mise en œuvre de ses obligations par l'État. Une autorité judiciaire pourrait ainsi conclure que le législateur doit agir sur la base des dispositions de la convention. Cependant, si la norme en cause impose uniquement une obligation d'agir, mais ne précise pas comment, le tribunal se limitera à une décision dite « incitative », pour demander à l'autorité législative d'agir.[13]

Quant à la question de savoir si la convention sur l'IA pourrait être invoquée devant la CourEDH, tel pourrait être le cas tout au plus en lien avec la violation d'un droit protégé par la CEDH. Ainsi, par exemple, la non-concrétisation d'une disposition de la convention sur l'IA demandant aux États Parties de prendre des mesures législatives ou autres dans un certain domaine pourrait conduire la CourEDH à constater la violation d'un droit garanti par la CEDH,

---

[10] La relation entre droit international et droit interne (n. 9), 2105 s.

[11] La relation entre droit international et droit interne (n. 9), 2105 s.

[12] La relation entre droit international et droit interne (n. 9), 2089.

[13] Voir ATF **137** I 305, consid. 6.6 s.

dans la mesure où l'omission d'agir s'analyserait en une violation des obligations positives de l'État, par exemple au titre de l'art. 8 CEDH. Cela ne concerne évidemment que les États Parties qui sont membres du Conseil de l'Europe. Cette situation pourrait conduire à un système à deux niveaux entre les États Parties qui sont membres du Conseil de l'Europe, et les États Parties qui ne le sont pas. Cependant, ce cas de figure est commun à toutes les conventions du Conseil de l'Europe ouvertes à la ratification aussi par des États non-membres et n'est donc pas nouveau.

4.1.3 Mise en œuvre en droit interne et fédéralisme

Selon l'art. 5, al. 4, Cst., la Confédération et les cantons doivent respecter le droit international.

Comme mentionné ci-dessus (cf. ch. 4.1.2), la convention sur l'IA s'adresse aux États et contient des règles et principes formulés en termes généraux. Ainsi, les obligations qu'elle prévoit doivent être mises en œuvre en droit interne.

Dans un État fédéral comme la Suisse, cette obligation de mise en œuvre peut incomber, selon l'obligation concernée, à la Confédération, aux cantons ou aux deux niveaux de l'État simultanément.[14] La mise en œuvre de la convention, si cette dernière venait à être ratifiée par la Suisse, imposerait d'examiner, pour chaque obligation, quelles sont les règles relatives à la répartition des compétences entre la Confédération et les cantons dans le domaine en question. Ainsi, en fonction du domaine du droit concerné par la mise en œuvre, il s'agira d'examiner si la Confédération dispose d'une compétence ou pas.

Dans le domaine de la protection des données par exemple, elle pourra s'appuyer sur les art. 95, al. 1, 97, al. 1, 122, al. 1, Cst. et sur sa compétence inhérente de régler l'organisation de ses autorités pour édicter des dispositions dans le secteur privé et pour les organes fédéraux. En revanche, sous réserve de compétences découlant de normes matérielles de protection des données contenues dans des lois fédérales spéciales[15], les traitements de données effectués par des organes cantonaux ou communaux relèvent du droit cantonal[16].

Toujours à titre d'exemple, en matière d'éducation, les compétences sont principalement cantonales (cf. art. 62 Cst.).

---

[14] JUDITH WYTTENBACH, Umsetzung von Menschenrechtsübereinkommen in Bundesstaaten, Zurich/St-Gall 2017, 299.

[15] Ainsi par exemple les art. 89*a* ss LCR, ou les art. 49*a* ss LAVS s'appliquent à tous les organes et personnes privées en charge de l'application des lois, y compris aux organes cantonaux.

[16] CR LPD-SYLVAIN MÉTILLE/LIVIO DI TRIA, art. 2 N 14 ; PHILIPPE MEIER, Protection des données – Fondements, principes généraux et droit privé, Berne 2010, N 356 ss.

La présente analyse se concentre sur la mise en œuvre potentielle de la convention sur l'IA au niveau de la Confédération. Comme indiqué ci-dessus, les cantons seraient toutefois responsables de la mise en œuvre dans leurs domaines de compétences.

4.2 Chapitre I : Dispositions générales

4.2.1 Article 1 – Objet et but

L'art. 1 règle l'objet et le but de la convention sur l'IA.

Il découle du par. 1 que la convention vise à garantir que les activités menées dans le cadre du cycle de vie des systèmes d'IA sont pleinement compatibles avec les droits de l'homme, la démocratie et l'État de droit. Cela s'inscrit dans le cadre du mandat du Conseil de l'Europe. La notion de « système d'IA » est définie à l'art. 2 (cf. ch. 4.2.2).

La convention vient compléter les mécanismes de protection des droits de l'homme en vigueur dans chaque État Partie, y compris les engagements nationaux et internationaux applicables. La convention ne créé toutefois pas de nouveaux droits fondamentaux. Elle ne réglemente pas non plus l'IA en tant que telle, mais seulement des aspects pertinents en matière de protection des droits de l'homme, de démocratie et d'État de droit.

Le par. 2 concerne la mise en œuvre de la convention. Il prévoit que « chaque Partie adopte ou maintient les mesures législatives, administratives ou autres appropriées pour donner effet aux dispositions de la présente Convention ». Cette formulation reflète le principe de droit international selon lequel les États sont libres quant à la manière dont ils souhaitent mettre en œuvre leurs obligations internationales. Ainsi, pour respecter les obligations prévues par la convention, les États disposent d'une grande liberté quant au choix des mesures : il s'agit d'une obligation de résultat et non de moyens. En théorie, l'État est donc libre de choisir si une disposition sera mise en œuvre, dans son ordre juridique interne, par le pouvoir législatif, exécutif ou judiciaire. L'art. 1, par. 2, prévoit en outre que ces mesures doivent être « graduées et différenciées, si nécessaire, en fonction de la gravité et de la probabilité de l'apparition d'impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit tout au long du cycle de vie des systèmes d'intelligence artificielle ». Cela signifie que les mesures doivent être adaptées au niveau de risque posé par un système d'IA.

En droit suisse, la question de savoir à quel niveau normatif en droit interne devrait être mise en œuvre la convention doit être examinée à l'aune des art. 5, al. 1, et 164 Cst. Cette dernière disposition prévoit que toutes les dispositions importantes qui fixent des règles de droit doivent être édictées sous la forme d'une loi fédérale. Appartiennent en particulier à cette catégorie les dispositions fondamentales relatives : a) à l'exercice des droits politiques ; b) à la restriction des droits constitutionnels; c) aux droits et aux obligations des personnes ; d) à la qualité de contribuable, à l'objet des impôts et au calcul du montant des impôts ; e) aux tâches et aux prestations de la Confédération ; f) aux obligations des cantons lors de la mise en œuvre et de l'exécution du droit fédéral ; g) à l'organisation et à la procédure des autorités fédérales.

L'art. 164 Cst. impose de prendre des mesures législatives dans certains domaines lorsque l'importance de la matière à régler ne permet pas l'adoption d'une autre forme de règles (or-

donnance, arrêté, etc.). Il s'agit du principe de la « réserve générale de la loi au sens formel »[17], selon lequel les actions importantes de l'État requièrent une base légale au sens formel. Ce principe permet d'assurer que les « grand arbitrages de valeurs et d'intérêts et les grands choix de politique et de technique juridiques »[18] soient adoptés par l'autorité législative, garantissant la représentativité et la participation politiques.

Ainsi, en droit interne, le cadre constitutionnel suisse devra être respecté s'agissant de la forme des actes à adopter pour mettre en œuvre la convention. Les principes et règles de base revêtant une grande importance devraient être concrétisés dans une loi au sens formel.

4.2.2 Définitions

4.2.2.1 Article 2 – Définition des systèmes d'intelligence artificielle

Selon la convention, un système d'IA est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influer sur des environnements physiques ou virtuels. Différents systèmes d'intelligence artificielle présentent des degrés variables d'autonomie et d'adaptabilité après leur déploiement.

Cette définition s'inspire de celle adoptée par l'OCDE le 8 novembre 2023[19], qui tient compte de la nécessité de renforcer la coopération internationale autour du thème de l'IA et de faciliter les efforts visant à harmoniser sa gouvernance au niveau mondial, y compris en harmonisant la terminologie pertinente.

La définition reflète une compréhension large de ce que sont les systèmes d'IA, notamment par opposition à d'autres types de systèmes de logiciel plus simples, basés sur des règles définies uniquement par des personnes physiques pour exécuter automatiquement des opérations. Elle est suffisamment abstraite et souple pour résister aux évolutions technologiques futures. La définition est également fonctionnelle en ce sens qu'elle a été rédigée spécifiquement pour les buts de la convention sur l'IA et qu'elle n'est pas censée donner une signification universelle du terme.

---

[17] BSK BV-JUDITH WYTTENBACH/KARL-MARC WYSS, art. 164 N 4 ; SGK BV-PIERRE TSCHANNEN, art. 164 N 4 ; CR Cst.-JACQUES DUBEY, art. 164 N 10.

[18] CR Cst.-JACQUES DUBEY, art. 164 N 11.

[19] OCDE, Recommandation du Conseil sur l'intelligence artificielle, 2024, OECD/LEGAL/0449, Chapitre I, disponible sous : https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449 (consulté le 26 août 2024).

Selon le rapport explicatif de la convention, les Parties peuvent préciser cette définition dans leur système juridique national pour plus de sécurité juridique, sans pour autant en limiter la portée.[20]

La définition correspond dans les grandes lignes à celle prévue par le règlement sur l'IA de l'UE (cf. ch. 5.2.4.1).

Le rapport explicatif de la convention sur l'IA donne peu d'éléments permettant d'interpréter la définition. Il renvoie à l'exposé des motifs accompagnant la définition actualisée de système d'IA de l'OCDE.[21] On peut dès lors se référer à ce dernier document pour plus de détails. Les éléments principaux à retenir sont les suivants :

- Objectifs « explicites ou implicites » du système d'IA

  Les objectifs du système d'IA peuvent être « explicites ou implicites ». Ceux-ci sont explicites lorsqu'un être humain les inscrit directement dans le système. Les objectifs sont implicites lorsqu'ils sont déduits d'un ensemble de règles spécifiées par un humain (p. ex. un système de conduite automatisé est programmé pour respecter le code de la route, mais implicitement il protège des vies), ou lorsqu'ils découlent du système lui-même, qui apprend alors de nouveau objectifs (p. ex. les systèmes d'IA générative comme ChatGPT).

- Le système d'IA « déduit comment générer des résultats », « à partir d'entrées reçues »

  Le système d'IA est alimenté avec des entrées, à savoir des données et des règles de programmation, qui peuvent être fournies par des êtres humains ou des machines. Il traite ces données à l'aide d'un modèle algorithmique capable de calculer un résultat.

  Par exemple, un système de reconnaissance visuelle d'objets déduit comment générer son résultat (c'est-à-dire l'identification de l'objet dans l'image), à partir d'une entrée constituée par les pixels de l'image.

- Le système produit des « résultats » susceptibles d'influencer des environnements physiques ou virtuels

  Les différentes fonctions exécutées par les systèmes d'IA se traduisent en des résultats. Il existe des grandes catégories de résultats, comme les recommandations, les prédictions et les décisions. Par exemple, si un système de conduite automatisé prédit

[20] Rapport explicatif de la convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, STCE 225, N 27 (disponible sous www.coe.int > Droits humains > Intelligence artificielle et droits humains > Convention-cadre).

[21] Rapport explicatif convention sur l'IA (n. 20), N 24 ; OCDE, Explanatory memorandum on the update OECD definition of an AI system. OECD Artificial intelligence papers, No. 8, OECD Publishing, mars 2024, disponible sous https://doi.org/10.1787/623da898-en (consulté le 26 août 2024).

qu'une zone de pixels dans l'entrée de sa caméra est un piéton, il peut recommander de freiner ou décider de freiner.

La définition mentionne explicitement les « contenus » en tant que catégorie de résultat. Ce terme se réfère en particulier aux systèmes d'IA dite générative, comme ChatGPT, qui produisent des contenus tels que des textes, des images, du son ou des vidéos.

Les environnements influencés par les systèmes d'IA peuvent être physiques ou virtuels. Ainsi, les résultats sont perçus soit par des humains, soit par des composants de machines ou d'appareils.

- Divers degrés d'autonomie et d'adaptabilité après déploiement

  La référence à l'autonomie se rapporte au degré de capacité d'un système d'IA à apprendre ou agir sans intervention humaine. Certains systèmes d'IA peuvent générer des résultats sans que ceux-ci ne soient expressément prévus dans les buts programmés des systèmes et sans instruction spécifique de la part d'un être humain.

  La mention de l'adaptabilité tient compte du fait que certains systèmes d'IA peuvent continuer à évoluer après leur conception et leur déploiement. Il peut s'agir par exemple d'un système de reconnaissance vocale qui s'adapte à la voix d'un individu ou d'un système de recommandation musicale personnalisé. Les systèmes d'IA peuvent être entraînés une seule fois, périodiquement ou continuellement et fonctionnent en déduisant des modèles et des relations dans les données. Grâce à cet entraînement, certains systèmes d'IA peuvent acquérir la capacité d'effectuer de nouvelles formes de déduction qui n'avaient pas été envisagées au départ par les programmateurs.

En l'état actuel, le droit suisse ne connaît pas de définition des systèmes d'IA. Cela étant, dans la mesure où le législateur suisse privilégie une approche technologiquement neutre, le cadre juridique actuel appréhende également les états de fait impliquant des systèmes d'IA, indépendamment de l'existence d'une définition de ces systèmes en droit suisse.

Il convient également de relever que le CNAI, rattaché à l'OFS, met à disposition une terminologie harmonisée en matière d'IA permettant de partager un langage et une compréhension communs dans l'administration fédérale dans ce domaine. La terminologie inclut une définition de ce qu'est un système d'IA.[22] Cette définition est proche de la définition proposée à l'art. 2 de la convention et de celle de l'OCDE.

[22] Cf. www.cnai.swiss > Services > Terminologie (état au 21 décembre 2023), 7, consulté le 26 août 2024 : « Un système IA (AI system) est un système automatique capable d'inférer, sur la base des «inputs» (entrées) qu'il reçoit et pour des objectifs explicites ou implicites, comment générer des «outputs» (sorties) tels que des prévisions, des contenus, des recommandations ou des décisions, et qui, ce faisant, peut exercer une influence sur des environnements physiques ou virtuels. Les systèmes IA peuvent être dotés d'une autonomie plus ou moins grande. »

Néanmoins, certaines obligations dérivant de la convention sur l'IA sont spécifiques aux systèmes d'IA, comme l'obligation de mettre en place un cadre de gestion des risques et des impacts (cf. ch. 4.3.4). Dans l'hypothèse d'une ratification de la convention sur l'IA, il paraît dès lors souhaitable qu'une définition des systèmes d'IA soit introduite en droit suisse également, afin de cibler le champ d'application de certaines dispositions de la convention.

4.2.2.2 Cycle de vie

Tout au long du texte de la convention, il est fait référence aux activités menées dans le cadre du « cycle de vie » des systèmes d'IA. Cette référence souligne le fait que les Parties doivent mettre en œuvre les dispositions de la convention en tenant compte de toutes les étapes des activités liées aux systèmes d'IA, allant de leur conception à leur mise hors service, quel que soit l'acteur concerné.[23]

En effet, bien que la phase d'utilisation des systèmes d'IA soit souvent vue comme la plus risquée, des atteintes aux droits de l'homme, à la démocratie et à l'État de droit peuvent également se produire à d'autres étapes. Ainsi, par exemple, l'art. 10 de la convention sur l'IA vise à garantir le respect de l'égalité et de la non-discrimination dans toutes les phases du cycle de vie d'un système d'IA, à savoir notamment non seulement lors du déploiement d'un système d'IA, mais aussi dans la phase de développement, s'agissant par exemple des exigences quant au jeu de données utilisées pour entraîner le système.

Le texte ne contient pas de définition du « cycle de vie », mais le rapport explicatif apporte des éclaircissements. Sans fournir une liste exhaustive des phases du cycle de vie d'un système d'IA, il donne un aperçu des phases possibles, en se référant aux travaux de l'OCDE: 1) planification et conception, 2) collecte et traitement des données, 3) développement de systèmes d'IA, y compris l'élaboration de modèles et/ou l'adaptation de modèles existants à des tâches spécifiques, 4) essais, vérification et validation, 5) fourniture/mise à disposition des systèmes, 6) déploiement, 7) exploitation et suivi, et 8) mise hors service. Ces activités se déroulent souvent de manière itérative, mais ne sont pas nécessairement séquentielles.[24]

[23] Rapport explicatif convention sur l'IA (n. 20), N 15.

[24] Rapport explicatif convention sur l'IA (n. 20), N 15.

4.2.3 Article 3 – Champ d'application

4.2.3.1 Secteur public et secteur privé

4.2.3.1.1 Généralités

La convention sur l'IA a un champ d'application large, afin de couvrir les activités menées dans le cadre du cycle des vie des systèmes d'IA qui sont susceptibles d'interférer avec les droits de l'homme, la démocratie et l'État de droit.[25]

Elle s'applique dans le secteur public et dans le secteur privé, mais avec des nuances. Dans les deux cas, le destinataire des obligations prévues par la convention est l'État et dans les deux cas, ce dernier dispose d'une marge de manœuvre dans la manière de mettre en œuvre ses obligations. Les négociateurs ont en effet souhaité préserver la diversité des différents systèmes juridiques, des traditions et pratiques, ainsi que des multiples contextes d'utilisation des systèmes d'IA, tant dans le secteur public que dans le secteur privé.[26]

Dans le secteur public, les États sont tenus d'appliquer la convention aux activités menées dans le cadre du cycle de vie des systèmes d'IA. Sont également couvertes ici les activités des acteurs privés agissant pour le compte des pouvoirs publics (art. 3, par. 1, let. a).[27] Des systèmes d'IA sont utilisés dans le secteur public par exemple pour l'octroi de prestations sociales, dans le domaine fiscal, ou dans le cadre de procédures d'asile.[28] En pratique, après un examen des risques et impacts des activités en question, les États Parties doivent décider de la manière dont ils souhaitent donner effet aux dispositions de la convention. Ils peuvent conserver leur législation, l'adapter, ou adopter de nouvelles règles. Les mesures à prendre peuvent selon les cas aussi consister en des mesures administratives ou des circulaires, la jurisprudence des tribunaux ou des mesures non-contraignantes, conformément à la graduation prévue à l'art. 1, par. 2.[29] Cependant, comme mentionné plus haut (cf. ch. 4.2.1), en Suisse le cadre constitutionnel doit en tout état être respecté s'agissant de la forme des actes à adopter.

Pour le secteur privé, il a fallu trouver un compromis durant les négociations entre les États en faveur de son inclusion dans le champ d'application de la convention et ceux qui y étaient

---

[25] Rapport explicatif convention sur l'IA (n. 20), N 26.

[26] Rapport explicatif convention sur l'IA (n. 20), N 16.

[27] Il s'agit notamment des activités des acteurs privés opérant dans le cadre d'un contrat avec une autorité publique ou une autre fourniture privée de service public, ainsi que les marchés et contrats publics, cf. Rapport explicatif convention sur l'IA (n. 20), N 28.

[28] Cf. NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung : Entwicklungen und Herausforderungen, Jusletter IT 4 juillet 2024, 5 N 9.

[29] Rapport explicatif convention sur l'IA (n. 20), N 18 ss.

opposés. Le texte prévoit finalement que les Parties doivent répondre aux risques et aux impacts découlant des activités menées dans le cadre du cycle de vie de système d'IA utilisés dans le secteur privé conformément à l'objet et au but de la convention. La référence à l'objet et au but de la convention a pour effet d'importer tous les concepts de l'art. 1. Cela implique que l'État doit également évaluer les risques et impacts des activités d'IA dans le secteur privé sur les droits de l'homme, la démocratie et l'État de droit et se déterminer ensuite sur la manière dont il souhaite y répondre.[30] Il peut soit décider d'appliquer les principes et obligations contenus aux Chapitre II à VI de la convention, soit prendre d'autres mesures.

La distinction entre les deux secteurs n'est pas dans le résultat à atteindre – il s'agira dans les deux cas de répondre aux risques et impacts liés aux systèmes d'IA pour les droits de l'homme, la démocratie et l'État de droit – mais dans la manière d'y parvenir. Il y a plus de liberté pour le secteur privé.

Selon le rapport explicatif, les « autres mesures », mentionnées à l'art. 3, par. 1, let. b, en référence au secteur privé, peuvent être des mesures administratives.[31] Sont visées des directives ou des circulaires. Cependant, on peut douter que le recours à de tels actes soit suffisant en Suisse. En effet, imposer des droits et obligations en matière d'IA dans le secteur privé requiert en principe une base légale formelle dans la mesure où cela touche les droits fondamentaux des personnes concernées (liberté économique notamment). Toujours selon le rapport explicatif, les mesures peuvent aussi être volontaires.[32] Il s'agit selon toute vraisemblance de mesures d'autorégulation. Il paraît cependant peu probable que les objectifs de la convention puissent être atteints sans mesures étatiques en Suisse, qui pourraient également consister en l'obligation à charge des acteurs privés de mettre en place des mécanismes d'autorégulation, ou en leur soutien de la part de l'État.

Dans tous les cas, l'État devra s'assurer que le système choisi permet d'appréhender les risques et impacts de l'IA dans le secteur privé de manière conforme à l'objet et au but de la convention. Par ailleurs, les Parties ne peuvent aller en-deçà des obligations internationales qui leur incombent (art. 3, par. 1, let. b *in fine*).

Les États doivent annoncer leur choix – modifiable en tout temps – à la Secrétaire générale ou au Secrétaire général lors de la ratification, l'acceptation, l'approbation ou l'adhésion à la Convention (art. 3, par. 1, let. b). Dans un délai de deux ans à compter de la date à laquelle il devient Partie, puis de manière périodique par la suite, chaque État doit fournir un rapport contenant les détails des activités qu'elle a entreprises pour donner effet à l'art. 3, par. 1, let. a et b, à l'attention de la Conférence des Parties (cf. art. 24 de la convention, ch. 4.4.2).

---

30 Rapport explicatif convention sur l'IA (n. 20), N 29.

31 Rapport explicatif convention sur l'IA (n. 20), N 29.

32 Rapport explicatif convention sur l'IA (n. 20), N 29.

4.2.3.1.2 Signification pour la Suisse

Concernant le secteur privé

Comme mentionné ci-dessus (cf. ch. 4.2.3.1.1), on peut douter que des mesures purement administratives ou volontaires soient suffisantes en droit suisse pour répondre aux risques et impacts découlant des activités menées dans le cadre du cycle de vie des systèmes d'IA dans le secteur privé d'une manière conforme à l'objet et au but de la convention. En conséquence, en l'absence d'autres mesures appropriées entrant en ligne de compte, l'analyse part de l'hypothèse que le législateur suisse mettra en œuvre l'art. 3, par. 1, let. b, de la convention en appliquant les dispositions de la convention au secteur privé également.

Cependant, il y a plus de liberté pour réglementer le secteur privé. En effet, des différences entre secteur public et secteur privé qui tiennent compte des particularités du droit suisse sont à envisager, notamment compte tenu du fait que les exigences légales sont en principe plus élevées dans le secteur public (cf. ci-dessous), et eu égard à l'effet en principe uniquement vertical des droits fondamentaux.

En droit suisse, cela signifie en pratique que les dispositions et principes de la convention devraient s'appliquer dans le secteur privé là où un effet horizontal direct ou indirect des droits fondamentaux est reconnu ou devait l'être à l'avenir. En effet, la convention n'a pas pour effet d'étendre la portée des droits fondamentaux entre privés, au-delà de ce que prévoit l'art. 35, al. 1 et 3, Cst. (cf. ch. 4.3.1.1). La reconnaissance d'un effet horizontal direct ou indirect des droits fondamentaux entre privés est donc une condition pour que la convention trouve application en Suisse dans les relations de droit privé.

L'impact de la convention dans les relations entre privés pourrait cependant être appelé à évoluer, notamment en raison des engagements internationaux pris par la Suisse (p. ex. convention internationale sur l'élimination de toutes les formes de discrimination raciale[33], convention internationale sur l'élimination de toutes les formes de discrimination à l'égard des femmes[34] ou encore convention relative aux droits des personnes handicapées[35]). L'évolution de la jurisprudence pourrait aussi jouer un rôle, y compris celle de la CourEDH s'agissant de l'interprétation de la CEDH et des obligations positives qui en découlent.

Il convient ici de préciser que certaines relations entre privés pourraient ne pas du tout être concernées par la mise en œuvre de la convention sur l'IA. Tel est le cas lorsqu'il n'existe pas de rattachement possible à l'un ou l'autre droit fondamental. L'on peut songer à la garantie classique du bon fonctionnement d'un appareil (p. ex. robot-tondeuse qui ne tond que la moitié du jardin, frigo intelligent qui adapte mal la température) ou la garantie pour les défauts en droit du bail. D'éventuels dommages économiques résultant de défauts de ces systèmes d'IA

[33] RS **0.104**.

[34] RS **0.108**.

[35] RS **0.109**.

ne vont a priori pas impliquer d'atteinte à des droits fondamentaux. Cependant, la réponse à cette question ne saurait être absolue : le fait de savoir si un droit fondamental se prête ou non à une réalisation dans les relations entre particuliers dépend au final de l'interprétation de chaque droit, du bien juridique qu'il protège, des fonctions qu'il remplit et des circonstances dans lesquelles il est appliqué.[36]

À noter que la présente analyse ne se détermine pas sur le siège de la matière lorsqu'elle évoque un éventuel besoin de légiférer dans le secteur privé. Il ne s'agira pas nécessairement de règles de droit privé (CC, CO, etc.). Des prescriptions de droit public applicables aux particuliers pourraient aussi entrer en ligne de compte, ou bien des réglementations mixtes comme la LPD ou la LEg, qui contiennent les deux types de normes.

Concernant le secteur public

S'agissant du secteur public, il convient de relever que l'État est soumis à des impératifs relevant de son fonctionnement, qui ne s'imposent pas au secteur privé et qui peuvent justifier des règles différentes. En droit suisse, l'impératif le plus important est certainement celui de la légalité, selon lequel l'activité de l'État doit se fonder sur une base légale (art. 5 et 164 Cst. ; cf. également l'art. 36 s'agissant de la restriction des droits fondamentaux). L'utilisation des systèmes d'IA n'échappe pas à ce principe.[37]

L'État peut se servir des systèmes d'IA dans différents domaines et à des degrés plus ou moins poussés. Les possibilités d'utilisation vont du simple soutien sur le plan interne, avec aucune ou très peu de répercussions extérieures (p. ex. applications qui distribuent automatiquement les tâches entre les membres du personnel, qui traduisent des documents ou qui rédigent des procès-verbaux de réunion), aux systèmes qui prennent eux-mêmes les décisions (automatisation complète), en passant par des systèmes qui soutiennent l'administration dans ses prises de décision (automatisation partielle).[38] Ce dernier cas de figure semble le plus fréquent.

La réponse à la question de savoir s'il faut une base légale expresse, de quel niveau et de quelle densité dépendra principalement de l'importance de la matière traitée (art. 164 Cst.),

---

[36] Voir CR Cst.-VINCENT MARTENET, art. 35 N 77.

[37] NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 7 N 17.

[38] OFJ, Révision totale de la loi fédérale sur la protection des données (LPD) – Aperçu des principales modifications en vue de l'élaboration des bases légales concernant le traitement de données par les organes fédéraux, octobre 2022, 23 s., disponible sous www.ofj.admin.ch > État & Citoyen > Protection des données > Informations destinées aux organes fédéraux > Révision totale de la loi fédérale sur la protection des données (LPD) – Aperçu des principales modifications en vue de l'élaboration des bases légales concernant le traitement de données par les organes fédéraux (consulté le 27 août 2024).

ou des restrictions possibles aux droits fondamentaux (art. 36 Cst.).[39] Les systèmes de soutien internes, avec pas ou peu de répercussions externes, relèvent en principe de l'administration auxiliaire et trouvent leur légitimité dans la base légale qui règle les tâches de l'autorité (p. ex applications qui distribuent automatiquement les tâches entre les membres du personnel, qui traduisent des documents ou qui rédigent des procès-verbaux de réunion).[40] Tel devrait être le cas également des *chatbots*, pour autant qu'ils se limitent à donner des informations et renseignements, et que, si des données personnelles sont traitées, la base légale correspondante existe. Dans ce cas-là en effet, on peut considérer qu'il s'agit du choix d'un canal de communication, pour lequel les autorités sont en principe libres.[41] En revanche, les décisions entièrement ou partiellement automatisées nécessitent une base légale formelle lorsqu'elles risquent de porter une atteinte grave aux droits fondamentaux des personnes concernées (art. 36, al. 1, et 164, al. 1, let. b Cst.)[42] et / ou constituent des dispositions fondamentales relatives à l'organisation et à la procédure des autorités fédérales (art. 164, al. 1, let. g, Cst.). Une base légale matérielle *a minima* devrait être nécessaire par exemple si le soutien va si loin que l'algorithme, dans le cadre d'une procédure administrative en particulier, prépare un texte que la personne humaine n'a plus qu'à vérifier.[43]

Dans le domaine de la protection des données, le législateur a concrétisé les exigences de base légale à l'art. 34 LPD. Selon cette disposition, qui cristallise l'art. 36 Cst., les organes fédéraux ne sont sauf exception (cf. art. 34, al. 4, et 36, al. 2, LPD) en droit de traiter des données personnelles que s'il existe une base légale. L'al. 2 prévoit que la base légale doit en principe être formelle pour les traitements de données sensibles (let. a), les profilages (let. b), et les traitements dont la finalité ou le mode de traitement sont susceptibles de porter gravement atteinte aux droits fondamentaux de la personne concernée (let. c). Les systèmes d'IA qui traitent des données personnelles sont évidemment soumis à ces exigences. Ceux qui conduisent à une automatisation complète ou partielle d'une décision, s'ils ne tombent pas déjà sous le coup des let. a et b de l'art. 34, pourront tomber dans le champ d'application de la let. c. Cela dépendra du risque d'atteinte aux droits fondamentaux, en particulier à la protection de la sphère privée et à la protection des données (art. 13 Cst.). De manière générale, plus le mode de traitement est opaque, plus la marge de manœuvre est importante et plus la collecte des données provient de sources différentes, plus ce risque sera considéré comme élevé et nécessitant une base légale formelle.[44]

---

[39] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz in der Verwaltung : rechtliche und ethische Fragen, Schlussbericht vom 28. Februar 2021 zum Vorprojekt IP6.4, 34.

[40] CATHERINE REITER, Künstliche Intelligenz im Verwaltungsverfahren – Ermessen als Stolperstein?, PJA 2022, 984 ss, 988 ; PHILIP GLASS, Datenschutzrecht für künstliche Intelligenz in der öffentlichen Verwaltung : eine Auslegeordnung am Beispiel des Kantons Zürich, in : Michael Widmer (éd.), Datenschutz : Rechtliche Schnittstellen, Zurich 2023, 177 ss, 208.

[41] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al. (n. 39), 58 s. Cela ne signifie cependant pas qu'il n'y a pas d'autres questions juridiques qui se posent, notamment en lien avec le respect de la bonne foi (art. 9 Cst.), cf. pour plus de développements sur ce thème NADJA BRAUN BINDER/LILIANE OBRECHT/GRACE WITTMER, Vertrauensschutz bei fehlerhaften Behördenauskünften durch Chatbots, IusNet DigR 2024 ; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 5 N 10 ss, 18 N 45 ss.

[42] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (n. 39), 59.

[43] CATHERINE REITER, Künstliche Intelligenz im Verwaltungsverfahren (n. 40), 984 ss, 988.

[44] MONIQUE COSSALI SAUVAIN, in : Yaniv Benhamou/Bertil Cottier (éds.), Petit commentaire LPD, art. 34 N 35 à 37 ; OFJ, Guide de législation en matière de protection des données – Conséquences de la nouvelle loi sur la protection des données sur l'élaboration de bases légales, Berne

4.2.3.2 Sécurité nationale

L'art. 3, par. 2, est le fruit d'un consensus entre les États qui souhaitaient exclure le secteur de la sécurité nationale du champ d'application de la convention sur l'IA, et les États qui étaient en faveur d'une inclusion complète.

Finalement, le texte prévoit que les Parties peuvent, mais ne doivent pas, appliquer la convention aux activités menées dans le cadre du cycle de vie des systèmes d'IA liées à la protection de leurs intérêts de sécurité nationale, sans égard à l'entité qui exerce ces activités. Dans tous les cas les activités en question doivent demeurer conformes au droit international applicable, dans la mesure où la sécurité nationale est incluse dans le champ d'application de nombreuses conventions. Le rapport explicatif cite à titre indicatif la CEDH et les deux Pactes de l'ONU.[45] Il est également précisé que ces activités doivent être menées dans le respect des institutions et processus démocratiques des Parties.[46]

La sécurité nationale est considérée comme un concept « essentiellement contesté », soit une notion sur le contenu de laquelle il n'est pas possible de trouver un consensus au niveau international.[47] Cette situation laisse aux États une marge de manœuvre considérable, et entraîne en doctrine une diversité de propositions de définitions.[48]

Sur le plan interne, le Conseil fédéral a développé la classification suivante, qui repose sur une notion large de la sécurité nationale : 1) prévention policière des menaces, protection de l'État et poursuite pénale ; 2) prévention, prévision et maîtrise de catastrophes naturelles et anthropiques ; 3) défense contre une attaque militaire ; 4) sauvegarde des intérêts de la Suisse à l'étranger et contribution à la gestion internationale des crises.[49]

2024, ch. 3.2.3, 21, disponible sous www.ofj.admin.ch > Nouveau droit de la protection des données > Informations destinées aux organes fédéraux > Guide de législation en matière de protection des données (consulté le 27 août 2024).

45 Pacte international relatif aux droits économiques, sociaux et culturels (Pacte I, RS **0.103.1**) et Pacte international relatif aux droits civils et politiques (Pacte II, RS **0.103.2**).

46 Rapport explicatif convention sur l'IA (n. 20), N 32.

47 BSK BK-OLIVER DIGGELMANN/TILMANN ALTWICKER, art. 57, N 9 et réf. citée.

48 FRANCK ELONG MBOULÉ, Le régime juridique des biens des organisations internationales, Zurich 2022, 294 ss, 296.

49 Rapport du Conseil fédéral à l'Assemblée fédérale sur la politique de sécurité de la Suisse du 23 juin 2010, FF **2010** 4681, 4707 s.

Pour l'Assemblée parlementaire du Conseil de l'Europe, la sécurité nationale consiste à combattre toute menace visible et réelle pour l'ordre démocratique de l'État et de la société.[50] Cette définition n'est toutefois pas reprise par les différents États européens, car elle ne tient pas compte des besoins des agences pour lutter contre les menaces occultes ou spéculatives.[51]

La convention sur l'IA n'a pas pour vocation de clarifier la situation, mais le rapport explicatif précise que les activités courantes de maintien de l'ordre en lien avec la prévention, la détection, l'investigation et la poursuite de crimes, y compris les menaces à la sécurité publique, ne sont pas concernées dans la mesure où la sécurité nationale des Parties n'est pas en jeu.[52] Selon l'interprétation de OFJ, ces activités régulières se distinguent des mesures secrètes des services de renseignement, qui s'inscrivent en principe dans l'exception de la sécurité nationale.

A noter que l'exception ne s'applique pas aux activités des systèmes d'IA à double usage dans la mesure où un autre but que la préservation de la sécurité nationale est également poursuivi.[53]

Durant les négociations autour de la convention sur l'IA, la délégation suisse s'est positionnée en faveur d'une inclusion totale du secteur de la sécurité nationale dans le champ d'application du texte. Il n'y a pas de raison de s'écarter de cette position dans le cadre de la présente analyse. Par ailleurs, en droit suisse, les droits fondamentaux s'appliquent aussi dans ce domaine, même si des restrictions sont possibles aux conditions usuelles de l'art. 36 Cst., notamment en présence d'un intérêt public prépondérant. L'analyse part donc de l'hypothèse que les activités des systèmes d'IA en lien avec la protection de la sécurité nationale, et en particulier les activités du SRC, entrent dans le champ d'application de la convention.

4.2.3.3 Recherche et développement

Les activités de recherche et de développement sont exclues du champ d'application de la convention à certaines conditions. Il faut que les systèmes d'IA en question n'aient pas encore été rendus disponibles à l'utilisation, et que les essais ou les activités similaires ne soient pas susceptibles d'interférer avec les droits de l'homme, la démocratie et l'État de droit (art. 3, par. 3).

---

[50] Recommandation 1402 (1999) de l'Assemblée parlementaire du Conseil de l'Europe, Contrôle des services de sécurité intérieure dans les Etats membres du Conseil de l'Europe, point A. 2.

[51] CR Cst.-OLIVIER BLEICKER, art. 57 N 34 et réf. en nbp 86.

[52] Rapport explicatif convention sur l'IA (n. 20), N 32 *in fine*.

[53] Rapport explicatif convention sur l'IA (n. 20), N 32.

D'après le rapport explicatif, ces activités devraient dans tous les cas respecter les droits de l'homme et le droit national applicables, ainsi que les normes éthiques et professionnelles reconnues en matière de recherche scientifique. Les systèmes d'IA qui sont rendus disponibles à l'utilisation après la phase de recherche et de développement devraient quant à eux en principe respecter la convention, y compris s'agissant de leur conception et leur développement.[54]

À noter que cette exception au champ d'application ne concerne pas les obligations découlant de l'art. 13 (« Innovation sûre »), et 25 par. 2 (« Coopération internationale »).

4.2.3.4 Défense nationale

Le domaine de la défense nationale est exclu du champ d'application de la convention (art. 3, par. 4). Il s'agit de l'expression du principe ancré à l'art. 1, let. d, du Statut du Conseil de l'Europe, qui prévoit que les questions relatives à la défense nationale ne sont pas de la compétence de cette organisation.

Cela ne sous-entend en aucun cas que les activités menées dans le cadre du cycle de vie des systèmes d'IA relatives à la défense nationale ne sont pas couvertes pas le droit international.[55] En Suisse, comme pour la sécurité nationale, les droits fondamentaux s'appliquent aussi dans ce domaine. Il ne s'agirait donc pas de prévoir des exclusions, mais tout au plus des exceptions ou des aménagements là où cela pourrait se justifier dans l'intérêt de la défense nationale de la Suisse.

4.2.3.5 En résumé

Les principaux éléments relatifs au champ d'application de la convention selon l'art. 3 peuvent être résumés comme suit :

- Secteur public / secteur privé (art. 3, par. 1) :

  La convention s'applique aux deux secteurs. Cependant, pour le secteur privé, les Parties ont une plus grande marge de manœuvre, pour autant qu'elles répondent, d'une manière conforme à l'objet et au but de la convention, aux risques et aux impacts découlant des activités menées dans le cadre du cycle de vie des systèmes d'IA.

  En cas de ratification, la Suisse appliquerait les principes et obligations énoncés aux Chapitres II à VI de la convention au secteur privé également. Des mesures purement administratives ou volontaires ne paraissent pas suffisantes eu égard à notre ordre juridique. Cependant, des différences entre les deux secteurs qui tiennent compte des particularités du droit suisse sont à envisager, notamment eu égard à l'effet en principe uniquement vertical des droits fondamentaux, et compte

[54] Rapport explicatif convention sur l'IA (n. 20), N 33 s.

[55] Rapport explicatif convention sur l'IA (n. 20), N 36.

tenu du fait que les exigences légales sont en principe plus élevées dans le secteur public.

- Sécurité nationale (art. 3, par. 2) :

  Les États ne sont pas obligés d'appliquer la convention dans le domaine de la sécurité nationale. Il n'existe pas de consensus global sur la portée de ce terme. Il se réfère surtout aux mesures secrètes des services de renseignement. Pour les besoins de la présente analyse, l'OFJ a choisi l'hypothèse selon laquelle les dispositions de la convention seraient appliquées aussi dans ce secteur.

- Recherche et développement (art. 3, par. 3) :

  La convention ne s'applique pas aux activités de recherche et développement, à moins que des essais soient effectués de manière telle qu'ils sont susceptibles de porter atteinte aux droits de l'homme, à la démocratie et à l'État de droit. Les art. 13 (« Innovation sûre ») et 25, par. 2 (« Coopération internationale ») de la convention ne sont pas concernés par cette exception.

- Défense nationale (art. 3, par. 4) :

  Les questions relatives à la défense nationale ne sont pas de la compétence du Conseil de l'Europe et donc pas non plus de la convention.

4.3 Obligations et principes

4.3.1 Chapitre II : Obligations générales

4.3.1.1 Article 4 – Protection des droits de l'homme

Selon l'art. 4, chaque Partie adopte ou maintient les mesures pour veiller à ce que les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle soient cohérentes avec les obligations de protection des droits de l'homme, telles qu'elles sont consacrées par le droit international applicable et par son droit interne.

Les Parties doivent veiller à ce que ce le cadre légal en vigueur fournisse une protection tout aussi valable et efficace des droits de l'homme dans le contexte spécifique des systèmes d'IA qu'en dehors.

Le but même de la convention sur l'IA est de garantir que les activités menées dans le cadre du cycle de vie des systèmes d'IA soient pleinement compatibles notamment avec le respect des droits de l'homme. Des obligations spécifiques sont prévues à cet effet, soit en particulier celle de prévoir un cadre de gestion des risques et des impacts des systèmes d'IA sur les droits de l'homme, la démocratie et l'État de droit (cf. ch. 4.3.4). Leur mise en œuvre permet ainsi aux Parties de se conformer indirectement à l'obligation générale de protection des droits de l'homme figurant à l'art. 4 de la convention sur l'IA.

En droit suisse, comme rappelé ci-dessus (cf. ch. 3), divers instruments internationaux de protection des droits de l'homme s'appliquent déjà en cas de recours à l'IA, en particulier la CEDH. En droit interne, les garanties de protection des droits fondamentaux s'appliquent déjà de la même manière dans le contexte spécifique des systèmes d'IA. Ainsi, en particulier, le

régime consacré à l'art. 36 Cst., qui subordonne en général la restriction d'un droit fondamental à quatre conditions cumulatives (base légale, intérêt public, proportionnalité et respect du noyau intangible du droit) trouve application dans le secteur public. L'origine de l'activité étatique dont procède la restriction (être humain ou système d'IA) n'altère pas la protection des droits fondamentaux.

Pour ce qui est du secteur privé, il convient de relever que selon l'art. 35, al. 1, Cst., les droits fondamentaux doivent être réalisés dans l'ensemble de l'ordre juridique. À l'art. 35, al. 3, Cst., il est précisé que les autorités veillent à ce que les droits fondamentaux, dans la mesure où ils s'y prêtent, soient aussi réalisés dans les relations qui lient les particuliers entre eux. Cette disposition confirme que les droits fondamentaux ne peuvent en principe pas déployer leurs effets directement entre individus, mais que cette application dépend d'une intervention spécifique de l'autorité.[56] Elle vise tant l'adoption de règles de droit que l'interprétation et l'application de celles-ci. Le seul cas dans lequel la Cst. reconnaît un effet horizontal direct est l'art. 8, al. 3, Cst., troisième phrase, qui garantit le principe de l'égalité des salaires entre les hommes et les femmes.

Ainsi, la transposition en droit interne de l'obligation de protection des droits humains prévue à l'art. 4 de la convention devrait tenir compte du fait qu'en droit suisse, les droits fondamentaux n'ont généralement qu'un effet horizontal indirect s'agissant des relations entre particuliers.

Il découle de ce qui précède que le cadre juridique suisse de protection des droits fondamentaux trouve déjà application en cas de recours à des systèmes d'IA, avant tout dans le secteur public. La convention exige que ce cadre soit suffisant et efficace pour répondre aux défis spécifiques de ces systèmes. C'est l'objectif même de toutes les règles et de tous les principes généraux contenus dans la convention sur l'IA que de renforcer la protection des droits fondamentaux. On se référera donc aux développements ci-après pour l'examen de l'efficacité de la protection offerte en droit suisse et les éventuels besoins d'agir.

4.3.1.2 Article 5 – Intégrité des processus démocratiques et respect de l'État de droit

Selon l'art. 5, par. 1, chaque Partie adopte ou maintient les mesures visant à garantir que les systèmes d'IA ne sont pas utilisés pour porter atteinte à l'intégrité, à l'indépendance et à l'efficacité des institutions et processus démocratiques. Cela englobe également la séparation des pouvoirs, le respect de l'indépendance de la justice et l'accès à la justice. Selon l'art. 5, par. 2, chaque Partie adopte ou maintient des mesures qui visent à protéger ses processus démocratiques dans le cadre des activités menées au cours du cycle de vie des systèmes d'IA, y compris l'accès équitable et la participation des personnes au débat public, ainsi que leur capacité à se forger librement une opinion.

[56] GIORGIO MALINVERNI/MICHEL HOTTELIER/MAYA HERTIG RANDALL/ALEXANDRE FLÜCKIGER, Droit constitutionnel suisse – Volume II : Les droits fondamentaux, Berne 2021, N 135.

La disposition ne définit pas la notion d'« institutions et processus démocratiques », ni celle d'État de droit, ce qui contribue à la rendre plutôt vague. L'analyse la définit comme l'obligation d'adopter ou de maintenir des mesures visant à protéger la démocratie et l'État de droit en tant que principes structurels à la base des pays démocratiques.

Les impacts des systèmes d'IA sur les institutions et processus démocratiques et l'État de droit peuvent être très variés. Comme relevé dans le rapport explicatif, en s'acquittant des obligations prévues à l'art. 5, les Parties peuvent se concentrer sur certains risques posés par l'IA, par exemple celui d'atteinte au pluralisme politique.[57] La convention n'indique pas un risque particulier à appréhender ni une mesure spécifique à prendre. Bien que vaste, la disposition a l'avantage de résister à l'épreuve du temps, puisqu'elle permettra d'englober des risques pas encore connus.

Sont exposés ci-après des exemples d'impacts négatifs que les systèmes d'IA peuvent avoir sur l'intégrité des processus démocratiques et le respect de l'État de droit, à l'aune de certaines dispositions du droit suisse.

- Désinformation et informations erronées

  Les systèmes d'IA peuvent amplifier le phénomène de la désinformation et de la diffusion d'informations erronées, dont le but est d'amener le public à croire des choses fausses. La désinformation peut orienter l'opinion publique et influencer les élections et les votations. Les fausses nouvelles peuvent prendre plusieurs formes : déclarations, expressions d'opinions infondées ou discours haineux contre des groupes sociaux ou des minorités. Des logiciels automatisés (« bots »), qui imitent le comportement humain sur les médias sociaux, en publiant des informations, en affichant leur approbation (« like ») ou encore en s'adressant à des personnes réelles, peuvent par exemple polariser l'opinion publique en diffusant des discours haineux. Les systèmes d'IA peuvent aussi être utilisés pour imiter des voix et des images afin de créer des réalités alternatives, en recourant ainsi aux « deep fakes ».[58] La distinction entre fiction et réalité continue de s'éroder. En outre, des informations inventées et trompeuses peuvent être diffusées par des chatbots, ce qui concourt à la désinformation et, en fonction des contextes, peut devenir un problème pour la démocratie.[59]

---

[57] Rapport explicatif convention sur l'IA (n. 20), N 46.

[58] YVES-MARIE DOUBLET, Désinformation et campagnes électorales, rapport à l'attention du Conseil de l'Europe, juin 2019, 5, 11 et 16 ; voir aussi MURAT KARABOGA/NULA FREI et al., Deepfakes und manipulierte Realitäten. Technologiefolgenabschätzung und Handlungsempfehlungen für die Schweiz, Ta-Swiss, TA 81/2024, Zollikon:vdf.

[59] Cf. l'étude réalisée par Algorithmwatch et Forensics qui a testé les réponses d'un chatbot en lien avec les élections en Suisse, en Bavière et en Hesse en octobre 2023, disponible sous https://algorithmwatch.ch > Publications > Le chatbot d'IA fournit des réponses erronées aux questions sur les élections démocratiques (consulté le 27 août 2024). Voir aussi la récente étude d'août 2024 en lien avec les élections régionales en

Le cadre légal suisse contient déjà des normes pertinentes qui, dans certaines circonstances, peuvent s'appliquer en matière de désinformation et informations erronées :

- Dans le contexte spécifique d'élections et votations, l'art. 34, al. 2, Cst. protège la libre formation de la volonté des électeurs et l'expression non faussée de celle-ci. La LDP contient des dispositions spécifiques s'agissant des élections et des votations au niveau fédéral. En particulier, selon l'art. 77, al. 1, let. b, LDP, un recours peut être déposé en cas d'irrégularités affectant les votations.

  Le TF a établi qu'aucun résultat de votation ne peut être reconnu s'il n'exprime pas de manière fiable et non faussée la libre volonté des citoyens. Si la volonté libre n'a pas été exprimée de manière fiable et non faussée, la votation doit être répétée. Ces règles visent en premier lieu à empêcher une influence inadmissible de l'État. Toutefois, les particuliers peuvent également être à l'origine de la désinformation et influencer la libre formation de la volonté des citoyens. Tel est le cas, par exemple, lorsque des informations manifestement inexactes ou fallacieuses sont diffusées à une date si proche du scrutin que les citoyens ne sont plus en mesure de se renseigner de manière fiable à d'autres sources. Compte tenu de la liberté d'expression, une atteinte de ce type à la formation de l'opinion des citoyens n'est pas reconnue à la légère. L'annulation d'un scrutin ne doit être envisagée qu'avec une grande retenue et seulement en cas de manquements particulièrement graves.[60]

  En outre, le TF reconnaît que la rectification d'informations manifestement fausses ou trompeuses peut constituer un motif valable justifiant une intervention des autorités dans la campagne de votation. Dans certaines circonstances, il existe même une obligation pour les autorités d'intervenir afin de garantir le droit fédéral à la libre formation de la volonté et à l'expression non faussée du vote. Les autorités disposent toutefois d'une grande marge d'appréciation à cet égard. Un devoir d'intervention, dont la violation peut conduire à l'annulation de la votation, n'est en principe admis que si la prise d'influence

Allemagne, disponible sous https://algorithmwatch.org/de > Blog > Chatbots bringen noch immer viele Falschinformationen in Umlauf (consulté le 28 août 2024).

[60] ATF **135** I 292, consid. 4.1.

d'acteurs privés entrave de manière très grave ou rend carrément impossible la formation de la volonté des électeurs.[61]

Les considérations qui précèdent s'appliquent de la même manière en cas d'utilisation de systèmes d'IA dans le but de désinformer. La diffusion d'une image faussée dans un contexte particulier peut par exemple permettre des manipulations politiques.[62]

Une tromperie grave quant à l'objet d'une votation sur les médias sociaux par des particuliers soulèvera en premier lieu la question de l'obligation pour les autorités d'intervenir pour corriger ou rectifier les informations concernées.[63] Ensuite, s'agissant de l'annulation du scrutin, rappelons que la jurisprudence n'exige pas la preuve de l'influence des irrégularités sur le résultat du vote ; il suffit au contraire qu'elle apparaisse possible. À défaut d'une constatation chiffrée de l'impact d'une irrégularité dans la procédure, son influence sur le résultat est déterminée selon l'ensemble des circonstances et, en principe, avec plein pouvoir d'examen.[64] Cet allègement pourrait s'avérer particulièrement utile lorsque la désinformation a lieu en ligne par le biais de systèmes d'IA.

Les dispositions en matière de droits politiques précitées n'interviennent toutefois qu'une fois que l'atteinte a été commise et ne permettent pas, à titre préventif, d'encadrer les structures mêmes de production et de diffusion des informations.

- Les plateformes en ligne jouent un rôle de plus en plus important dans la communication et la formation d'opinion. Elles ne se contentent pas de contribuer à la liberté d'expression et d'information, elles influencent également le débat public. De plus, la facilité avec laquelle des contenus peuvent être produits et diffusés peut entraîner des conséquences néfastes.

  Outre les règles de droit générales, il n'existe aujourd'hui aucune norme spécifique ni de pratique juridique établie sur la question de la responsabilité des

---

[61] Arrêt du TF, 1C_472/2010, consid. 4.3 ; NADJA BRAUN BINDER/MANUELA KÄLIN, Rechtliche Aspekte der politischen Meinungsbildung, in : Urs Bieri et al. (éds.), Digitalisierung der Schweizer Demokratie. Technologische Revolution trifft auf traditionelles Meinungsbildungssystem, TA-Swiss, TA 75/2021, Zurich:vdf, 125 ss, 129.

[62] Pour un exposé des divers dangers possibles, cf. TOM LEBRUN, La liberté d'expression aux Etats-Unis et au Canada face au risque de propagande générée par l'intelligence artificielle, in : Céline Castets-Renard/Jessica Eynard (éds.), Un droit de l'intelligence artificielle. Entre règles sectorielles et régime général. Perspectives comparées, Bruxelles, 2023, 409 ss.

[63] NADJA BRAUN BINDER/MANUELA KÄLIN, Rechtliche Aspekte der politischen Meinungsbildung (n. 61), 145.

[64] ATF **135** I 292, consid. 4.4.

intermédiaires pour les contenus illégaux publiés par des tiers sur leurs plateformes. De même, il n'existe pratiquement aucune prescription visant à renforcer les droits des utilisateurs vis-à-vis des intermédiaires ou à obliger les plateformes à faire preuve de davantage de transparence.

Pour cette raison, le Conseil fédéral a chargé le DETEC (OFCOM) de rédiger un avant-projet visant à réguler les grandes plateformes en ligne en leur imposant un devoir de diligence. Celles-ci devraient par exemple mettre en place des procédures de prévention et de suppression des discours de haine illicites, offrir un point de contact et désigner un représentant en Suisse. L'avant-projet aura également pour objectif de renforcer les droits des utilisateurs, notamment en obligeant les plateformes à justifier les mesures prises à l'encontre de leurs utilisateurs et à leur offrir des voies de recours.

Plusieurs de ces aspects concernent notamment les systèmes algorithmiques employés par les plateformes (p. ex. à des fins de modération ou de classement des contenus). La réglementation prévue contribuera dès lors aussi à la protection de la démocratie (en l'occurrence de la liberté de formation de l'opinion) et de l'État de droit.

- L'identification des contenus générés par des systèmes d'IA constitue aussi un moyen de lutter contre la désinformation en ligne. Cela ne limite pas en soi la diffusion de la désinformation, mais permet aux destinataires de mieux évaluer et de contextualiser les informations qu'ils reçoivent. En ce sens, la mise en œuvre de l'obligation de transparence prévue à l'art. 8 de la convention sur l'IA est particulièrement pertinente dans le contexte de la protection de l'intégrité des processus démocratiques et de l'État de droit. L'analyse y revient ci-dessous (cf. ch. 4.3.2.3). Un marquage permettrait notamment d'informer les utilisateurs qu'un compte ou une publication ne sont pas l'œuvre d'un être humain.

- Le droit pénal offre aussi un cadre légal qui, à certaines conditions, peut être pertinent et fixer des limites à certains comportements pénalement répréhensibles. En effet, les dispositions du CP sont en principe neutres sur le plan technologique et applicables quelle que soit la méthode utilisée par l'auteur d'une infraction. Le droit pénal permet donc de sanctionner certains crimes et délits contre la paix publique (art. 258 ss CP) et délits contre la volonté populaire (art. 279 ss CP). Par exemple, l'art. 261bis CP sur la discrimination et l'incitation à la haine s'applique, à certaines conditions, en présence de discours haineux sur les réseaux sociaux. Il en va de même des art. 173 ss CP sur les délits contre l'honneur, ou encore de l'art. 179^{decies} CP sur l'usurpation d'identité (cf. sur ces dispositions ch. 6.6). En droit civil, le régime de protection de la personnalité (art. 28 ss CC) peut aussi trouver application (cf. pour des considérations en lien avec les deepfakes ch. 6.3.3).

  Lors de l'application de ces normes, la liberté d'expression ancrée à l'art. 16, al. 2, Cst. doit aussi être respectée. Celle-ci couvre également les affirmations

fausses et trompeuses.[65] Des restrictions demeurent possibles, mais devront satisfaire les exigences de l'art. 36 Cst.[66]

- Ciblage du public à des fins politiques

En lien avec les risques pour la libre formation de l'opinion des citoyens, il convient aussi de mentionner la question du ciblage du public à des fins politiques par le biais du recours aux systèmes d'IA et les aspects de protection des données. Grâce à l'usage d'algorithmes, il est possible d'influencer le comportement des citoyens, par exemple en envoyant des informations ciblées à des groupes dont on a établi le profil, en supposant que leurs membres réagissent particulièrement à un certain type de message.

a) Cadre légal posé par la LPD

En droit suisse, le cadre posé par la nouvelle LPD entrée en vigueur le 1er septembre 2023 doit en particulier être respecté dans ce contexte.[67]

Le traitement de données en lien avec les opinions politiques des citoyens correspond à un traitement de données sensibles (art. 5, al. 1, let. c, LPD). Ces données bénéficient donc d'une protection qualifiée. Par ailleurs, le fait de croiser, à l'aide de méthodes d'analyse automatisées, des données provenant de plusieurs sources dans le but d'identifier les opinions et les tendances politiques des électeurs correspond à une activité de profilage (art. 5, al. 1, let. f, LPD). Selon les circonstances, il pourrait même s'agir d'un profilage à risque élevé (art. 5, al. 1, let. g, LPD).

Les principes de bonne foi, transparence, proportionnalité, finalité, exactitude et sécurité des données doivent notamment être respectés. Selon les circonstances, lorsque le traitement envisagé est susceptible d'entraîner un risque élevé pour la personnalité ou les droits fondamentaux de la personne concernée, le responsable du traitement doit procéder au préalable à une analyse d'impact relative à la protection des données personnelles (art. 22, al. 1, LPD). L'existence d'un risque élevé, en particulier lors du recours à de nouvelles technologies, comme l'IA, dépend de la nature, de l'étendue, des circonstances et de la finalité du traitement. Un tel risque existe notamment en cas de traitement de données sensibles à grande échelle (art. 22, al. 2, let. b, LPD),

---

65 FLORENT THOUVENIN/STEPHANIE VOLZ/MARK EISENEGGER/DANIEL VOGLER/MARIELA JAFFÉ, Governance von Desinformation in digitalisierten Öffentlichkeiten, Jusletter 5 février 2024, 9 N 24, et les réf. citées.

66 RAPHAELA CUENI, Falsche und irreführende Informationen im Verfassungsrecht der Schweiz, ex ante 1/2019, 3 ss, en particulier 9 ss.

67 Cf. à ce sujet le Guide du 15 décembre 2022 des autorités de protection des données de la Confédération et des cantons concernant le traitement numérique de données personnelles dans le cadre d'élections et de votations en Suisse, disponible sous https://www.edoeb.admin.ch > Protection des données > Internet & Technologie > Guide relatif aux élections et votations (consulté le 27 août 2024) ; voir aussi MICHAEL MONTAVON, Big Data, un outil d'influence en période électorale, 2023, www.swissprivacy.law/258.

en particulier en cas de profilage à grande échelle visant à identifier les opinions et les tendances politiques des électeurs.

b) Publicité politique

Au niveau suisse, les art. 10 LRTV et 17, al. 3, ORTV interdisent la publicité politique à la radio et à la télévision. Le champ d'application de ces dispositions est toutefois restreint et n'englobe pas les autres médias, y compris les médias sociaux.[68]

Dans l'UE, la publicité à caractère politique est régie par des règles spécifiques depuis le 1er avril 2024[69]. Celles-ci incluent notamment des obligations précises pour la publicité à caractère politique en ligne. Dans le cadre de la rédaction de l'avant-projet visant à réguler les plateformes de communication, l'OFCOM étudie la pertinence d'obligations similaires en Suisse.

- Risques pour l'exercice des droits fondamentaux, en particulier les libertés de la communication

La protection de la démocratie et de l'État de droit va aussi de pair avec la protection des droits fondamentaux, en particulier les libertés de la communication (notamment la liberté d'opinion et d'information [art. 16. Cst.], la liberté des médias [art. 17 Cst.], la liberté de réunion [art. 22 Cst.]), qui leur sont consubstantielles. Les systèmes d'IA peuvent notamment être utilisés à des fins de surveillance, par exemple moyennant des outils de reconnaissance faciale dans l'espace public, et porter ainsi potentiellement atteinte notamment à la liberté d'expression, à la liberté de réunion et à la sphère privée d'un grand nombre de personnes.

Dans son arrêt du 4 juillet 2023 Glukhin v. Russie, déjà mentionné[70], la CourEDH a condamné la Fédération de Russie pour violation de l'art. 8 CEDH (droit au respect de la vie privée) et de l'art. 10 CEDH (liberté d'expression), en raison du recours à la technologie de reconnaissance faciale pour identifier et localiser M. Glukhin, après que celui-ci se fut livré à une manifestation en solo dans le métro de Moscou. La Cour a conclu que le traitement des données personnelles de M. Glukhin dans le contexte de sa manifestation pacifique, laquelle n'avait menacé ni l'ordre ni la sécurité publics, s'est révélé particulièrement intrusif. Le recours à la technologie de reconnaissance faciale dans son cas a été jugé incompatible avec les idéaux et valeurs d'une société démocratique régie par la prééminence du droit. En droit suisse, les art. 8 et 10 CEDH

---

[68] NADJA BRAUN BINDER/MANUELA KÄLIN, Rechtliche Aspekte der politischen Meinungsbildung (n. 61), 134 ss.

[69] Règlement (UE) 2024/900 du Parlement européen et du Conseil du 13 mars 2024 relatif à la transparence et au ciblage de la publicité à caractère politique, PE/90/2023/REV/1 ; JO L, 2024/900, 20 mars 2024.

[70] CourEDH, *Glukhin c. Russie* (n. 6).

se retrouvent aux art. 13 et 16, al. 2, Cst. Les principes dégagés par la CEDH sont contraignants pour la Suisse.

Le cadre légal suisse en vigueur en matière de protection des droits fondamentaux s'avère donc pertinent aussi s'agissant de la mise en œuvre de l'art. 5 de la convention sur l'IA.

Il convient de conclure que l'objet de protection de l'art. 5, à savoir l'intégrité des processus démocratiques et le respect de l'État de droit, est appréhendé en droit suisse par un large éventail de normes. Ces dernières vont de la protection des droits fondamentaux aux règles en matière d'exercice des droits politiques, en passant par la protection de la personnalité et la protection des données. La future réglementation envisagée des grandes plateformes de communication devrait par ailleurs renforcer les droits des utilisateurs en Suisse et la transparence de la part de ces plateformes, concourant aussi au bon fonctionnement de la démocratie.

Selon les premières analyses ci-dessus, d'autres modifications législatives ne paraissent pas nécessaires. Compte tenu du fait que les développements dans le contexte des dangers pour la démocratie sont particulièrement rapides, avec des conséquences potentiellement importantes, il convient de continuer à observer la situation de près ainsi que l'évolution de la jurisprudence et de sa portée dans le contexte numérique afin de réagir à temps en cas de nécessité. Le rapport du Conseil fédéral en réponse au postulat 22.3006 CPS-N « État des lieux relatif à la menace que constituent pour la Suisse les campagnes de désinformation »[71] est un exemple de suivi utile.

Il convient aussi de relever que la convention sur l'IA contient d'autres dispositions qui peuvent être pertinentes en matière de protection de la démocratie et de l'État de droit, en particulier le principe de transparence (art. 8) et le cadre de gestion des risques et des impacts des systèmes d'IA (art. 16). La mise en œuvre de ces dispositions à vocation générale permettrait également de compléter et renforcer le cadre juridique en vigueur en matière de protection de la démocratie et de l'État de droit.

4.3.2 Chapitre III : Principes relatifs aux activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle

4.3.2.1 Article 6 – Approche générale

L'art. 6 prévoit que les principes qui figurent au Chapitre III de la convention doivent être mis en œuvre par les Parties de manière adaptée à leur ordre juridique interne et aux autres obligations de la convention.

[71] Activités d'influence et désinformation. Rapport du Conseil fédéral en réponse au postulat 22.3006 CPS-N, 19 juin 2024, disponible sous www.admin.ch > Documentation > Communiqués > Activités d'influence et désinformation : le Conseil fédéral insiste sur la résilience et renforce l'analyse et la coordination (consulté le 27 août 2024).

Selon le rapport explicatif de la convention, cet article est particulièrement important dans la mesure où chaque Partie dispose d'un régime juridique détaillé de protection des droits de l'homme avec son propre ensemble de règles, de principes et de pratiques concernant le champ d'application, le contenu des droits substantiels et des restrictions, dérogations ou exceptions possibles à ces droits, ainsi que le fonctionnement des mécanismes de surveillance et de mise en œuvre applicables.[72]

En conclusion, cette disposition reflète le principe de droit international public – déjà exposé plus haut (cf. ch. 4.2.1) – selon lequel les États sont libres quant à la manière dont ils souhaitent mettre en œuvre leurs obligations internationales. En droit suisse, là où une compétence fédérale est donnée, l'art. 164 Cst. prévoit que toutes les dispositions importantes qui fixent des règles de droit doivent être édictées sous la forme d'une loi fédérale.

4.3.2.2 Article 7 – Dignité humaine et autonomie personnelle

Selon l'art. 7, les Parties adoptent ou maintiennent des mesures pour le respect de la dignité humaine et de l'autonomie personnelle en ce qui concerne les activités menées dans le cadre du cycle de vie des systèmes d'IA.

Cette disposition souligne l'importance de la dignité humaine et de l'autonomie personnelle dans le cadre d'une réglementation centrée sur l'humain. Ainsi, les activités menées dans le cadre du cycle de vie des systèmes d'IA ne doivent pas contribuer à la déshumanisation des individus.[73]

L'autonomie personnelle est un aspect important de la dignité humaine qui renvoie à la capacité d'autodétermination des individus, c'est-à-dire leur capacité à faire des choix et à prendre des décisions, y compris sans subir d'influence ou de coercition indue. Dans le contexte de l'IA, l'autonomie personnelle exige que les individus aient le contrôle sur l'utilisation et l'impact des technologies d'IA dans leur vie.[74] Ce principe est donc étroitement lié au principe de transparence et contrôle (art. 8 de la convention, cf. ch. 4.3.2.3).

En droit suisse, au niveau de la Constitution, la dignité humaine est garantie par l'art. 7 Cst. La jurisprudence reconnaît à cette disposition la portée d'un droit fondamental à part entière, permettant d'assurer la protection, par définition absolue, d'éventuels aspects de la dignité humaine que les autres droits fondamentaux ne couvriraient pas.[75] Cependant, en pratique,

---

[72] Rapport explicatif convention sur l'IA (n. 20), N 51.

[73] Rapport explicatif convention sur l'IA (n. 20), N. 53.

[74] Rapport explicatif convention sur l'IA (n. 20), N 55.

[75] CR Cst.-JACQUES DUBEY, art. 7, N 25, et les réf. citées.

la dignité humaine est surtout invoquée à l'appui d'un autre droit fondamental et n'a semble-t-il jamais encore été à elle seule décisive dans un jugement d'une autorité judiciaire fédérale.[76] Ainsi, elle a plutôt le caractère d'un droit de réserve ou de dernier recours. Cela tient à la définition vague de son domaine de protection, puisque l'état de fait visé n'est autre que le fait même d'être humain.[77] Le domaine de protection de la dignité humaine se confond avec son noyau intangible (art. 36, al. 4, Cst.). Toute atteinte de l'État doit ainsi s'entendre comme une violation.[78] Quant à l'auto-détermination et la protection de l'autonomie, elle est couverte tant par l'art. 10, al. 2, Cst. (liberté personnelle) que par l'art. 13 Cst. (protection de la sphère privée).

Compte tenu des défis technologiques liés à l'IA, notamment le risque de déshumanisation des individus[79], le rôle de garant de la dignité humaine va vraisemblablement encore s'accroître. Certains usages, comme le recours à des outils d'IA par les autorités qui prévoiraient une surveillance constante des individus avec un « social scoring », ou l'utilisation de la reconnaissance émotionnelle dans certaines situations, pourraient constituer une atteinte à la dignité humaine, *per se* injustifiable. Le droit constitutionnel permet ainsi déjà d'interdire de tels usages.

Sont aussi pertinentes dans ce cadre les lignes directrices sur l'IA, dont la Confédération s'est dotée en novembre 2020. Ces lignes directrices, non contraignantes, doivent servir de cadre d'orientation général pour le développement de l'IA dans l'administration fédérale et doivent à ce titre garantir une pratique cohérente. La première ligne de conduite consiste à placer l'humain au centre. Dans ce contexte, il est mentionné que la protection des droits fondamentaux prend une dimension particulière dans l'utilisation de l'IA.[80]

---

[76] CR Cst.-JACQUES DUBEY, art. 7, N 26 ; BSK BV-EVA MARIA BELSER/EVA MOLINARI, art. 7 N 39 s. ; *contra* SGK BV-RAINER J. SCHWEIZER/CHRISTOPH SPENLÉ, art. 7 N 22.

[77] CR Cst.-JACQUES DUBEY, art. 7, N 27.

[78] CR Cst.-JACQUES DUBEY art. 7 N 56 ss, en particulier N 59 ; BSK BV-EVA MARIA BELSER/EVA MOLINARI, art. 7 N 63.

[79] Rapport explicatif convention sur l'IA (n. 20), N 53.

[80] Intelligence artificielle – lignes directrices pour la Confédération, Cadre d'orientation en matière d'IA dans l'administration fédérale, 2020, disponible sous www.sbfi.admin.ch > Politique FRI > Encouragement de la formation, de la recherche et de l'innovation 2021-2024 > Thèmes transversaux > Numérisation dans le domaine FRI > Intelligence artificielle (consulté le 26 août 2024). Les lignes directrices doivent être évaluées tous les deux ans. La prochaine évaluation est prévue cette année dans le cadre de l'état des lieux sur les approches de régulation en matière d'IA. Celle de 2022 a montré que les lignes directrices en matière d'IA sont bien connues dans l'administration fédérale et qu'elles sont prises en considération et appliquées par les collaborateurs qui travaillent avec l'IA (que ce soit en utilisant l'IA, au niveau de la réglementation, etc.). Tout en relevant qu'il est nécessaire discuter plus en détail des lignes directrices et de leur application concrète au sein de la Confédération, l'évaluation n'a pas révélé de besoin de les adapter ou de les actualiser, cf. Rapport OFCOM, Suivi des lignes directrices intelligence artificielle

Se pose la question de la portée de l'art. 7 de la convention dans le cadre de la mise en œuvre de cette disposition dans le secteur privé (cf. art. 3, par. 1, let. b de la convention, ch. 4.2.3.1). Dans l'ordre juridique suisse, il existe déjà des règles réalisant la garantie de la dignité humaine et l'autonomie individuelle entre privés. Tel est le cas en particulier de l'art. 28 CC sur la protection de la personnalité (cf. pour des développements en lien avec ce point ch. 6.3.3).

Compte tenu de ce qui précède, il apparaît que la Constitution protège déjà le droit à la dignité humaine et à l'autonomie individuelle vis-à-vis de l'État.

Dans le secteur privé, il existe également des règles en la matière. A priori, le droit civil suisse, avec ses clauses générales ouvertes, est bien placé pour faire face aux problèmes qui se posent. Tel est le cas de l'art. 28 CC sur la protection de la personnalité. La question de savoir si ces règles devraient être précisées dans certains cas pourra être approfondie en cas de ratification de la convention.

4.3.2.3 Article 8 – Transparence et contrôle

L'art. 8 prévoit l'obligation pour les Parties d'adopter ou maintenir des mesures pour veiller à ce que les exigences de transparence et de contrôle adaptées aux contextes et aux risques spécifiques sont en place en ce qui concerne les activités au sein du cycle de vie des systèmes d'IA, y compris en ce qui concerne l'identification de contenu généré par des systèmes d'IA.

La disposition couvre deux aspects :

- Le premier concerne la transparence en lien avec les systèmes d'IA. La transparence doit être assurée par des informations appropriées sur le système d'IA, telles que le ou les objectifs, les limitations connues, les hypothèses, les choix techniques effectués lors de la conception, les caractéristiques, les détails des modèles ou les algorithmes sous-jacents, les méthodes d'apprentissage et les processus d'assurance de qualité. En outre, la transparence peut impliquer d'informer, le cas échéant, les personnes concernées ou le grand public des détails des données utilisées pour l'apprentissage du système et la protection des données personnelles, ainsi que sur l'objectif du système et la manière dont il a été conçu et déployé.[81]

  Le principe de transparence vise également à ce que les systèmes d'IA soient compréhensibles et accessibles aux acteurs appropriés de l'IA. Ce principe implique l'explicabilité et l'interprétabilité des systèmes d'IA. L'*explicabilité* renvoie à la capacité de fournir, sous réserve de la faisabilité technique, des explications suffisamment compréhensibles sur les raisons pour lesquelles un système d'IA fournit des informations,

pour la Confédération, Evaluation de l'application et de l'actualité des lignes directrices, 9 décembre 2022, disponible sous www.cnai.swiss > Ressources > Lignes directrices IA (consulté le 26 août 2024).

[81] Rapport explicatif convention sur l'IA (n. 20), N 58.

produit des prédictions, des contenus, des recommandations ou des décisions, ce qui est particulièrement crucial dans des domaines sensibles tels que les soins de santé, la finance et la justice pénale, où il est essentiel de comprendre le raisonnement qui sous-tend les décisions prises ou assistées par un système d'IA.[82]

L'*interprétabilité* renvoie à la capacité de comprendre comment un système d'IA fait ses prédictions ou prend ses décisions. Il s'agit de rendre le fonctionnement interne, la logique et les processus décisionnels des systèmes d'IA compréhensibles et accessibles aux utilisateurs humains, y compris les développeurs, les parties prenantes et les utilisateurs finaux, ainsi qu'aux personnes concernées.[83]

Sous l'angle de l'explicabilité et de l'interprétabilité, les systèmes d'IA présentent des défis particuliers, qui se manifestent avec plus ou moins d'intensité en fonction du type de système d'IA.[84] Si l'algorithme est basé sur des règles, cela signifie que des personnes physiques programment des étapes afin que la machine parvienne à un résultat. Il en résulte qu'un même *input* conduira toujours au même *output*. Afin d'expliquer le résultat, il faudra notamment mettre en évidence les différentes étapes du raisonnement. Certes, ces algorithmes peuvent aussi avoir une certaine complexité, ce qui peut rendre plus difficile la justification de décisions basées sur ceux-ci.

Toute autre est la situation des algorithmes d'apprentissage automatique (« machine learning ») basés sur des données. Ces algorithmes analysent d'énormes quantités de données et identifient des corrélations dans ces données afin de définir eux-mêmes une règle généralisable. De tels algorithmes sont basés sur des probabilités. Il en résulte que la même entrée ne produit pas toujours le même résultat. En outre, l'origine des résultats n'est souvent pas compréhensible. Dans ce cas, comme les résultats ne reposent pas sur des relations de cause à effet, mais sur des corrélations entre les différentes données, il peut être plus difficile d'expliquer quelles données factuelles et quelles bases juridiques ont été déterminantes pour le résultat du système dans le cas particulier. Il est donc très difficile de garantir l'explicabilité et l'interprétabilité de ces systèmes (effet boîte noire).

- Le deuxième aspect visé à l'art. 8 de la convention concerne la nécessité de prévoir des mécanismes de contrôle conçus pour surveiller, évaluer et orienter les activités menées dans le cadre du cycle de vie des systèmes d'IA. Selon le rapport explicatif, le contrôle peut consister en toute sorte de cadres et processus. Sont mentionnés par exemple les cadres de gestion des risques et des impacts, les lignes directrices

---

[82] Rapport explicatif convention sur l'IA (n. 20), N 60.

[83] Rapport explicatif convention sur l'IA (n. 20), N 61.

[84] Voir en lien avec le devoir de motiver des autorités, NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (n. 39), 37 ; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 10 N 23 ; NADJA BRAUN BINDER/LILIANE OBRECHT, Die Begründung von Verfügungen beim Einsatz algorithmischer Systeme, RSJ 2024, 707 ss.

> éthiques, les processus de certification, les outils de détection et d'atténuation des biais, le rôle des autorités compétentes telles que les autorités de supervision sectorielle, les autorités de protection des données, etc.[85] Dans ce contexte, l'obligation de transparence vise aussi à garantir que les systèmes d'IA soient conçus, développés et utilisés de manière à ce que les mécanismes de contrôle humain puissent être utilisés efficacement au cours du cycle de vie des systèmes d'IA.[86]

La disposition fait également référence à la transparence et au contrôle en lien avec l'identification des contenus générés par l'IA. Le but est de diminuer le risque de tromperie et de permettre la distinction entre les contenus authentiques générés par un humain et les contenus générés par l'IA. Les mesures à prendre pourraient inclure des techniques telles que l'étiquetage et le filigrane, qui impliquent généralement l'intégration d'une signature reconnaissable dans les résultats du système d'IA, sous réserve de la disponibilité de ces technologies et de leur efficacité prouvée, de l'état de l'art généralement reconnu et des spécificités des différents types de contenu.[87]

En droit suisse, s'agissant du secteur public et de l'activité de l'État, la transparence peut notamment être comprise comme un élément de l'État de droit (art. 5 Cst.).[88] Le mandat d'information du Conseil fédéral (art. 180, al. 2, Cst.) constitue un autre fondement constitutionnel pertinent.[89]

Au niveau législatif, la LTrans vise à améliorer le contrôle de la population sur l'administration, en permettant à toute personne intéressée de consulter les documents des autorités fédérales. Elle suppose que des documents existent et que l'accès ne soit pas exclu en vertu des exceptions prévues par la LTrans. Le devoir d'information actif du Conseil fédéral est quant à lui réglé notamment à l'art. 10 LOGA.

Il convient en outre de relever que la transparence de l'activité de l'État peut aussi être assurée par le principe de légalité (cf. pour des développements sur ce principe, ch. 4.2.3.1.2).

Le principe de la transparence liée aux activités de l'État est parfois concrétisé dans certains domaines précis. Par exemple, les tribunaux doivent garantir l'accès au dossier aux parties concernées pour qu'elles puissent exercer leur droit d'être entendues (art. 29, al. 2, Cst.). En outre, l'autorité publique a le devoir de motiver ses décisions. Pour les développements en

---

85 Rapport explicatif convention sur l'IA (n. 20), N 63.

86 Rapport explicatif convention sur l'IA (n. 20), N 65.

87 Rapport explicatif convention sur l'IA (n. 20), n. 59.

88 ROLF H. WEBER/SIMON HENSELER, Regulierung von Algorithmen in der EU und in der Schweiz, Zeitschrift für Europarecht 2020, 28 ss, 36.

89 DOMINIQUE HÄNNI, Vers un principe d'intégrité de l'administration publique – La prévention de la corruption en droit administratif, Genève/Zurich/Bâle 2019, 305 N 794.

lien avec ces aspects et les exigences posées par la convention sur l'IA en matière de recours, il est renvoyé au commentaire ci-dessous en lien avec les art. 14 et 15 de la convention (cf. ch. 4.3.3).

La LPD vise aussi à instaurer le principe de transparence, en mettant l'accent sur la transparence lors du traitement de données personnelles. L'art. 6, al. 3, LPD prévoit que les données personnelles ne peuvent être collectées que pour des finalités déterminées et reconnaissables pour la personne concernée et doivent être traitées ultérieurement de manière compatible avec ces finalités. L'art. 12 LPD prévoit l'obligation pour les responsables du traitement et les sous-traitants de tenir chacun un registre de leurs activités de traitement (pour plus de détails, cf. ch. 4.3.2.5).

En outre, le devoir d'information de l'art. 19 LPD prévoit qu'en principe la personne est informée de manière adéquate de la collecte de ses données personnelles. En matière de décisions individuelles automatisées, qui peuvent être prises par des systèmes d'IA, l'art. 21, al. 1, LPD prévoit que le responsable du traitement, à savoir l'exploitant du système, doit informer la personne concernée de toute décision qui est prise exclusivement sur la base d'un traitement de données personnelles automatisé et qui a des effets juridiques pour elle ou l'affecte de manière significative. L'art. 21, al. 4, LPD prévoit notamment qu'une décision individuelle automatisée émanant d'un organe fédéral doit être qualifiée comme telle. La personne concernée a en outre en principe le droit de faire valoir son point de vue et d'exiger que la décision soit revue par une personne physique. L'art. 25, al. 2, let. f, LPD prévoit un droit d'accès correspondant. Ce dernier prescrit que la personne concernée doit au moins, le cas échéant, recevoir des informations sur l'existence d'une décision individuelle automatisée ainsi que sur la logique sur laquelle se base la décision. Ce qui précède ne vaut cependant qu'en cas de décision exclusivement automatisée, et non lorsque les systèmes d'IA sont utilisés comme support – même très important – à la prise de décision.[90] Finalement, l'art. 56 LPD prévoit que le PFPDT tient un registre des activités de traitement des organes fédéraux et que ce registre est publié.

Enfin, certaines ordonnances techniques prévoient ponctuellement des exigences de transparence en lien avec l'utilisation d'algorithmes.[91]

Pour ce qui est du secteur privé, le principe de transparence et de contrôle devrait être aussi mis en œuvre dans les relations entre privés, là où un effet horizontal direct ou indirect des

---

[90] La CJUE semble avoir une interprétation plus large de la notion de décision individuelle automatisée. Elle a en effet récemment considéré que le « scoring » effectué par une société qui procède à un examen de la solvabilité, constitue déjà une telle décision si le destinataire du renseignement se fonde de manière déterminante sur ce résultat pour se prononcer sur une demande d'octroi d'un prêt de la personne concernée (CJUE, affaire C□634/21 du 7 décembre 2023, SCHUFA Holding [Scoring]). Les effets de cette jurisprudence en Suisse mériteront d'être suivis. Il conviendra dans tous les cas de tenir compte des spécificités du droit suisse.

[91] Cf. l'art. 24 al. 1 de l'ordonnance du DFI sur la radioprotection s'appliquant aux accélérateurs de particules utilisés à des fins médicales (RS **814.501.513**) : « Le fournisseur du système de planification des traitements par irradiation doit donner, dans le descriptif technique, des indications précises sur les algorithmes utilisés pour le calcul des répartitions de dose […] ».

droits fondamentaux est reconnu ou devait l'être à l'avenir (cf. ch. 4.2.3.1). À cet égard, la LPD s'applique également aux personnes privées et assure une protection s'agissant de la transparence des traitements (voir en particulier les art. 6, al. 3, 12, 19, 21 et 25, al. 2, let. f, LPD). Pour ce qui est de l'éventuelle obligation de fournir, devant un tribunal, des explications sur le fonctionnement d'un système d'IA utilisé dans le cadre d'une relation juridique en cas de litige, voir ch. 6.3.1.

En conclusion, le principe de transparence et de contrôle est déjà ponctuellement réalisé au sein de l'ordre juridique suisse. L'analyse conclut cependant que le cadre légal actuel ne suffit pas eu égard aux obligations découlant de l'art. 8 de la convention, en tout cas pour le secteur public. Au niveau de l'administration, une possibilité serait notamment l'adoption d'un registre public des systèmes d'IA utilisés par les autorités publiques, avec une obligation d'annonce. Pour lutter contre l'effet boîte noire, des exigences telles une documentation détaillée, des audits et tests réguliers des systèmes d'IA pourraient être prévus. Il s'agit ici de garder le contrôle sur les systèmes en s'assurant que l'on continue à comprendre comment ils fonctionnent ou à défaut que les phénomènes d'hallucination[92] inhérents à l'IA générative puissent être détectés et corrigés.

Dans les deux domaines, public et privé, se pose la question d'un éventuel élargissement des obligations d'information, respectivement des informations à livrer en cas d'exercice du droit d'accès selon la LPD, aux décisions partiellement automatisées.

S'agissant de l'identification des contenus générés par des systèmes d'IA, le droit suisse n'impose en principe pas d'obligation d'étiquetage. L'introduction d'une mesure de ce type pourrait être indiquée afin de permettre l'identification des contenus générés par l'IA.

À noter que les mesures adoptées devront tenir compte des intérêt opposés qui pourraient entrer en collision avec l'intérêt à la divulgation de certaines informations, p. ex. la protection des données de tiers, les secrets commerciaux, ou des intérêts liés à la poursuite d'infractions pénales.

4.3.2.4 Article 9 – Obligation de rendre des comptes et responsabilité

L'art. 9 prévoit l'obligation pour les États d'adopter ou maintenir des mesures pour garantir l'obligation de rendre des comptes et d'assumer la responsabilité pour les impacts négatifs qui résultent des activités menées dans le cadre du cycle de vie des systèmes d'IA.

Selon le rapport explicatif de la convention, cette obligation implique que les personnes, les organisations ou les entités responsables des activités menées dans le cadre du cycle de vie

[92] Les hallucinations se manifestent par des réponses qui ne correspondent en rien à la réalité, cf. Aide-mémoire du 26 avril 2024 de sensibilisation en matière de grands modèles de langage (large language models, LLM) au sein de l'administration fédérale, 2, disponible sous www.cnai.swiss > Services > Autres services > Aide-mémoire pour l'utilisation de l'IA au sein de l'administration fédérale (consulté le 26 août 2024).

des systèmes d'IA doivent répondre des impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit résultant de ces activités. La disposition exige la mise en place de cadres et de mécanismes pour donner effet à cette obligation.[93]

L'art. 9 exige par ailleurs d'établir des lignes de responsabilité claires pour être en mesure de retracer les actions et les décisions jusqu'à des personnes ou des entités spécifiques. Il faut garantir la possibilité d'identifier un impact négatif et d'attribuer les responsabilités de manière appropriée.[94] Il peut s'agir de nouveaux cadres et mécanismes, mais aussi de mesures judiciaires et administratives, régimes de responsabilité civile, pénale et autres, et, dans le secteur public, de procédures administratives et autres permettant de contester les décisions, ou de responsabilité et d'obligations spécifiques imposées aux opérateurs qui participent aux activités du cycle de vie des systèmes d'IA.[95]

Cette obligation est indissociable du principe de transparence et de contrôle (art. 8) puisque la transparence permet indirectement d'expliquer comment les systèmes d'IA fonctionnent et produisent des décisions. Elle est aussi liée à l'obligation de créer un cadre de gestion des risques et des impacts (art. 16), qui prévoit des mesures permettant d'être proactif dans la prévention et l'atténuation des risques et des impacts.[96]

Pour ce qui est du secteur privé, le principe de responsabilité tel que décrit à l'art. 9 devra être mis en œuvre dans les relations entre privés, là où un effet horizontal direct ou indirect des droits fondamentaux est donné ou devait être reconnu à l'avenir (cf. ch. 4.2.3.1). Pour des considérations relatives au droit civil et pénal, en particulier en matière de responsabilité, il est renvoyé aux développements ci-dessous (cf. ch. 6.3 et 6.6).

En droit suisse, les règles instaurant des régimes de responsabilité sont formulées de telle sorte qu'elles trouvent application indépendamment du fait qu'une violation procède d'un système d'IA. La LPD prévoit en outre des mesures d'« accountability » (obligation de rendre compte), indépendamment de l'IA (cf. ch. 4.3.2.6).

Dans le secteur public, les normes en matière de responsabilité de l'État, ainsi que les règles qui encadrent les activités de ses organes, sont applicables, peu importe que l'activité implique l'usage d'un système d'IA. Cependant, il convient de s'assurer que le cadre de règles, normes juridiques et autres mécanismes permette une attribution efficace des responsabilités. L'analyse montre que les mesures prévues en lien avec la mise en œuvre

93 Rapport explicatif convention sur l'IA (n. 20), N 66.

94 Rapport explicatif convention sur l'IA (n. 20), N 68.

95 Rapport explicatif convention sur l'IA (n. 20), N 66.

96 Rapport explicatif convention sur l'IA (n. 20), N 69 s.

des art. 8 (« Transparence et contrôle »), 14 (« Recours »), 15 (« Garanties procédurales ») et 16 (« Cadre de gestion des risques et des impacts »), par rapport auxquelles un besoin d'agir est constaté, concourront à renforcer l'efficacité des normes en vigueur.

Pour les considérations relatives à la responsabilité civile et pénale, il est renvoyé aux développements ci-dessous (cf. ch. 6.3 et 6.6).

4.3.2.5 Article 10 – Égalité et non-discrimination

L'art. 10, par. 1, prévoit que chaque Partie adopte ou maintient des mesures pour garantir le respect de l'égalité, y compris l'égalité de genre, et l'interdiction de la discrimination dans les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle, comme le prévoit le droit international et interne applicable. Cette formulation renvoie spécifiquement à l'ensemble des instruments juridiques internationaux et nationaux pertinents pour les Parties. Ces derniers fournissent ensemble une base juridique solide et des orientations permettant à chaque État de prendre des mesures pour garantir l'égalité de tous et interdire la discrimination dans le contexte des activités menées dans le cadre du cycle de vie des systèmes d'IA.[97]

Le par. 2 prévoit que chaque Partie s'engage à adopter ou à maintenir des mesures qui visent à supprimer les inégalités, afin d'obtenir des résultats impartiaux, justes et équitables, conformément aux obligations nationales et internationales qui lui incombent en matière de droits de l'homme, en ce qui concerne les activités menées dans le cadre du cycle de vie des systèmes d'IA. Selon le rapport explicatif, l'approche requise ici ne doit pas se limiter à exiger qu'une personne ne soit pas traitée de manière moins favorable « sans justification objective et raisonnable » sur la base d'une ou plusieurs caractéristiques protégées qu'elle possède. Les Parties s'engagent au contraire à adopter de nouvelles mesures ou à maintenir les mesures existantes visant à surmonter les inégalités structurelles et historiques.[98]

L'utilisation de l'IA présente un défi majeur en matière d'égalité de traitement et d'interdiction des discriminations.[99] En effet, à mesure que les algorithmes présentent les résultats de calculs définis par des humains et basés sur des données collectées auprès d'humains, de machines ou une combinaison des deux, ils peuvent refléter et traiter les biais humains qui sont incorporés lorsqu'ils sont programmés et lorsqu'ils traitent des données.[100] Ces biais, le plus souvent involontaires[101], peuvent découler des données utilisées (manque de représentativité, données obsolètes, traitement insuffisant des données), de l'algorithme lui-même (quels paramètres sont pris en compte dans le modèle, lesquels ne le sont pas[102]) ou de la manière

---

97 Rapport explicatif convention sur l'IA (n. 20), N 71.

98 Rapport explicatif convention sur l'IA (n. 20), N 77.

99 Rapport défis de l'intelligence artificielle (n. 1), 38 s.

100 Conseil de l'Europe, Commission pour l'égalité de genre (GEC) et Comité directeur sur l'anti-discrimination, la diversité et l'inclusion (CDADI), Etude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination, préparée par IVANA BARTOLETTI/RAPHAËLE XENIDIS, Strasbourg 2023, 17, disponible sous https://rm.coe.int/prems-107623-fra-2530-etude-sur-l-impact-de-ai-a5-web/1680ac99e2 (consulté le 15 août 2024).

101 GRÉGOIRE LOISEAU, Intelligence artificielle et droit des personnes, in : Alexandra Bensamoun/Grégoire Loiseau (éds.), Droit et intelligence artificielle, Paris 2022, 35 ss, 49.

102 La discrimination par proxy est particulièrement pertinente dans le contexte de l'IA. Les proxys (ou substituts) sont des critères en apparence anodins, mais qui peuvent être fortement corrélés avec des critères sensibles. Par exemple, la donnée « 30 ans d'expérience professionnelle » indique que la personne en question doit être âgée de 45 ans au minimum, le code postal permet de formuler des hypothèses sur le statut socioéconomique ou sur l'origine d'un individu, cf. NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 12 N

dont il est appliqué (à quels besoins le système est-il censé répondre et comment est-il utilisé dans la pratique).[103] Ces risques peuvent encore être accrus par les boucles de rétroactions et le manque de transparence dans les processus[104] (sur la question de la transparence voir ch. 4.3.2.3).

Le secteur du droit du travail, et en particulier la phase de recrutement, est souvent cité dans ce contexte en tant que domaine sensible. Les typologies de systèmes d'IA dans ce domaine sont en effet très variées et vont de la « simple » sélection de CV à des programmes qui évaluent les expressions faciales d'un candidat lors d'un entretien vidéo, ou qui prédisent les performances professionnelles sur la base d'un CV.[105] Un exemple fameux dans ce domaine est celui d'Amazon. En 2018, l'agence de presse Reuters a rapporté que l'entreprise avait développé un programme s'appuyant sur l'apprentissage automatique pour repérer les meilleurs CV. Or, ce programme désavantageait systématiquement les CV des femmes, car il reflétait l'écart entre les sexes parmi le personnel recruté au cours des dix dernières années. La suppression de la rubrique du genre à cocher n'y a rien changé, car le programme arrivait à déduire le sexe de la personne sur la base d'autres recoupements.[106]

Des inégalités et des discriminations sont susceptibles de se produire en raison de l'utilisation de l'IA dans tous les domaines. Des exemples existent notamment quant à : la migration; l'accès aux biens et services, aux banques et aux assurances ; l'accès aux prestations sociales[107] ; l'évaluation des risques dans le domaine de la sécurité, de la prévention du crime, du maintien de l'ordre et du système judiciaire ; l'accès aux services publics et administratifs ; l'éducation ; les médias et moteurs de recherche ; ou encore dans le secteur de la santé.[108] Dans ce dernier secteur en particulier, des exemples montrent que les systèmes sont parfois entraînés sur la base de données non représentatives. Cela peut conduire à des mauvais

30 ; AlgorithmWatch/CH, Papier de position : Protection contre la discrimination algorithmique, septembre 2023, 6, disponible sous https://algorithmwatch.ch > Positions > Discrimination par des algorithmes : comment assurer la protection ? (consulté le 15 août 2024).

[103] Rapport explicatif convention sur l'IA (n. 20), N 75 et 76.

[104] Rapport défis de l'intelligence artificielle (n. 1), 38 s. Voir également pour un aperçu étoffé des problématiques : FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung beim Einsatz von Künstlicher Intelligenz (KI), Jusletter IT 4 juillet 2024.

[105] AVI ASHER-SCHAPIRO, AI is taking over job hiring, but can it be racist ?, Reuters 2021, disponible sous https://www.reuters.com/article/business/healthcare-pharmaceuticals/ai-is-taking-over-job-hiring-but-can-it-be-racist-idUSL5N2NF5ZC/ (consulté le 15 août 2024).

[106] JEFFREY DASTIN, Insight – Amazon scraps secret AI recruiting tool that showed bias against women, Reuters 2018, disponible sous https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ (consulté le 15 août 2024).

[107] On citera ici par exemple le scandale des allocations familiales au Pays-Bas. Des critères relevant du profilage racial ont été intégrés dans l'élaboration du système algorithmique utilisé pour déterminer si des demandes d'allocations familiales devaient être considérées comme erronées et potentiellement frauduleuses. Ceci a conduit à ce que des dizaines de milliers de parents et de personnes ayant la charge d'enfants, appartenant pour la plupart à des familles à faibles revenus, soient accusées à tort de fraude par les autorités fiscales néerlandaise, cf. https://www.amnesty.org/fr/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/ (consulté le 15 août 2024).

[108] Conseil de l'Europe, Etude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination (n. 100), 23 ss (consulté le 15 août 2024) ; AlgorithmWatch/CH, Papier de position : Protection contre la discrimination algorithmique (n. 102) (consulté le 15 août 2024).

diagnostics et des discriminations envers les femmes[109], les minorités et les personnes de couleur[110], ou encore les personnes âgées[111].

Il est important de noter qu'un système d'IA bien conçu peut également permettre de venir en aide aux personnes discriminées ou défavorisées. Les systèmes d'IA peuvent être utilisés, par exemple, pour améliorer l'accessibilité aux informations ou aux biens et services existants. Le recours à des systèmes automatisés de traduction dans les langues régionales ou minoritaires pourrait améliorer l'accès à des services essentiels. L'IA pourrait également servir à promouvoir l'égalité dans le secteur policier, par exemple lorsqu'elle est utilisée pour prévenir les risques de violence fondée sur le genre (voir en Espagne le logiciel VioGen). Dans le secteur de la santé, l'IA pourrait être utilisée pour améliorer l'accès aux soins dans les zones défavorisées et renforcer les capacités de diagnostic des groupes traditionnellement sous-représentés.[112] Par ailleurs, les systèmes d'IA peuvent aussi corriger des erreurs humaines (erreurs de rétroaction, informations incomplètes). Dans de tels cas, le système d'IA peut aider à mettre en évidence des inégalités de traitement et offre ainsi la possibilité de les corriger et de lutter contre.[113]

Le droit suisse ne connaît pas de dispositions légales relatives au respect de l'égalité et à la lutte contre les discriminations spécialement dédiées à l'IA. Il convient donc d'apprécier si le cadre réglementaire existant, qui s'applique aussi en cas d'utilisation de l'IA, appréhende de manière suffisante les enjeux découlant de cette technologie.

Les obligations internationales de la Suisse dans le domaine de l'égalité et de la lutte contre la discrimination résultent notamment des deux Pactes des Nations Unies (art. 2, par. 2, Pacte I et art. 2, par. 1, Pacte II, art. 3 des deux Pactes et art. 26 Pacte II) ainsi que de la CEDH (art. 14). Pour différentes raisons, les dispositions précitées n'ont cependant pas de portée indépendante et ne peuvent être invoquées en Suisse qu'en relation avec d'autres dispositions des textes.[114] La Suisse a par ailleurs ratifié plusieurs conventions de l'ONU qui garantissent l'égalité dans des domaines particuliers, telle que la convention sur l'élimination de toutes les formes de discrimination raciale[115], et la convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes.[116] La majorité des dispositions de ces textes

---

[109] ANNE-SOPHIE MORAND, A WEIRD AI system?, Jusletter 20 septembre 2021, 9 N 21 ss.

[110] ANNE-SOPHIE MORAND, A WEIRD AI system? (n. 109), 7 N 16 ss.

[111] Voir JUSTYNA STYPINSKA, AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies, AI & SOCIETY 2023, 665 ss.

[112] Conseil de l'Europe, Etude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination (n. 100), 89.

[113] Universität Zürich, Digital Society Initiative, Positionspapier, Ein Rechtsrahmen für Künstliche Intelligenz, November 2021, 4, disponible sous https://www.dsi.uzh.ch > Forschung > Strategy Lab > 1. DSI Strategy Lab (consulté le 19 août 2024) ; FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung (n. 104), 17.

[114] GIORGIO MALINVERNI et. al., Droit constitutionnel suisse – Volume II (n. 56), N 1103 ss ; s'agissant de l'art. 14 CEDH cf. ATF **123** II 472, consid. 4c.

[115] RS **0.104**.

[116] RS **0.108**.

sont invocables devant le TF.[117] À citer encore la convention relative aux droits des personnes handicapées[118], dont l'art. 5 est directement justiciable[119].

Au niveau de la Constitution, le principe d'égalité est ancré à l'art. 8, al. 1, et l'interdiction de discrimination, à l'art. 8, al. 2. L'al. 3 de cette disposition garantit l'égalité de droit et de fait entre hommes et femmes et l'al. 4 impose une obligation d'éliminer les inégalités qui frappent les personnes handicapées. Le principe d'égalité s'applique dans toutes les relations que l'État peut entretenir. Il est étroitement lié à l'exigence de justice et d'équité. Hormis l'égalité salariale entre hommes et femmes de l'al. 3, le principe d'égalité et l'interdiction de discrimination ne lient pas directement les privés. L'État est toutefois tenu, en vertu de l'art. 35, al. 1, Cst., de les réaliser dans l'ensemble de l'ordre juridique, y compris lorsque cela s'y prête dans les relations entre particuliers (art. 35, al. 3, Cst.) (cf. ch. 4.3.1.1). D'autres dispositions constitutionnelles rappellent le principe d'égalité dans des domaines plus spécifiques. Tel est les cas par ex. de l'art. 27 Cst. qui garantit la libre concurrence, ou l'art. 15 Cst. qui interdit à l'État de prendre parti pour ou contre une religion ou une conviction (principe de neutralité religieuse).[120]

Le législateur suisse a concrétisé le principe d'égalité dans plusieurs textes légaux. Ces derniers s'appliquent comme indiqué également en présence d'un système d'IA :

- La LEg : Elle vise à garantir l'égalité de droit et de fait entre femmes et hommes dans les rapports de travail, dans le secteur public mais aussi privé. Concrètement, elle prévoit notamment l'interdiction de toute discrimination directe et indirecte à raison du sexe, fondée notamment sur l'état civil, la situation familiale ou la grossesse (art. 3, al. 1 et 2). L'interdiction de toute discrimination s'applique notamment à l'embauche, à l'attribution des tâches, à la rémunération et à la résiliation des rapports de travail. Par ailleurs, compte tenu des difficultés à prouver les cas de discrimination, la loi introduit un allègement du fardeau de la preuve (art. 6). Elle prévoit également un droit d'action spécial des organisations lorsqu'un nombre considérable de rapports de travail est susceptible d'être affecté (art. 7).

- La LHand et ses ordonnances d'application : Cette loi vise à prévenir, réduire et éliminer les inégalités qui frappent les personnes handicapées. Elle s'applique en particu-

[117] Pour la première de ces deux conventions voir ATF **123** IV 202, consid. 2 ; ATF **123** II 472, 479, consid. 4d ; pour la seconde, ATF **125** I 21, consid.4a ; ATF **145** I 308, consid. 3.1.

[118] RS **0.109**.

[119] Arrêt du TF 8C_633/2021, consid. 4.2.

[120] GIORGIO MALINVERNI et al., Droit constitutionnel suisse – Volume II (n. 56), N 1100, 1113 s.

lier aux différents types de constructions accessibles au public, ou à l'accès à la formation. Elle s'adresse en priorité à l'État, sous réserve des constructions visées à l'art. 3.

- Il convient également de citer l'art. 261bis CP, introduit dans la foulée de la ratification de la convention sur l'élimination de toutes les formes de discrimination raciale[121] afin de transposer les engagements internationaux dans le droit national. Cet article vise en premier lieu à protéger la dignité humaine des personnes visées par le dénigrement raciste. Il protège en second lieu la paix publique, qui se trouve menacée par des actes pouvant conduire les gens à se dresser les uns contre les autres.[122] Hormis l'al. 5 (refus d'une prestation destinée à l'usage public en raison de l'appartenance raciale, ethnique ou religieuse), le comportement doit avoir lieu en public. Le TF considère assez largement ce qui est public : il s'agit de tous les propos ou comportements qui n'ont pas lieu dans le cadre privé, c'est-à-dire par exemple dans le cercle familial, le cercle des amis, ou dans un environnement de relations personnelles ou de confiance particulière.[123]

- Enfin, on peut encore citer la loi fédérale sur l'analyse génétique humaine[124], dont l'art. 4 prévoit que nul ne doit être discriminé en raison de son patrimoine génétique ou l'art. 2 de l'Accord sur la libre circulation des personnes[125], qui interdit les discriminations fondées sur la nationalité.

Dans les domaines qui échappent à ces textes, les personnes qui subissent une discrimination pourront se fonder, en fonction des circonstances, directement sur la Constitution, ou sur des dispositions protectrices plus générales telles que l'art. 6 de la LTr.[126] En droit suisse, le législateur a en effet fait le choix de réaliser le principe d'égalité de traitement et de non-discrimination entre particuliers par type de discrimination ou dans des domaines spécifiques. En droit privé, certaines normes protègent indirectement contre la discrimination. Par exemple, en cas de discrimination atteignant la personnalité, on peut se fonder sur les art. 28

[121] RS **0.104**.

[122] CR CP II-MIRIAM MAZOU, art. 261bis N 3 et les réf. citées.

[123] ATF **130** IV 111, consid. 5.2.1.

[124] RS **810.12**.

[125] Accord entre la Confédération suisse, d'une part, et la Communauté européenne et ses États membres, d'autre part, sur la libre circulation des personnes du 21 juin 1999, RS **0.142.112.681**.

[126] Voir notamment, SECO, commentaire, art. 2 OLT 3, 302-4, disponible sous https://www.seco.admin.ch > Travail > Loi sur le travail et Ordonnances > Commentaires relatifs à la loi sur le travail et ses ordonnances > Commentaire de l'OLT 3 (consulté le 19 août 2024).

CC[127] et 328 CO[128] pour requérir une cessation de cette discrimination ou une indemnité. Il existe également une protection contre le licenciement discriminatoire, qui entre dans la catégorie des licenciements abusifs motivés par une caractéristique inhérente à sa personnalité ou par l'exercice d'un droit constitutionnel (art. 336, al. 1, let. a et b CO).

Outre les dispositions susmentionnées, il convient également de mentionner la LPD s'agissant de la protection des personnes physiques. Cette loi, bien que ne visant pas, en premier lieu, à lutter contre les discriminations et les inégalités, contient de nombreuses dispositions susceptibles d'offrir une protection indirecte à cet égard dans le secteur privé et le secteur public. Les traitements de données personnelles qui interviennent dans le cycle de vie d'un système d'IA doivent en effet respecter les règles fixées par ce texte. Sous l'angle de la thématique traitée ici, les éléments suivants paraissent pertinents :

- *Données sensibles* : la LPD prévoit une protection accrue lorsque des données sensibles sont traitées, de même qu'en cas de profilage et de profilage à risque élevé (art. 5, let. c, f et g), ce qui peut protéger contre les discriminations. Font notamment partie des données sensibles les données sur les opinions ou les activités religieuses, philosophiques, politiques ou syndicales, les données sur la santé, la sphère intime ou l'origine raciale ou ethnique, les données génétiques et les données sur des mesures d'aide sociale.

- *Exigence de base légale* : en principe les organes fédéraux ne peuvent traiter des données personnelles que si une base légale le permet. La décision de recourir à l'IA incombera ainsi au Conseil fédéral en cas de base légale matérielle et, cas échéant, au Parlement en cas de base légale formelle, à moins qu'il existe déjà une base légale suffisante (voir ch. 4.2.3.1). Cette exigence constitue également une forme de protection indirecte contre les inégalités et les discriminations dans la mesure où l'État est lié par le principe d'égalité dans l'élaboration de la loi. Par ailleurs, s'agissant de l'élaboration de projets législatifs, il convient de rappeler que le droit actuel prévoit aussi une obligation spécifique à la charge du législateur de procéder à une analyse d'impact de la réglementation sur l'égalité femmes-hommes dans certains cas, en plus de l'analyse d'impact de la réglementation (AIR).

- *Principes de proportionnalité, d'exactitude et de protection des données dès la conception et par défaut* (art. 6, al. 2, et 7 LPD) : la mise en œuvre de ces principes peut permettre de minimiser la collecte et le traitement de données, et d'assurer leur exactitude et leur sécurité, voire de les anonymiser, ce qui peut réduire le risque de discriminations fondées sur ces données. Cela peut également permettre d'assurer la

[127] Cf. par exemple : FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper : Diskriminierung, juin 2024, 7, disponible sous https://www.itsl.uzh.ch > Wissenstransfer > Publikationen > Positionspapiere > Nachvollziehbare Algorithmen : ein Rechtsrahmen für den Einsatz von künstlicher Intelligenz (consulté le 19 août 2024); ELEONOR KLEBER, La discrimination multiple – Etude de droit international, suisse et européen, Zurich 2015, 210 et les réf. citées.

[128] CR CO I-LEMPEN, art. 328 N 1 ; BSK OR I-PORTMANN/RUDOLPH, art. 328 N 45 s. et art. 320 N 2a pour la phase des pourparlers précontractuels.

transparence des traitements de données et permettre aux personnes concernées d'exercer leurs droits, et par là de constater d'éventuelles discriminations.

- *Obligation d'effectuer une analyse d'impact* : lorsque le traitement envisagé est susceptible d'entraîner un risque élevé pour la personnalité ou les droits fondamentaux, le responsable du traitement doit procéder à une analyse d'impact relative à la protection des données (art. 22 LPD). Dans l'évaluation du risque, il doit tenir compte des possibles dommages ou conséquences pour la personne concernée liées à une éventuelle violation de ses droits de la personnalité et de son droit à l'autodétermination informationnelle. Les possibles dommages englobent aussi par exemple des refus de crédits ou un refus d'embauche.[129] Cela peut permettre dans certains cas de détecter de possibles discriminations. L'analyse d'impact contient une description du traitement envisagé, une évaluation des risques pour la personnalité ou les droits fondamentaux de la personne concernée, ainsi que les mesures prévues pour protéger sa personnalité et ses droits fondamentaux. À noter que dans le secteur privé il existe des exemptions de réaliser des études d'impact (art. 22, al. 4 et 5, LPD).

- *Devoir d'information en général et devoir d'information en cas de décisions automatisées et droit d'accès* (art. 19, 21 et 25, al. 2, let. f, LPD) : Le responsable du traitement doit sauf exception informer la personne concernée de la collecte de ses données. Par ailleurs, il doit en principe l'informer de toute décision prise exclusivement sur la base d'un traitement de données personnelles automatisé ayant des effets juridiques pour elle ou l'affectant significativement. La personne concernée a en outre en principe le droit de faire valoir son point de vue et d'exiger que la décision soit revue par une personne physique. Ces obligations ne s'appliquent pas lorsque la décision est en relation avec la conclusion ou l'exécution directe d'un contrat et que la demande de la personne est satisfaite, ou si elle a expressément consenti à ce que la décision soit prise de manière automatisée. Dans le cadre du droit d'accès, la personne concernée reçoit les informations nécessaires pour qu'elle puisse faire valoir ses droits selon la LPD et pour que la transparence du traitement soit garantie. Le cas échéant, elle doit être informée de l'existence d'une décision individuelle automatisée ainsi que de la logique sur laquelle se base la décision (art. 25, al. 2, LPD). Ces dispositions peuvent permettre de mettre en lumière des discriminations. Ce qui précède ne vaut cependant qu'en cas de décision exclusivement automatisée, et non lorsque les systèmes d'IA sont utilisés comme support – même très important – à la prise de décision.[130]

---

[129] CR LPD-PHILIPPE GILLIÉRON, art. 22 N 24 ; DAVID ROSENTHAL/SAMIRA STUDER/ALEXANDRE LOMBARD (pour la traduction), La nouvelle loi sur la protection des données, Jusletter 16 novembre 2020, 58, N 149.

[130] La CJUE semble avoir une interprétation plus large de la notion de décision individuelle automatisée. Elle a en effet récemment considéré que le « scoring » effectué par une société qui procède à un examen de la solvabilité, constitue déjà une telle décision si le destinataire du renseignement se fonde de manière déterminante sur ce résultat pour se prononcer sur une demande d'octroi d'un prêt de la personne concernée

- *Devoirs de documentation et de journalisation* : en vertu des art. 12 LPD et 3, 4 et 5 OPDo, le responsable du traitement a une obligation de tenir un registre des activités de traitements et de journalisation dans certaines hypothèses. Il doit également prendre des mesures techniques et organisationnelles pour assurer la confidentialité des données, leur disponibilité, leur intégrité et leur traçabilité. Ces mesures pourraient donc permettre d'assurer un contrôle sur les traitements effectués[131], et mettre en lumière des irrégularités susceptibles de causer des discriminations.

- *Principe d'exactitude* : en vertu de l'art. 6, al. 5, LPD, celui qui traite des données personnelles doit s'assurer qu'elles sont exactes. Il prend toute mesure appropriée permettant de rectifier, d'effacer ou de détruire les données inexactes ou incomplètes au regard des finalités pour lesquelles elles sont collectées ou traitées.

- *Principe de la bonne foi* : selon certains auteurs, un traitement de données personnelles conduisant à des discriminations pourrait être qualifié de violation du principe de la bonne foi en matière de protection des données (art. 6, al. 2, LPD). Le fait que le but de la LPD soit de protéger les droits fondamentaux et la personnalité des personnes concernées (art. 1 LPD) plaiderait en faveur d'une telle interprétation.[132]

Nonobstant ce qui précède, plusieurs auteurs estiment qu'en l'état le droit suisse anti-discrimination est lacunaire sur certains aspects, que l'on recourt ou non à l'IA.[133] Des recommandations visant à renforcer la législation en la matière avaient par ailleurs été émises à l'époque par le Centre suisse de compétence pour les droits humains (CSDH) dans le cadre d'une étude mandatée par le Conseil fédéral suite à l'adoption du postulat Naef 12.3543.[134] Ces recommandations n'ont pour la plupart pas été suivies par le Conseil fédéral, qui a notamment renoncé à inscrire la protection contre la discrimination dans une norme de droit privé explicite qui viendrait compléter les art. 27 ss CC, à alléger de manière générale le fardeau de la preuve en cas de discrimination ou à étendre la portée de l'art. 261bis CP.[135] Le Conseil fédéral a choisi de privilégier une approche par domaine.

Dans le cadre des réflexions en lien avec la discrimination algorithmique, on tend à conclure que le recours à l'IA exacerbe les faiblesses déjà constatées. Certaines voix se positionnent ainsi en faveur de l'élaboration d'une loi générale sur l'égalité de traitement, respectivement

(CJUE, affaire C□634/21 du 7 décembre 2023, SCHUFA Holding [Scoring]). Les effets de cette jurisprudence en Suisse mériteront d'être suivis. Il conviendra dans tous les cas de tenir compte des spécificités du droit suisse.

131 YVES POULLET, L'IA un défi pour nos législations vie privée, in : Astrid Epiney/Sophia Rovelli (éds.) : Künstliche Intelligenz und Datenschutz / L'intelligence artificielle et la protection des données, Fribourg 2021, 1 ss, 40, en lien avec le principe de responsabilité prévu par le RGPD.

132 En ce sens : FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper : Diskriminierung (n. 127), 7.

133 Notamment CR Cst.-VINCENT MARTENET, art. 8, N 141 ss et les réf. citées ; PREVITALI ADRIANO/MICHAEL MONTAVON, L'interdiction des discriminations, in : Oliver Diggelmann/Maya Hertig Randall/Benjamin Schindler (éds.), Verfassungsrecht der Schweiz Bd. II / Droit constitutionnel suisse Vol. II, 2020, 1453 ss ; SAMANTHA BESSON, L'égalité horizontale : l'égalité de traitement entre particuliers, Fribourg 1999, 1352 ss.

134 CENTRE SUISSE DE COMPÉTENCE POUR LES DROITS HUMAINS, Accès à la justice en cas de discrimination, Rapport de synthèse, Berne 2015, disponible sous www.ofj.admin.ch > Société > Egalité des genres et protection contre la discrimination > Protection contre la discrimination > Etudes et rapports > Postulat Naef 12.3543 (consulté le 19 août 2024).

135 Rapport du Conseil fédéral du 25 mai 2016 en réponse au postulat Naef 12.3543, Le droit à la protection contre la discrimination, en particulier les ch. 4.2.1 et 4.3.1, 4.2.3, 4.2.6, disponible sous www.ofj.admin.ch > Société > Egalité des genres et protection contre la discrimination > Protection contre la discrimination > Etudes et rapports > Postulat Naef 12.3543 (consulté le 27 août 2024).

anti-discrimination[136], de l'introduction d'un principe dans la LPD qui interdirait les traitements de données conduisant à une discrimination[137], l'instauration d'un principe de *Non-Discrimination by Design*[138] ou encore d'un renversement, ou d'un allégement du fardeau de la preuve.[139]

En conclusion, la problématique de l'égalité et de la non-discrimination en lien avec l'utilisation de l'IA est appréhendée par la législation actuelle dans certains de ses aspects.

Néanmoins, on peut se demander si le cadre légal actuel doit être complété ou non. On peut constater à cet égard que les critiques et revendications formulées à l'égard de l'IA concernant la capacité de la législation actuelle à répondre aux problèmes posés ne sont pas nouvelles. Elles correspondent aux revendications déjà émises de manière générale.

Toutefois, l'IA est susceptible d'exacerber et d'accélérer les problématiques déjà constatées, et de leur donner une dimension collective. La protection contre les discriminations ne devrait en principe pas être différente selon que la discrimination est le fait d'un système d'IA ou non. En cas de volonté politique d'agir, tout renforcement de la législation anti-discrimination de manière générale, c'est-à-dire sans égard à la technologie utilisée, sera bénéfique dans le contexte de l'IA.

Indépendamment de cette question, une intervention du législateur paraît nécessaire en matière de renforcement de la transparence et d'analyse d'impact en lien avec des possibles discriminations, dans la mesure où il demeure difficile de reconnaître et de prouver les biais dans le contexte des systèmes d'IA. On pourrait imaginer les interventions suivantes :

- Obligation d'appréhender les aspects d'égalité et de non-discrimination lors de l'analyse des risques et des impacts des systèmes d'IA (voir ch. 4.3.4, art. 16 de la convention).

- Devoir d'information en cas de décision individuelle automatisée : actuellement ce devoir n'existe qu'en cas de décision fondée exclusivement sur la base d'un traitement de données personnelles automatisé (cf. art. 21 et 25 LPD). Or, bien souvent les systèmes d'IA sont utilisés comme support ou aide à la décision uniquement et échappent à cette disposition. Une extension du devoir d'information aux décisions partiellement automatisées pourrait donc être bénéfique.

---

[136] FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper : Diskriminierung (n. 127), 6 ; AlgorithmWatch/CH, Papier de position : Protection contre la discrimination algorithmique (n. 102), 8.

[137] FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper : Diskriminierung (n. 127), 6.

[138] FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA/WEINER/CHRISTOPH HEITZ, Diskriminierung (n. 104), 26 s.

[139] En ce sens : FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper : Diskriminierung (n. 127), 6 s.

- Annonce des systèmes d'IA utilisés dans le secteur public : actuellement le CNAI tient un registre des systèmes d'IA utilisés dans le secteur public. Les annonces se font pour le moment sur une base volontaire. L'annonce pourrait devenir obligatoire.

- Lutte contre l'effet boîte noire : documentation détaillée, audits et tests réguliers des systèmes d'IA. Il s'agit ici de garder le contrôle sur les systèmes en s'assurant que l'on continue à comprendre comment ils fonctionnent ou que l'on puisse détecter des hallucinations. Les mesures adoptées devront cependant tenir compte des intérêt opposés qui pourraient entrer en collision avec l'intérêt à la divulgation de certaines informations, par ex. la protection des données de tiers, les secrets commerciaux, ou des intérêts liés à la poursuite d'infractions pénales.

4.3.2.6 Article 11 – Respect de la vie privée et protection des données à caractère personnel

L'art. 11 de la convention prévoit que chaque Partie adopte ou maintient des mesures pour garantir que les droits à la vie privée des personnes et les données personnelles sont protégés, notamment par des lois, normes et cadres nationaux et internationaux applicables. Cet article prévoit également que des garanties et des protections effectives doivent avoir été mises en place conformément aux obligations juridiques nationales et internationales applicables.

Selon le rapport explicatif, la protection de la vie privée et des données à caractère personnel est un principe commun nécessaire au respect de la quasi-totalité des autres principes de la convention. La collecte et le traitement de données personnelles sont omniprésents dans le domaine de l'IA et les technologies et systèmes automatisés utilisés ont un impact direct sur la vie des personnes. Les systèmes d'IA étant principalement axés sur les données, en l'absence de garanties appropriées, les activités entrant dans le cadre du cycle de vie de ces systèmes pourraient présenter de graves risques dans ce domaine.[140]

Des défis existent à tous les stades du cycle de vie des systèmes d'IA, en particulier en cas d'IA générative. Les problématiques sont différentes selon que l'on se trouve dans la phase de conception, d'entraînement, d'utilisation ou de feed-back du système.[141]

Il convient d'apprécier si le cadre réglementaire existant en droit interne, qui s'applique aussi en matière d'IA, appréhende de manière suffisante les enjeux découlant de cette technologie.

---

[140] Rapport explicatif convention sur l'IA (n. 20), N. 79.

[141] SAMUEL KLAUS, KI trifft Datenschutz – Risiken und Lösungsansätze, in : Astrid Epiney/Sophia Rovelli (éds.) : Künstliche Intelligenz und Datenschutz (n. 131), 81 ss, 83 s. ; DAVID ROSENTHAL, Datenschutz und KI : Worauf in der Praxis zu achten ist, Jusletter 26 avril 2022, 3 ; voir également European Data Protection Board, Report of the work undertaken by the ChatGPT Taskforce, 23 mai 2024, 4 N 14, disponible sous https://www.edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-chatgpt-taskforce_en (consulté le 19 août 2024).

Au plan constitutionnel, les droits fondamentaux à la vie privée et à la protection des données sont protégés par l'art. 10, al. 2, Cst. (liberté personnelle) ainsi que, plus spécifiquement, par l'art. 13 Cst. (dans une moindre mesure par l'art. 7 Cst.).[142] L'art. 13 prévoit, d'une part, que toute personne a droit au respect de sa vie privée et familiale, de son domicile, de sa correspondance et des relations qu'elle établit par la poste et les télécommunications (al. 1) et, d'autre part, que toute personne a le droit d'être protégée contre l'emploi abusif des données qui la concernent (al. 2). Ce second alinéa consacre le droit à l'autodétermination en matière informationnelle, à savoir le droit pour la personne concernée de pouvoir déterminer elle-même si et dans quels buts des informations à son sujet peuvent être traitées.[143]

Au niveau législatif, l'art. 13 Cst. est principalement garanti s'agissant des personnes physiques par la LPD et ses ordonnances d'application, l'OPDo et l'OCPD. Cette législation s'applique dans le secteur privé et dans le secteur public (fédéral). La législation en matière de protection des données a été refondue récemment et est entrée en vigueur le 1er septembre 2023. De nombreuses lois fédérales spéciales viennent compléter ce cadre juridique dans des secteurs particuliers. On relèvera ici la loi sur les télécommunications[144], qui concrétise le secret des télécommunications, et les loi énumérées l'art. 1 de la loi fédérale sur la surveillance de la correspondance par poste et télécommunication, qui déterminent les conditions auxquelles une surveillance peut être opérée.[145] D'autres dispositions, telles que les art. 28 ss CC, qui protègent notamment le droit à la vie privée et le droit à l'image[146], ou les art. 179 ss CP (infractions contre le domaine secret ou le domaine privé) sont aussi pertinentes.

La LPD et ses ordonnances s'appliquent aux systèmes d'IA lorsque ceux-ci utilisent des données personnelles, ce qui est très souvent le cas.[147] Les responsables du traitement et dans certains cas les sous-traitants devront ainsi se soumettre à de nombreuses obligations prévues par ces textes. Il s'agit en particulier du respect des principes généraux (licéité, reconnaissabilité, finalité, exactitude, proportionnalité, bonne foi, protection des données dès la conception et par défaut, sécurité ; cf. art. 6, 7 et 8 LPD), des obligations en matière d'annonce des traitements automatisés et registres (art. 12 LPD et art. 31 OPDo), des obligations en matière de documentation et journalisation (art. 3 à 5 OPDo), des obligations en matière de devoir d'information (art. 19 LPD), spécialement en cas de décision automatisée (art. 21 LPD), des informations à donner en cas d'exercice du droit d'accès (art. 25 LPD), du devoir

---

142 CR LPD-BERTIL COTTIER, art. 1 N 19, BSK DSG/BGÖ-MATTHIAS R. SCHÖNBÄCHLER/URS MAURER-LAMBROU/SIMON KUNZ, art. 1 N 5.

143 ATF **145** IV 42, consid. 4.2 ; JULIEN FRANÇAIS, in : Petit commentaire LPD (n. 44), art. 1 N 12.

144 RS **784.10**.

145 RS **780.1**.

146 BSK ZGB I-ANDREAS MEILI, art. 28 N 17.

147 FLORENT THOUVENIN/STEPHANIE VOLZ, White Paper : Datenschutz, juin 2024, 3, disponible sous https://www.itsl.uzh.ch > Wissenstransfer > Publikationen > Positionspapiere > Nachvollziehbare Algorithmen : ein Rechtsrahmen für den Einsatz von künstlicher Intelligenz (consulté le 19 août 2024).

d'effectuer une analyse d'impact (art. 22). Ces différents éléments ont été détaillés sous l'art. 10 de la convention (« Égalité et non-discrimination ») (cf. ch. 4.3.2.5).

Dans le secteur public, les traitements de données personnelles sont en outre soumis au principe de base légale (ch. 4.2.3.1.2). La LPD a concrétisé ce principe. Le recours par l'État à un algorithme de reconnaissance faciale qui traite des données personnelles biométriques (données sensibles) ou qui formule des prédictions basées sur un profilage, requerra ainsi une base légale formelle (art. 34, al. 2, let. a et b, LPD). En outre, le simple fait de recourir à l'IA pourrait aussi être soumis à cette condition si la finalité du traitement ou le mode de traitement des données personnelles est susceptible de porter gravement atteinte aux droits fondamentaux de la personne concernée (art. 34, al. 2, let. c, LPD).[148] L'approche de la LPD est fondée sur le « risque ».

En cas de violation non justifiée des obligations ci-dessus, la personne concernée pourra notamment exiger du responsable du traitement qu'il cesse les traitements, rectifie des données inexactes ou efface les données (art. 30 et 41 LPD). Elle pourra également dénoncer son cas au Préposé fédéral à la protection des données et à la transparence (PFPDT), qui pourra interdire les traitements de données contraires à la LPD. Le PFPDT peut aussi intervenir d'office.

Cependant, on notera ici que bien souvent le développement de systèmes d'IA entre en contradiction avec les principes généraux de la LPD. En effet, la collecte de masse de données personnelles issues de nombreuses sources pour l'entraînement des systèmes d'IA s'accorde mal avec les principes de minimisation des données, de reconnaissabilité, d'exactitude ou de finalité.[149] La question de savoir si le responsable du traitement privé peut dans ce cas invoquer un intérêt prépondérant, en particulier le motif justificatif lié au traitement de données à des fins ne se rapportant pas à des personnes, notamment dans le cadre de la recherche, de la planification ou de la statistique (cf. art. 31, al. 2, let. e LPD), devra être résolue de cas en cas, en fonction notamment des intérêts en présence et des types de système d'IA.[150] Dans le secteur public, on appréciera également de cas en cas l'applicabilité de l'art. 39 LPD.[151]

---

[148] OFJ, Révision totale de la loi fédérale sur la protection des données (LPD) – Aperçu des principales modifications (n. 38), 10 ss, ch. 2.2.1; voir aussi OFJ, Guide de législation en matière de protection des données (n. 44).

[149] Dans le même sens, en lien avec les outils de reconnaissance faciale, cf. NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 16 N 39.

[150] DAVID ROSENTHAL, Datenschutz beim Einsatz generativer künstlicher Intelligenz, Jusletter 6 novembre 2023, 12 N 34 ss ; ROLF H. WEBER, Künstliche Intelligenz und Datenschutz, Jusletter IT 4 juillet 2024, 7 N 19 ss.

[151] À noter que les problématiques liées à la réutilisation des données sont examinées dans le cadre des travaux en lien avec la réalisation de la motion 22.3890 « Elaboration d'une loi-cadre sur la réutilisation des données » du 22 août 2022.

Relevons encore que, dans le cadre de la révision totale de la LPD, le CP a été enrichi d'une disposition sur l'usurpation d'identité (art. 179^{decies} CP), qui vient s'ajouter aux nombreuses dispositions existantes protégeant la sphère privée et les données. Cette disposition est certainement amenée à jouer un rôle grandissant avec l'IA et les *deep fakes* notamment (cf. ch. 6.6.1).

Enfin, les lignes directrices sur l'IA de la Confédération prévoient que les technologies d'IA utilisées par la Confédération doivent être conçues de telle sorte que la sphère privée soit protégée selon les standards et que les dispositions sur la protection des données soient respectées en tout temps.[152]

En conclusion, l'IA pose de nombreux défis pour la sphère privée et la protection des données que le cadre législatif en vigueur permet toutefois d'appréhender de manière relativement complète.

Néanmoins, une intervention du législateur n'est pas exclue sous l'angle de la transparence pour permettre une mise en œuvre plus efficace des garanties en vigueur :

- Devoir d'information en cas de décision individuelle automatisée : actuellement ce devoir n'existe qu'en cas de décision fondée exclusivement sur la base d'un traitement de données automatisé. Or, bien souvent les systèmes d'IA sont utilisés comme support ou aide à la décision uniquement et échappent à cette disposition.

- Annonce des systèmes d'IA utilisés : actuellement le CNAI tient un registre des systèmes d'IA utilisés dans le secteur public. Les annonces se font pour le moment sur une base volontaire. L'annonce pourrait devenir obligatoire.

- Lutte contre l'effet boîte noire : documentation détaillée, audits et tests réguliers des systèmes d'IA. Il s'agit ici de garder le contrôle sur les systèmes en s'assurant que l'on continue à comprendre comment ils fonctionnent ou que l'on puisse détecter des hallucinations. Comme déjà relevé, les mesures adoptées devront tenir compte des intérêt opposés qui pourraient entrer en collision avec l'intérêt à la divulgation de certaines informations, p. ex. la protection des données de tiers, les secrets commerciaux, ou des intérêts liés à la poursuite d'infractions pénales.

- En outre, le législateur devra régler la coordination entre une éventuelle régulation de l'intelligence artificielle et la LPD, en particulier concernant l'obligation d'effectuer une analyse d'impact.

[152] Intelligence artificielle – lignes directrices pour la Confédération (n. 80), 3.

4.3.2.7 Article 12 – Fiabilité

L'art. 12 prévoit que chaque Partie prend des mesures appropriées pour promouvoir la fiabilité des systèmes d'IA et la confiance en leurs résultats. Cela pourrait comprendre des exigences en matière de qualité et de sécurité adéquates tout au long du cycle de vie des systèmes d'IA.

Selon le rapport explicatif, cette disposition met en évidence le rôle important que peuvent jouer notamment les normes, les spécifications techniques, ou les techniques d'assurance pour contrôler et vérifier la fiabilité des systèmes d'IA et pour documenter et communiquer de manière transparente les éléments de preuve de ce processus. Elle met l'accent sur des éléments clés spécifiques du fonctionnement des système d'IA, tels que la fiabilité, la robustesse, la sûreté, la sécurité, l'exactitude et la performance, ainsi que sur des conditions fonctionnelles préalables, telles que la qualité, l'intégrité et la sécurité des données et la cybersécurité. Il s'agit avec ces normes techniques de créer une confiance justifiée du public. Pour autant que leur processus d'élaboration soit transparent et inclusif, ces normes peuvent en effet contribuer à établir une assurance et une conformité mutuellement comprises et évolutives en matière d'IA.[153]

Les mesures à adopter en vertu de cette disposition devraient viser à garantir que, comme tout autre système logiciel, les systèmes d'IA soient « sécurisés et sûrs dès la conception ». La sécurité et la sûreté devraient être considérés comme des exigences fondamentales et non comme de simples caractéristiques, à préserver tout au long du cycle de vie des systèmes d'IA. Les mesures peuvent selon les cas inclure la fourniture d'informations claires sur la manière dont les acteurs de l'IA ont mis en œuvre ces normes en pratique. Cela implique de garantir la responsabilité de bout en bout grâce à une transparence suffisante des processus et à des protocoles de documentation. Il existe un lien évident entre cette disposition et l'art. 8 (« Transparence et contrôle ») et avec l'art. 9 (« Obligation de rendre des comptes et responsabilité »). Les normes techniques jouent un rôle important dans l'évaluation et l'atténuation des risques lorsque les règles existantes ne donnent pas suffisamment d'indications. Cela est en particulier pertinent pour le cadre de gestion des risques et des impacts.[154]

En droit suisse, la législation en matière de protection des données prévoit des obligations relativement vastes en matière de sécurité des données personnelles. En effet, le respect des principes énoncés à l'art. 6 LPD, les principes de protection des données dès la conception et par défaut (art. 7 LPD), de même que les règles en matière de sécurité (art. 8 LPD, complété par les art. 1 à 6 OPDo), de certification (art. 13 LPD) et d'annonce des violations des données (art. 24 LPD), permettent d'assurer une protection cohérente. Pour ce qui est de la certification en particulier, l'OCPD (art. 6) prévoit que le PFPDT émet des directives sur les exigences minimales qu'un système de gestion doit remplir en tenant compte notamment des normes techniques suivantes : SN EN ISO 9001, SN EN ISO/IEC 27001 et SN EN ISO/IEC 2770.

---

[153] Rapport explicatif convention sur l'IA (n. 20), N 84 ss.

[154] Rapport explicatif convention sur l'IA (n. 20), N 87 ss.

Par ailleurs la loi fédérale sur la sécurité de l'information[155], et ses ordonnances[156] entrées en vigueur le 1er janvier 2024, prévoient des règles pour garantir la sécurité du traitement des informations relevant de la compétence de la Confédération et la sécurité de ses moyens informatique (art. 1, al. 1, LSI). Cette loi met l'accent sur la standardisation des mesures. Ce texte peut également contribuer à une protection des systèmes d'IA dans ce cadre, susceptible de fonctionner sans données personnelles.[157]

Les lignes directrices sur l'IA de la Confédération indiquent au surplus que les systèmes d'IA doivent, dès leur conception, être sûrs, robustes et résilients afin de déployer des effets positifs sur les êtres humains et l'environnement, et à ne pas pouvoir être détournés à des fins abusives ni être utilisés de manière erronée. Des mesures appropriées doivent être prévues afin d'éviter les erreurs de décision graves.[158]

Une question qui se pose dans ce contexte est celle du rôle des normes harmonisées s'agissant de l'évaluation de la conformité d'un système d'IA par rapport au cadre légal. Comme il sera exposé ci-dessous, le règlement sur l'IA de l'UE prévoit, à certaines conditions, une présomption de conformité au règlement en cas de respect de normes harmonisées (cf. ch. 5.2.11).

En conclusion, la Suisse applique déjà certains standards en matière de protection de données et de sécurité de l'information, qui, en présence d'un système d'IA, permettent d'améliorer sa fiabilité, ainsi que ses conditions fonctionnelles préalables, telles que la qualité, l'intégrité, la sécurité des données ainsi que la cybersécurité.

En cas de ratification de la convention sur l'IA, la Suisse devra poursuivre ses prises de contacts et collaborations avec les organismes de normalisation en vue du développement de normes et standards en matière d'IA.

Actuellement, le droit suisse ne prévoit pas de mécanisme permettant de déduire du respect de certaines normes standardisées en matière d'IA la conformité au cadre légal applicable. Dans l'hypothèse d'une réglementation suisse sur l'IA, la question de la portée juridique des standards ne manquera pas de se poser.

---

[155] Loi fédérale sur la sécurité de l'information au sein de la Confédération du 18 novembre 2020 (RS **128**).

[156] Ordonnance sur la sécurité de l'information dans l'administration fédérale et l'armée du 8 novembre 2023 (RS **128.1**) ; ordonnance sur la procédure de sécurité relative aux entreprises du 8 novembre 2023 (RS **128.41**) ; ordonnance sur les systèmes de gestion des données d'identification et les services d'annuaires de la Confédération du 19 octobre 2016 (RS **172.010.59**).

[157] Universität Zürich, Digital Society Initiative, Positionspapier, Ein Rechtsrahmen für Künstliche Intelligenz (n. 113), 6, qui demandait à l'époque d'examiner la nécessité d'introduire une loi générale sur la sécurité informatique.

[158] Intelligence artificielle – lignes directrices pour la Confédération (n. 80), 5.

4.3.2.8 Article 13 – Innovation sûre

L'art. 13 prévoit que chaque Partie est appelée, en vue de favoriser l'innovation tout en évitant les impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit, à permettre la mise en place d'environnements contrôlés pour le développement, l'expérimentation et l'essai de systèmes d'IA sous la surveillance des autorités compétentes.

Il s'agit d'inciter dès les premiers stades de la conception des systèmes d'IA à prendre en compte de questions comme la qualité du système et des données, de la sécurité et de la sûreté et des préoccupations en matière de droits de l'homme. Cela est d'autant plus important que certains risques ne peuvent être identifiés qu'après des tests, et/ou ne peuvent être traités efficacement qu'au stade de la conception.[159]

La convention laisse aux Parties le soin de définir les modalités de mise en œuvre de la disposition. Selon le rapport explicatif, il peut notamment s'agir de « bacs à sable réglementaires » (« *regulatory sandboxes* »), d'orientations réglementaires informelles ou de lettres de non-intervention[160] pour clarifier comment les régulateurs aborderont les différents cycles de vie de l'IA.[161]

La notion de « bac à sable réglementaire » est utilisée pour couvrir une très large palette d'instruments, dont les caractéristiques se mélangent très souvent. L'analyse se rallie ici à la catégorisation de l'étude commandée par le SECO en 2022 sur les *regulatory sandboxes*. Selon cette dernière, les bacs à sables réglementaires au sens large englobent deux outils principaux : les « projets pilotes » impliquant une règlementation expérimentale (ci-après projets pilotes) et les bacs à sable au sens strict.[162] Dans les deux cas un cadre est fixé et supervisé par une autorité.

---

[159] Rapport explicatif convention sur l'IA (n. 20), N 90 ss.

[160] Selon le rapport explicatif, les lettre de non-intervention visent à clarifier la manière dont les régulateurs aborderont la conception, le développement ou l'utilisation des systèmes d'IA dans des contextes nouveaux (cf. Rapport explicatif convention sur l'IA (n. 20), N 92). Elle peut aussi consister une promesse des autorités de ne pas prendre de mesures de surveillance tant que les entreprises respectent les accords conclus ou les aides à l'interprétation des lois applicables données par les autorités. Cela donne aux personnes concernées une sécurité juridique et de planification, notamment dans les cas où il n'est pas certain de la manière dont un modèle d'entreprise doit être traité sur le plan juridique. La lettre de non- intervention confirme aux entreprises qu'il n'y aura pas d'application de la loi pendant la période d'essai dans le sandbox (cf. STEPHANIE VOLZ, KI Sandboxen für die Schweiz ?, RSDA 2022, 51 ss, 63 s.).

[161] Rapport explicatif convention sur l'IA (n. 20), N 92.

[162] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN/ANNICK PIETZONKA/GUIDO SAURER, Prüfauftrag zu Regulatory Sandboxes, in SECO (éd.), Grundlagen für die Wirtschaftspolitik, juin 2022, 6.

D'après l'étude précitée, les projets pilotes servent à tester de nouvelles technologies ou de nouveaux processus en situation réelle pendant une durée limitée, en s'écartant de la réglementation existante sur certains aspects spécifiques. Leurs résultats montrent dans quelle mesure la réglementation doit être révisée.[163] Les bacs à sable au sens strict permettent quant à eux aux entreprises d'expérimenter des procédés, des produits et des services sur le marché à l'aide d'exemptions temporaires de certaines réglementations. Les infractions à ces règlementations sont corrigées mais pas sanctionnées. De cette façon, les entreprises peuvent savoir si leurs modèles commerciaux innovants peuvent exister dans le cadre légal en place. Elles peuvent ainsi se rendre compte de ce qui est possible (ou non) avec la réglementation en vigueur.[164] À noter que les essais dans le cadre de bacs à sables paraissent pouvoir se faire en conditions réelles ou dans un cadre simulé et/ou contrôlé.[165]

S'agissant des conditions pour leur développement, l'étude du SECO précitée conclut que les exigences de base légale sont moins strictes pour les bacs à sable. Alors que les projets pilotes requièrent souvent une base légale formelle, les bacs à sables au sens strict peuvent selon les cas se fonder sur une ordonnance, voir sur un changement de pratique d'une autorité.[166]

Une alternative aux bacs à sable au sens large réside dans ce que l'on appelle les « réglementations basées sur le risque ». Dans ce cas-là la réglementation est ajustée proportionnellement au risque pris par les entreprises. Les simplifications réglementaires mises en place sont durables. Les entreprises avec des risques relativement faibles pour les clients et l'économie profitent ainsi de prescriptions simplifiées.[167] Un autre outil important qui s'inscrit aussi dans cette perspective d'innovation sûre, et qui constitue un complément voire une alternative aux bacs à sable au sens strict, sont les Hub d'innovation. Il s'agit de plateformes

---

[163] Voir également OFJ, Guide de législation, 4e édition, 2019, avec mises à jour 2023, N 1044 ss, disponible sous www.ofj.admin.ch > État & Citoyen > Légistique > Instruments de légistique (consulté le 25 juillet 2024).

[164] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN et al., Prüfauftrag zu Regulatory Sandboxes (n. 162), N 2.1.2 s.

[165] Voir aussi l'art. 3 point 55 du règlement sur l'IA.

[166] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN et al., Prüfauftrag zu Regulatory Sandboxes (n. 162), N 2.2.2 et 2.2.3.

[167] Tel est le cas par exemple en Suisse de la règlementation fintech qui permet de dispenser de l'obligation d'autorisation les entreprises gérant des dépôts du public d'un montant maximal d'un million de francs (cf. YVES SCHNEIDER/PATRICK ZENHÄUSERN/PETER HETTICH/ROGER KÜTTEL, Quelles réglementations pour favoriser l'innovation, La Vie économique – Plateforme de politique économique, 23 juin 2022, disponible sous https://dievolkswirtschaft.ch/fr/2022/06/quelles-reglementations-pour-favoriser-linnovation/ [consulté le 25 juillet 2024]).

permettant l'échange d'idées entre autorités, entreprises, universités et investisseurs sur les tendances technologiques et les connaissances spécifiques à un secteur.[168]

La législation suisse contient des dispositions autorisant les projets pilotes dans de nombreux domaines sectoriels. La plupart de ces dispositions sont suffisamment neutres pour s'appliquer aussi à l'IA.

Dans le domaine de la protection des données tout d'abord, l'art. 35 LPD prévoit que le Conseil fédéral peut autoriser les traitements automatisés de données sensibles ou d'autres traitements au sens de l'art. 34, al. 2, let. b et c avant l'entrée en vigueur d'une loi au sens formel si les conditions suivantes sont réunies : 1) les tâches qui nécessitent ce traitement sont réglées dans une loi au sens formel déjà en vigueur; 2) des mesures appropriées sont prises aux fins de réduire au minimum les atteintes aux droits fondamentaux de la personne concernée et 3) la mise en œuvre du traitement rend indispensable une phase d'essai avant l'entrée en vigueur de la loi au sens formel, en particulier pour des raisons techniques. Il s'agit de conditions cumulatives.[169] La procédure d'approbation par le PFPDT et les modalités sont détaillées dans l'OPDo (art. 32 ss). Les essais pilotes permettent de tester le fonctionnement de nouvelles technologies impliquant le traitement de données personnelles[170], donc possiblement aussi des systèmes d'IA.

Un autre exemple de projet pilote susceptible de trouver application en matière d'IA est fourni par l'art. 15 de la loi sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités[171], entrée partiellement en vigueur au 1er janvier 2024. Cette loi vise à promouvoir le traitement électronique des processus de la Confédération (principe de la priorité au numérique). Ces processus comprennent l'interaction des autorités de tous les échelons de l'État entre elles, ainsi que celle de ces autorités avec la population et les entreprises. Pour l'essentiel, la LMETA fixe les conditions générales de plusieurs activités, à savoir le développement de la cyberadministration à l'échelon de la Confédération, la collaboration de la Confédération avec d'autres collectivités et organisations dans le domaine de la cyberadministration et la

---

[168] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN ET AL., Prüfauftrag zu Regulatory Sandboxes (n. 162), N 2.2.4. La Suisse est dotée de nombreux Hub d'innovation qui permettent l'échange d'information et de conseils entres les différents acteurs de l'IA. Au plan fédéral, citons par exemple Innosuisse, le Parc suisse d'innovation, ou le centre IA de l'EPFL. D'autres pôles importants existent au plan cantonal.

[169] BSK DSG/BGÖ-ANDREAS STÖCKLI/CHRISTOPH GRÜNINGER, art. 35 N 8 ; CR LPD-ASTRID EPINEY/SAMAH POSSE, art. 35 N 2.

[170] MONIQUE COSSALI SAUVAIN, in : Petit commentaire LPD (n. 44), art. 35 N 20.

[171] Loi fédérale sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités (LMETA ; RS **172.019**).

fourniture de prestations administratives électroniques.[172] L'art. 15 LMETA prévoit également la possibilité de mener des projets pilotes. Si ces derniers impliquent des traitements tombant dans le champ d'application de l'art. 35 LPD, il ne devrait pas être possible déroger aux conditions imposées par cette disposition.[173]

On trouve d'autres exemples dans le droit sectoriel. Par exemple, l'art. 106, al. 5, LCR, autorise le Conseil fédéral à octroyer des autorisations en vue de projet pilotes dans le domaine des véhicules automatisés. Plus d'une dizaine de projets ont déjà été réalisés sur cette base.[174] Dans le domaine de l'approvisionnement de l'électricité également, des projets pilotes visant le développement de technologies, de modèles d'affaires ou de produits innovants dans le secteur de l'énergie peuvent être autorisés à certaines conditions.[175] On trouve également des projets pilotes dans le domaine de l'assurance-maladie[176] afin de freiner l'augmentation des coûts de la santé et de renforcer la qualité ou de promouvoir la numérisation, dans l'assurance-invalidité, pour faciliter la réadaptation[177] ou encore dans la formation professionnelle.[178]

Au vu de ce qui précède, on peut conclure que le droit fédéral prévoit des dispositions autorisant des projets pilotes susceptibles de s'appliquer à l'IA. Le panorama suisse en matière de pôles d'innovations est également très riche. En revanche, on remarque que les « bacs à sable » au sens strict, qui permettraient à des entreprises de tester des systèmes innovants dans des conditions réelles ou simulées ne sont pas développés. De tels bacs à sables seraient, selon un sondage mené dans le cadre de l'étude commandée par le SECO sur l'examen des *regulatory sandboxes*, particulièrement utiles en lien avec l'IA et la protection des données. En effet, les règles de protection des données, respectivement les incertitudes quant à leur application concrète, empêcheraient le développement d'utilisations potentiellement utiles pour la société.[179]

Compte tenu de ce qui précède, il conviendrait d'examiner plus en détails dans le cadre d'une éventuelle mise en œuvre de la convention sur l'IA si les dispositions existantes en matière de projets pilotes offrent déjà une base suffisante pour développer des « bacs à sable » pour l'IA (le cas échéant de manière sectorielle). On examinera cas échéant comment mettre en place ces outils. Il convient de noter que certains auteurs voient les *sandboxes* comme une

---

[172] Communiqué de presse du 22 novembre 2023, Le Conseil fédéral met en vigueur la loi fédérale sur l'utilisation des moyens électroniques pour l'exécution des tâches des autorités, disponible sous www.admin.ch > Documentation > Communiqués > Le Conseil fédéral met en vigueur la loi fédérale sur l'utilisation des moyens électroniques pour l'exécution des tâches des autorités (consulté le 27 août 2024).

[173] MONIQUE COSSALI SAUVAIN, Petit commentaire LPD (n. 44), art. 35 N 11 ss.

[174] Voir la liste des projets publiés sur le site de l'OFROU, « Mobilité intelligente en Suisse : vue d'ensemble des projets pilotes achevés », état au 3 février 2023, disponible sous www.astra.admin.ch > Thèmes > Mobilité intelligente > Essais pilotes (consulté le 25 juillet 2024).

[175] Art. 23*a* de loi fédérale sur l'approvisionnement en électricité (RS **734.7**).

[176] Art. 59*b* de la loi fédérale sur l'assurance-maladie (RS **832.10**).

[177] Art. 68^{quater} de la loi fédérale sur l'assurance-invalidité (RS **831.20**).

[178] Art. 4 de la loi fédérale sur la formation professionnelle (RS **412.10**).

[179] YVES SCHNEIDER/PETER HETTICH/PATRICK ZENHÄUSERN ET AL., Prüfauftrag zu Regulatory Sandboxes (n. 162), N 3.1. Egalement: STEPHANIE VOLZ, KI Sandboxen für die Schweiz ? (n. 160), 67 s., qui plaide pour l'adoption d'une clause expérimentale dans la LPD.

« mesure technique et organisationnelle » permettant de réduire les atteintes aux droits fondamentaux dans le cadre de projets pilotes menées par la Confédération.[180]

Une source d'inspiration dans ce cadre pourrait être l'Innovation Sandbox[181] lancée en 2022 par le canton de Zurich. Le projet crée un environnement de test dans lequel les acteurs peuvent mettre en œuvre des projets d'IA dans un cadre clairement défini. Différentes organisations telles que des start-ups, des PME, de grandes entreprises ou des instituts de recherche ont accès au savoir-faire réglementaire et à de nouvelles sources de données grâce à la Sandbox. En contrepartie, toutes les connaissances et tous les résultats sont partagés publiquement. Contrairement à de nombreuses approches étrangères, l'Innovation Sandbox pour l'IA va plus loin en ce sens que certains des projets soumis ne sont pas seulement examinés, mais aussi mis en pratique. L'objectif est de faire progresser l'innovation responsable en tenant compte de critères juridiques et éthiques et de soutenir la diffusion de l'IA dans l'administration, l'économie et la recherche.

Le cadre législatif suisse prévoit des possibilités sectorielles en matière de projets pilotes qui peuvent aussi s'appliquer à l'IA. Les pôles d'innovation sont aussi bien présents. En revanche, il semble que la création d'environnements contrôlés permettant aux entreprises de tester leur produits et systèmes dans des conditions réelles (*bac à sables au sens strict*) ne soit pas ou peu développée. Il conviendra d'affiner les besoins en la matière dans le cadre de la mise en œuvre de la convention.

4.3.3 Chapitre IV : Recours

4.3.3.1 Généralités

Le Chapitre IV est consacré aux recours. L'art. 14 contient des prescriptions s'agissant des mesures à prendre afin de garantir un recours accessible et effectif contre des violations des droits de l'homme résultant des systèmes d'IA. L'art. 15 traite plus généralement des garanties de procédure.

4.3.3.2 Article 14 – Recours

Art. 14, par. 1 – Recours accessibles et effectifs

Selon l'art. 14, par. 1, chaque Partie adopte ou maintient, dans la mesure où des voies de recours sont requises par ses obligations internationales et conformément à son système juridique interne, des mesures garantissant la disponibilité de voies de recours accessibles et effectives contre les violations des droits de l'homme résultant des activités menées dans le cadre du cycle de vie des systèmes d'IA.

[180] SANDRA HUSI-STÄMPFLI/ANNE-SOPHIE MORAND, Datenschutzrecht, Zurich 2024, 180 N 336.

[181] Innovation Sandbox for Artificial Intelligence (AI), disponible sous https://www.zh.ch > Wirtschaft & Arbeit > Wirtschaftsstandort > Innovation-Sandbox für KI (consulté le 25 juillet 2024).

La disposition ne requiert pas la création de nouvelles voies de droit. Elle demande cependant aux Parties de renforcer les voies de droit existantes en matière de violation des droits de l'homme avec des mesures permettant d'appréhender les enjeux posés par les systèmes d'IA. Les systèmes d'IA sont en effet d'une grande complexité technique et leur fonctionnement est souvent opaque. Les personnes potentiellement affectées par l'utilisation des systèmes d'IA peuvent donc avoir des difficultés à exercer leurs droits. En particulier, l'accès aux informations pertinentes et leur compréhension sont difficiles pour elles.[182]

C'est la raison pour laquelle l'art. 14 de la convention souligne la nécessité d'avoir des voies de recours à la fois accessibles et effectives. La terminologie est axée notamment sur celle de l'art. 2 du Pacte II de l'ONU et de l'art. 13 CEDH. Pour être effectif, le recours doit permettre de remédier directement aux situations contestées. Pour être accessible, il doit être disponible avec des garanties procédurales suffisantes pour que le recours soit utile à la personne concernée.[183] En général, concernant l'accessibilité pratique d'un recours, il convient par exemple de considérer la possibilité d'avoir accès aux informations nécessaires et à des conseils éclairés. S'agissant de l'accessibilité en droit, il faut notamment que le recours offre des perspectives raisonnables de succès.[184]

Se pose la question de la portée de l'art. 14 dans les relations entre privés. L'art. 14 de la convention s'applique en effet en matière de recours *contre les violations des droits de l'homme* résultant des activités menées dans le cadre du cycle de vie des systèmes d'IA. Les obligations dérivant de l'art. 14 de la convention sur l'IA devraient s'appliquer là où un effet horizontal direct ou indirect des droits fondamentaux entre privés est ou devait être reconnu à l'avenir (cf. les considérations développées sous ch. 4.2.3.1). Dans ce cadre limité, les voies de droit civiles donnent déjà certaines garanties (voir notamment les considérations sous ch. 6.3).

Art. 14, par. 2 – Mesures afin de renforcer la portée de l'art. 14, par. 1

La convention prévoit déjà des obligations connexes pertinentes en matière de recours au Chapitre III, à savoir l'art. 8 (« Transparence et contrôle »), et l'art. 9 (« Obligation de rendre des comptes et responsabilité »). L'art. 14, par. 2, va plus loin et introduit l'obligation d'adopter ou maintenir trois mesures spécifiques permettant de renforcer la portée de l'art. 14, par. 1 :

- La let. a) contient l'obligation de documenter. Plus précisément, la documentation doit porter sur les « informations pertinentes concernant les systèmes d'IA susceptibles d'avoir une incidence significative sur les droits de l'homme et leur utilisation perti-

[182] Rapport explicatif convention sur l'IA (n. 20), N 95 s.

[183] Rapport explicatif convention sur l'IA (n. 20), N 98.

[184] OLIVIER BIGLER, in : Luc Gonin/Olivier Bigler (éds.), Commentaire de la Convention européenne des droits de l'homme (CEDH), Berne 2018, art. 13 N 21 s, et les réf. citées.

nente ». Ces informations doivent être fournies « aux organismes autorisés à avoir accès à ces informations et, si nécessaire et applicable, mises à disposition des personnes concernées ou communiquées à ces dernières ».

La disposition vise à garantir la transparence en lien avec l'usage des systèmes d'IA. Le but ultime de cet article est toutefois de permettre aux personnes concernées de contester les décisions préparées ou prises à l'aide des systèmes d'IA, voire, si approprié, l'utilisation du système (cf. art. 14, par. 2, let. b et son examen ci-dessous).

En droit suisse, les obligations de documenter et de mettre les informations à disposition des personnes concernées sont des corollaires du droit d'être entendu prévu à l'art. 29, al. 2, Cst. En outre, la LTrans permet d'accéder, à certaines conditions, aux documents officiels de l'administration fédérale.

En matière de protection des données, la LPD connaît en outre l'obligation de tenir un registre des activités de traitement des données personnelles (cf. art. 12 LPD). S'agissant de la mise à disposition d'informations aux personnes concernées, il convient de mentionner également les art. 19, 21 et 25 LPD sur le devoir d'information et le droit d'accès (cf. ch. 4.3.2.5 et 4.3.2.6). Les dispositions en vigueur ne paraissent toutefois pas suffisantes eu égard aux obligations prévues à l'art. 14, par. 2, let. a), qui sont plus étendues.

En conclusion, il apparaît que le droit suisse contient déjà des dispositions auxquelles il est possible de rattacher l'obligation prévue à l'art. 14, par. 2, let. a) de la convention. Le cadre juridique en vigueur ne semble toutefois pas suffisant.

À noter que les termes « pertinentes », « significative » et « si applicable » laissent une marge d'appréciation aux Parties dans le cadre de la mise en œuvre. En particulier, seuls les systèmes susceptibles d'avoir une incidence « significative » sur les droits de l'homme sont visés par la disposition, qui n'impose donc pas des exigences applicables à tous les systèmes d'IA tels que définis à l'art. 3 de la convention. Il appartiendra au législateur de déterminer le seuil pertinent.

En outre, le devoir de fournir des informations pertinentes pourrait connaître des limites en fonction d'autres intérêts prépondérants, par exemple liés à la protection des données de tiers, ou encore aux secrets d'affaires.

L'art. 14, par. 2, let. a), nécessite donc de devoir préciser le cadre juridique applicable, si bien qu'une intervention du législateur dans ce sens paraît indiquée. En outre, les informations ne seront généralement pas en mains des autorités, mais des développeurs des systèmes d'IA.

La disposition ne prescrit pas l'obligation de tenir des registres recensant les informations pertinentes sur les systèmes d'IA, mais telle pourrait être une possible forme de mise en œuvre de cette obligation.

- La let. b) requiert que les informations visées à l'art. 14, par. 2, let. a) soient « suffisantes pour permettre aux personnes concernées de contester la ou les décisions

prises par le biais de l'utilisation du système ou fondées en grande partie sur celle-ci, et, si nécessaire et approprié, de contester l'utilisation du système ». Ce dernier cas de figure peut couvrir les situations où, par hypothèse, il serait interdit d'utiliser un certain système. On peut aussi songer aux situations où l'on reproche à l'autorité d'utiliser un système d'IA en l'absence d'une base légale adéquate et que cette utilisation est contestée en tant que telle.

Les systèmes d'IA posent des défis particuliers en matière de devoir de motiver.[185] Ces défis se présentent de manière différente en fonction du type de système d'IA utilisé. Comme expliqué plus haut (cf. *ad* art. 8, ch. 4.3.2.3), les algorithmes basés sur les règles sont moins problématiques s'agissant de leur explicabilité et de l'interprétabilité de leurs résultats que les algorithmes d'apprentissage automatique (« machine learning »). Satisfaire au devoir de motiver en lien avec ces derniers s'avère particulièrement difficile.

En tout état de cause, le devoir de motivation ne saurait porter sur le fonctionnement en général d'un système d'IA, mais sur le *raisonnement sur lequel se base la décision dans le cas d'espèce*. Le contenu des informations transmises doit être adapté au contexte, suffisamment clair et, plus important encore, permettre à la personne concernée d'utiliser effectivement les informations en question dans le cadre de la procédure de recours.[186]

En droit suisse, en matière de décisions administratives et judiciaires, le droit à la motivation découle du droit d'être entendu de l'art. 29, al. 2, Cst. La motivation d'une décision doit porter à la fois sur l'état de fait retenu et le raisonnement juridique suivi. Le TF souligne que « l'essentiel est que la décision indique clairement les faits qui sont établis et les déductions juridiques qui sont tirées de l'état de fait déterminant »[187]. Le droit d'être entendu est concrétisé dans le droit de procédure applicable. Les dispositions de la PA, en particulier l'art. 35 PA, sont notamment déterminantes pour la procédure administrative au niveau fédéral. La motivation doit être adéquate et varie en fonction du cas concret. Ainsi, plus les faits sont complexes, plus la motivation doit être dense. En outre, si l'autorité dispose d'une grande marge de manœuvre, elle devra motiver plus en détail sa décision. Dans les domaines où un grand nombre de décisions similaires sont rendues, il est admis que la motivation peut être d'une moindre

---

[185] Cf. sur les défis dans le secteur public, NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (n. 39), 37 ; NADJA BRAUN BINDER/LILIANE OBRECHT, White Paper : Transparenz durch Begründung von Verfügungen, juin 2024, 3, disponible sous https://www.itsl.uzh.ch > Wissenstransfer > Publikationen > Positionspapiere > Nachvollziehbare Algorithmen : ein Rechtsrahmen für den Einsatz von künstlicher Intelligenz (consulté le 27 août 2024) ; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 10 N 23 ; NADJA BRAUN BINDER/LILIANE OBRECHT, Die Begründung von Verfügungen (n. 84), 707 ss.

[186] Rapport explicatif convention sur l'IA (n. 20), N 99.

[187] ATF **142** II 154, consid. 4.2 ; JACQUES DUBEY, Droits fondamentaux, Volume II : Liberté, garanties de l'État de droit, droits sociaux et politiques, Bâle 2018, 815.

intensité, dans un souci d'économie de procédure. Il convient de relever que le fait de pouvoir motiver de manière adéquate l'action de l'État permet aussi de garantir la sécurité juridique, puisque l'activité étatique devient ainsi prévisible et une pratique uniforme est favorisée. Ainsi, si l'autorité utilise des systèmes d'IA pour prendre des décisions administratives ou judiciaires, les garanties déduites du droit d'être entendu devront être respectées.

Il convient aussi de mentionner dans ce contexte la LPD, qui connaît des dispositions pertinentes applicables en matière de décisions individuelles automatisées (cf. art. 21 et 25, al. 2, let. f, LPD). Il y a décision individuelle automatisée lorsqu'un traitement de données a lieu sans intervention humaine et qu'il en résulte une décision, ou un jugement, à l'égard de la personne concernée qui produit des effets juridiques pour elle ou l'affecte de manière significative. Il peut s'agir d'une décision prise par une entité privée ou un organe public (fédéral). Cette décision doit cependant présenter un certain degré de complexité. Les décisions simples du genre de celles qui sont prises lors d'un retrait au bancomat n'en font pas partie.[188] En droit public, une décision au sens de l'art. 5 PA découlant exclusivement d'un traitement de données personnelles automatisé produit des effets juridiques sur la personne concernée et doit ainsi être qualifiée de décision individuelle automatisée au sens de l'art. 21 LPD (cf. par exemple l'art. 38, al. 2, LD).[189]

Dans le cadre du droit d'accès, l'art. 25, al. 2, let. f, LPD prévoit que la personne concernée doit au moins, le cas échéant, recevoir des informations sur l'existence d'une décision individuelle automatisée ainsi que sur la logique sur laquelle se base la décision. Selon le Message du Conseil fédéral, il n'y a pas lieu de révéler à la personne les algorithmes utilisés, qui relèvent souvent du secret d'affaires, mais plutôt les hypothèses de base qui sous-tendent la logique algorithmique sur laquelle repose la décision individuelle automatisée.[190]

Il sied toutefois de relever que ces dispositions ne s'appliquent que si la décision est prise sur la base d'un traitement de données personnelles *exclusivement* automatisé (cf. art. 21, al. 1, LPD). L'art. 14, par. 2, let. b) de la convention sur l'IA a en revanche une portée plus large, car il inclut également les cas où la décision est fondée « en grande partie » sur l'utilisation du système d'IA. Sont donc visées aussi les situations où les systèmes d'IA sont utilisés en tant qu'aide à la décision, laquelle est ensuite prise par une personne physique. En outre, à la différence de l'art. 21 LPD, l'art. 14,

---

[188] Message du 15 septembre 2017 concernant la loi fédérale sur la révision totale de la loi fédérale sur la protection des données et sur la modification d'autres lois fédérales (LPD), FF **2017** 6565, 6674.

[189] LISA JACCOUD/SÉBASTIEN FANTI/ALEXANDRE STAGER, in : Petit commentaire LPD (n. 44), art. 21 N 22.

[190] Message LPD (n. 188), 6684.

par. 2, let. b) ne se limite pas aux cas de décisions automatisées comprenant un traitement de données personnelles. En effet, bien que la majorité des systèmes d'IA implique le traitement de données personnelles, tel n'est pas toujours le cas.[191]

Par ailleurs, il convient de rappeler que les dispositions de la LPD ne s'appliquent pas aux traitements de données personnelles effectués dans le cadre de procédures devant des tribunaux ou dans le cadre de procédures régies par des dispositions fédérales de procédure. En revanche, la LPD s'applique aux procédures administratives de première instance (art. 1, al. 3, LPD).

En conclusion, le droit suisse contient déjà diverses dispositions pertinentes en matière de devoir de motivation dans le contexte des décisions automatisées, notamment dans le cadre de la LPD. Cependant, celle-ci ne couvre pas toutes les constellations, à savoir en particulier les décisions prises *à l'aide* des systèmes d'IA. Une intervention du législateur paraît nécessaire afin de couvrir cette hypothèse également et de définir les exigences en matière de devoir de motiver dans ce contexte.

En ce qui concerne les décisions administratives de première instance, il va de soi que les exigences du droit d'être entendu doivent être respectées. Cela étant, il paraît judicieux d'approfondir si le cadre légal doit être précisé s'agissant de la manière de garantir le respect de ces exigences dans le contexte des procédures administratives automatisées.

Aucun besoin d'agir ne semble être donné en matière de décisions judiciaires, puisqu'aucune base légale n'autorise ce type de décisions, en l'état actuel des lois de procédure. Si de telles bases légales devaient être introduites, il conviendra d'approfondir les besoins éventuels d'adaptation.

- L'art. 14, par. 2, let. c), prévoit qu'une possibilité effective soit donnée aux personnes concernées de former un recours auprès des autorités compétentes. Selon le rapport explicatif, cela peut inclure les mécanismes de contrôle visés à l'art. 26 de la convention (cf. ch. 4.4.5).[192]

Selon l'analyse de l'OFJ, la disposition ne requiert pas la création de nouvelles voies de droit. Elle demande aux Parties de garantir l'effectivité des voies de droit existantes en matière de violation des droits de l'homme, dans le contexte des activités menées dans le cadre du cycle de vie des systèmes d'IA. Les mesures prévues aux art. 14, par. 2, let. a) et b) de la convention ont précisément ce but. Il est

[191] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (n. 39), 37, nbp 246.

[192] Rapport explicatif de la convention sur l'IA (n. 20), N 100.

donc renvoyé aux développements ci-devant s'agissant d'éventuels besoins d'agir en droit suisse.

4.3.3.3 Article 15 – Garanties procédurales

Art. 15, par. 1 – Garanties protections et droits procéduraux effectifs

Selon l'art. 15, par. 1, chaque Partie veille à ce que, lorsqu'un système d'IA a un impact significatif sur la jouissance des droits de l'homme, les personnes affectées par celui-ci disposent de garanties, de protections et de droits procéduraux effectifs, conformément au droit international et au droit interne applicables.

La disposition s'applique uniquement aux systèmes d'IA qui ont un impact « significatif » sur la jouissance des droits de l'homme, et non pas à tous les systèmes d'IA au sens de l'art. 3 de la convention.

S'agissant de l'application de la disposition au secteur privé, il est renvoyé aux remarques ci-dessus en lien avec l'art. 14 (cf. ch. 4.3.3.2). L'exigence d'un impact « significatif » limitera d'autant plus les situations de droit privé concernées.

La disposition vise les garanties de procédure en général. Elle a pour but d'assurer que les garanties de procédure usuelles aient la même effectivité en cas de recours à l'IA que sans.

S'agissant du droit suisse, il convient en particulier de citer le droit à un procès équitable, garanti aux art. 29 à 32 Cst., et par l'art. 6 CEDH. Les éléments suivants méritent d'être relevés :

- Une personne a le droit, découlant du *droit d'être entendu*, de *faire valoir son point de vue* envers l'autorité, avant que celle-ci ne rende une décision qui la touche. Elle a le droit de participer à l'instruction de la cause, d'alléguer des faits, d'offrir des preuves et de se déterminer sur les preuves qui sont administrées.

  En droit suisse, l'art. 21, al. 1 et 2, LPD prévoit le devoir d'information en cas de décision individuelle automatisée ainsi que la possibilité, pour la personne concernée qui le demande, de faire valoir son point de vue. Des restrictions sont prévues aux art. 21, al. 3 et 4, LPD. En particulier, s'agissant des organes fédéraux, si selon l'art. 30, al. 2, PA, la personne ne doit pas être entendue, les droits découlant de l'art. 21, al. 2, LPD ne s'appliquent pas.

  Comme déjà mentionné, le champ d'application des dispositions de la LPD ne couvre toutefois pas toutes les hypothèses visées par l'art. 15, par. 1, de la convention, notamment le cas de décisions partiellement automatisées. En outre, selon l'art. 1, al. 3, LPD, les traitements de données personnelles effectués dans le cadre de procédures devant des tribunaux ou dans le cadre de procédures régies par des dispositions fédérales de procédure, ainsi que les droits des personnes concernées, obéissent au droit de procédure applicable à l'exclusion de la LPD. Ces autres lois prévoient à leur tour des garanties en matière de droit d'être entendu.

- Le droit de *consulter le dossier* joue aussi un rôle important. Les droits des parties à la procédure restent sans effet si celles-ci n'ont pas connaissance des documents sur lesquels l'autorité fonde sa décision.

  Dans les procédures automatisées, la personne concernée devrait avoir accès aux informations particulières d'un cas traitées par un système d'IA (*input*) et au résultat qui en découle (*output*). De plus, il faut qu'elle puisse savoir comment ces informations sont traitées par la machine. Or, selon les circonstances, ces contenus ne sont toutefois que difficilement compréhensibles pour la personne touchée. Cela étant, le Tribunal fédéral a indiqué, certes dans un contexte différent, que d'éventuelles difficultés factuelles résultant de la quantité de documents à consulter ne portent pas atteinte au droit de consulter le dossier.[193] La limite pourrait toutefois être atteinte en matière de systèmes d'IA, puisque la personne concernée pourrait ne pas être en mesure de tirer de conclusions sur la base des informations données.[194]

- En procédure administrative, le *devoir de l'autorité d'établir les faits d'office* (maxime inquisitoire, cf. art. 12 PA) et *l'obligation des parties de collaborer* (cf. art. 13 PA) sont aussi pertinents.

  En cas de procédures totalement ou partiellement automatisées, il faudra notamment garantir que tous les faits pertinents ont pu être pris en considération.[195] Cela suppose, d'une part, que les données utilisées par un système d'IA soient correctes et complètes et, d'autre part, que toutes les données dont la pertinence semble douteuse dans le contexte spécifique de la prise de décision soient écartées.[196]

  Les données factuelles peuvent également être obtenues par d'autres moyens, par exemple de la part du requérant. La saisie nécessairement standardisée des données factuelles dans l'optique d'une automatisation peut toutefois avoir pour conséquence que le requérant ne puisse pas fournir toutes les indications qui, de son point de vue, seraient nécessaires à la procédure. Si un formulaire ne contient que des champs

---

[193] ATF **144** II 427, 436, consid. 3.2.3.

[194] Dans ce sens aussi REGINA WEDER, Verfahrensgrundrechtliche Anforderungen an automatisierte Verwaltungsverfahren, in : Monika Simmler (éd.), Smart Criminal Justice. Der Einsatz von Algorithmen in der Polizeiarbeit und Strafrechtspflege, Basel 2021, 237 ss, 251 s.

[195] NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 10 N 25 s. ; voir aussi NADJA BRAUN BINDER, Der Untersuchungsgrundsatz als Herausforderung vollautomatisierter Verfahren, zsis 2/2020, N 28.

[196] Conseil de l'Europe, L'administration et vous – Un manuel. Principes de droit administratif concernant les relations entre l'Administration et les personnes, Strasbourg 2024, 18, disponible sous https://www.coe.int/fr/web/cdcj/-/publication-of-the-new-handbook-the-administration-and-you-that-takes-into-account-the-increasing-use-of-ai (consulté le 26 août 2024).

dans lesquels des valeurs numériques peuvent être inscrites, il n'est pas possible d'ajouter des explications supplémentaires qui pourraient être nécessaires pour pouvoir évaluer correctement le dossier dans le cas d'espèce.[197]

- La *garantie de l'accès au juge* (art. 29*a* Cst.) garantit à toute personne le droit d'être jugée par une autorité judiciaire. Il résulte de cette règle qu'une décision entièrement automatisée ne devrait pas pouvoir être prononcée en instance unique dans une cause individuelle.[198]

Il découle de ce qui précède que les garanties procédurales existantes s'appliquent déjà pleinement en matière de systèmes d'IA. Elles sont concrétisées dans les lois de procédure administrative, civile et pénale. En outre, l'art. 21 LPD offre déjà une certaine protection en procédure administrative de première instance en cas de décision exclusivement automatisée.

Les règles actuelles paraissent suffisantes, sous réserve des considérations relatives au devoir de motiver exposées plus haut (cf. ch. 4.3.3.2). La jurisprudence pourra préciser les contours de ces garanties dans le contexte des systèmes d'IA. Si malgré cela des problèmes ou lacunes devaient apparaître, l'on pourra examiner les cas où le recours à l'IA empêcherait les personnes concernées de faire valoir leurs droits en justice tels que prévus par la législation actuelle.

Il conviendra en tout état de cause de tenir compte du fait qu'on peut se heurter à des intérêts opposés, notamment celui privé du concepteur de l'algorithme, qui peut faire valoir le secret d'affaires, ou un autre intérêt public comme l'intérêt à la poursuite pénale, ce qui pourrait limiter la portée des garanties en question. En procédure administrative par exemple, il est déjà possible aujourd'hui, en vertu de l'art. 27 PA, de tenir compte d'intérêts privés contraires au droit de consulter le dossier lors du traitement des demandes dans ce sens.

[197] NADJA BRAUN BINDER/MATTHIAS SPIELKAMP/CATHERINE EGLI et al., Einsatz Künstlicher Intelligenz (n. 39), 38 ; NADJA BRAUN BINDER/NINA LAUKENMANN/LILIANE OBRECHT, KI in der Verwaltung (n. 28), 10 N 25 s. ; NADJA BRAUN BINDER, Der Untersuchungsgrundsatz (n. 185), N 28.

[198] MICHAEL MONTAVON, Cyberadministration et protection des données – Etude théorique et pratique de la transition numérique en Suisse du point de vue de l'État, des citoyen-ne-s et des autorités de contrôle, Fribourg 2021, 667.

Art. 15, par. 2 – Information en cas d'interaction avec un système d'IA

Selon l'art. 15, par. 2, chaque Partie cherche à veiller à ce que, en fonction du contexte, les personnes qui interagissent avec des systèmes d'IA soient informées du fait qu'elles interagissent avec de tels systèmes et non avec un humain. La disposition a pour but d'éviter le risque de manipulation et tromperie, notamment.[199]

En droit suisse, une obligation dans ce sens existe dans le cadre de la LPD, en lien avec les décisions individuelles automatisées. En effet, l'art. 21, al. 1, LPD prévoit que le responsable du traitement informe la personne concernée de toute décision individuelle automatisée. Si la décision individuelle automatisée émane d'un organe fédéral, ce dernier doit la qualifier comme telle (art. 21, al. 4, LPD). L'art. 25, al. 2, let. f, LPD prévoit, dans le contexte du droit d'accès, que la personne concernée soit informée de l'existence d'une décision individuelle automatisée.

Comme déjà relevé, l'art. 21 LPD ne couvre toutefois pas toutes les hypothèses, notamment car il ne concerne que les cas où une décision exclusivement automatisée est rendue. En revanche, l'art. 15, par. 2, de la convention déclencherait probablement l'obligation de notification par exemple aussi en cas d'interaction avec des robots de conversation dotés d'IA sur les sites web de l'administration publique.[200]

En conclusion, la mise en œuvre en droit interne de l'art. 15, par. 2, de la convention sur l'IA nécessite une intervention du législateur visant à élargir le devoir d'information existant dans la LPD. Cette information permettra par ailleurs à la personne concernée de faire valoir ses droits procéduraux.

La disposition laisse une marge de manœuvre au législateur, puisque l'expression « en fonction du contexte » permettra de tenir compte des situations où, notamment, il est évident que l'interaction a lieu avec une machine et qu'il n'y a donc pas de nécessité d'informer. Cette obligation n'est également pas destinée, par exemple, à couvrir les situations où l'objectif même de l'utilisation du système d'IA serait contrecarré par la notification, par exemple dans le contexte de la poursuite pénale.

4.3.4 Chapitre V : Évaluation et atténuation des risques et des impacts négatifs

Ce chapitre contient un seul article (art. 16), dont le titre est « Cadre de gestion des risques et des impacts ».

L'art. 16, par. 1, prévoit que chaque Partie, compte tenu des principes énoncés au chapitre III, adopte ou maintient des mesures afin d'identifier, d'évaluer, de prévenir et d'atténuer les

[199] Rapport explicatif convention sur l'IA (n. 20), N 104.

[200] Rapport explicatif convention sur l'IA (n. 20), N 104.

risques posés par les systèmes d'IA en tenant compte des impacts réels et potentiels sur les droits de l'homme, la démocratie et l'État de droit.

La disposition contient l'obligation d'établir un cadre de gestion des risques et des impacts. Son but est d'évaluer *ex ante*, c'est-à-dire avant le déploiement d'un système d'IA et, le cas échéant, de manière itérative tout au long du cycle de vie du système, les risques pertinents du système d'IA pour les droits de l'homme, la démocratie et l'État de droit, suivant une méthodologie assortie de critères concrets. Elle constitue un outil essentiel pour garantir le respect des exigences de la convention, en particulier celles du Chapitre III dédié aux « Principes ».

L'art. 16 garantit que les Parties à la convention adoptent une approche commune dans l'identification, l'analyse et l'évaluation des risques et des impacts des systèmes d'IA. En même temps, il repose sur l'idée que les Parties sont mieux placées pour faire les choix réglementaires pertinents, et leur laisse donc une marge d'appréciation en ce qui concerne la mise en œuvre. C'est la raison pour laquelle l'art. 16, par. 2, indique que les mesures à prendre doivent être, le cas échéant, graduées et différenciées, et tenir dûment compte du contexte et de l'utilisation prévue des systèmes d'IA, ainsi que de la gravité et de la probabilité des impacts potentiels (cf. art. 16, par. 2, let. a et b).

L'art. 16, par. 2, donne d'autres précisions en lien avec les mesures à adopter ou maintenir. La let. c prescrit l'obligation de prendre en compte, dans le cadre de la gestion des risques et des impacts, le point de vue des parties prenantes pertinentes, en particulier les personnes dont les droits pourraient être affectés. D'autres parties prenantes peuvent entrer en ligne de compte, tels des experts techniques externes ou des représentants de la société civile.[201]

La let. d prévoit le caractère itératif des mesures, en ce sens que celles-ci doivent être répétées tout au long du cycle de vie des systèmes d'IA. La let. e indique que les mesures doivent en outre comprendre un suivi des risques et des impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit. Ces caractéristiques permettent d'identifier et évaluer aussi les impacts *ex post* des systèmes d'IA, c'est-à-dire après leur déploiement. La let. f inclut un devoir de documentation des risques, des impacts réels et potentiels, et de l'approche de gestion des risques. À la let. g, il est précisé que le cadre de gestion des risques devra exiger, le cas échéant, l'essai préalable des systèmes d'IA avant leur mise à disposition pour première utilisation et lorsqu'ils subissent des modifications significatives.

L'art. 16, par. 3, prévoit quant à lui l'obligation pour les Parties d'agir afin de traiter les impacts négatifs avérés des systèmes d'IA de manière adéquate. Il s'agit donc de remédier à ces impacts par le biais de mesures appropriées. Cette phase devra être documentée et incluse dans le cadre de gestion des risques et des impacts au sens de l'art. 16, par. 2.

Finalement, l'art. 16, par. 4, prévoit que si une Partie considère que l'utilisation d'un système d'IA et incompatible avec le respect des droits de l'homme, la démocratie et l'État de droit, elle évalue la nécessité d'un moratoire, d'une interdiction ou d'autres mesures appropriées.

---

[201] Rapport explicatif convention sur l'IA (n. 20), N 108.

Compte tenu de leur gravité, des mesures telles qu'un moratoire ou une interdiction ne devraient être envisagées que dans des circonstances où une Partie estime qu'une utilisation particulière des systèmes d'IA pose un risque inacceptable et où, après un examen approfondi, il n'y a pas d'autres mesures disponibles pour atténuer ce risque. Ces mesures devraient également être assorties de procédures d'examen appropriées afin de permettre leur retrait une fois que les risques pertinents ont été suffisamment réduits ou que des mesures d'atténuation appropriées sont devenues disponibles.

Le Conseil de l'Europe va développer une méthodologie non contraignante permettant d'illustrer une manière possible de mettre en œuvre les obligations découlant de l'art. 16. Il s'agit de fournir un modèle permettant d'aider les Parties dans l'élaboration de leur propre cadre de gestion des risques et des impacts.

En droit suisse, le cadre de gestion des risques et des impacts des systèmes d'IA selon l'art. 16 de la convention n'est pas sans rappeler notamment l'analyse d'impact relative à la protection des données (AIPD) au sens de l'art. 22 LPD. Il s'agit toutefois de deux instruments différents.

Selon l'art. 22, al. 1, LPD, le responsable du traitement procède au préalable à une AIPD lorsque le traitement envisagé est susceptible d'entraîner un risque élevé pour la personnalité ou les droits fondamentaux de la personne concernée. Selon le Message LPD, l'exigence de mener une analyse d'impact si certaines conditions sont remplies vaut pour les responsables privés comme pour les organes fédéraux, raison pour laquelle la disposition mentionne un risque élevé non seulement pour la personnalité de la personne concernée, mais aussi pour ses droits fondamentaux.[202]

L'AIPD est un instrument moins étendu que le cadre de gestion des risques et des impacts prévu à l'art. 16 de la convention sur l'IA. En outre, s'agissant du champ d'application, il n'est pas le même : bien que la majorité des systèmes d'IA implique le traitement de données personnelles, tel n'est pas toujours le cas. Par ailleurs, les risques visés par le cadre de gestion des risques et des impacts sont plus étendus que pour l'AIPD.

En lien avec les traitements de données par les organes fédéraux, l'art. 22, al. 1, LPD indique qu'une AIPD doit être réalisée lorsque le traitement envisagé est susceptible d'entraîner *un risque élevé pour les droits fondamentaux*. Les droits fondamentaux visés par la LPD sont principalement le droit à la liberté personnelle (art. 10, al. 2, Cst.), le droit à la vie privée (art. 13, al. 1, Cst.) et le droit à l'autodétermination informationnelle au sens de l'art. 13, al. 2, Cst., aussi consacré à l'art. 8 CEDH.[203] En pratique, l'AIPD selon l'art. 22 LPD évalue généralement les conséquences négatives que le traitement de données personnelles pourrait avoir

---

[202] Message LPD (n. 188), 6676.

[203] JULIEN FRANÇAIS, in : Petit commentaire LPD (n. 44), art. 1 N 11 ; CR LPD-BERTIL COTTIER, art. 1 N 19 ; BSK DSG/BGÖ-MATTHIAS R. SCHÖNBÄCHLER/URS MAURER-LAMBROU/SIMON KUNZ, art. 1 N 18 ; MARCO FREY, in : Bruno Baeriswy/Kurt Pärli/Dominika Blonski (éds.), Datenschutzgesetz, Stämpfli Handkommentar, art. 1 N 29 ; DAVID ROSENTHAL, in : David Rosenthal/Yvonne Jöhri (éds.), Handkommentar zum Datenschutzrecht, art. 1 N 3 s.

avec une certaine vraisemblance sur la personne concernée, notamment ses conséquences physiques (p. ex. des données erronées peuvent conduire à un mauvais traitement médical), matérielles (p. ex. une personne se voit refuser un emploi ; une carte de crédit est utilisée de manière frauduleuse), ou immatérielles (p. ex. le fait de se savoir observée peut amener la personne à modifier son comportement).[204] Or, l'art. 16 de la convention sur l'IA vise à évaluer les risques et les impacts des systèmes d'IA sur tous les droits de l'homme, ainsi que la démocratie et l'État de droit, indépendamment du traitement de données personnelles.

En ce qui concerne le secteur privé, la mise en œuvre de l'art. 16 aurait pour conséquence que les acteurs privés concernés devraient réaliser une analyse d'impact en matière de droits fondamentaux en tout cas dans les domaines où un effet horizontal direct ou indirect des droits fondamentaux entre privés est donné ou devait être reconnu à l'avenir (cf. les considérations développées au ch. 4.2.3.1). Cela signifie par exemple de devoir se référer à la législation sur la protection des données, aux normes sur la protection de la personnalité telles que déclinées dans le CC et le CO, à la LEg, à la LHand, etc. En effet, la convention sur l'IA n'a pas pour effet d'étendre le champ de protection des droits fondamentaux.

Selon l'interprétation qui est faite ici, l'art. 16 laisse une certaine marge de manœuvre aux États, y compris en leur donnant la possibilité d'aller moins loin dans le secteur privé dans certains cas de figure. Il serait par exemple imaginable de limiter l'obligation d'effectuer une analyse d'impact des systèmes d'IA sur les droits fondamentaux à certains prestataires de services essentiels, à l'image de ce qui est prévu dans le règlement sur l'IA de l'UE (cf. art. 27 du règlement, ch. 5.2.7.3.3).

Au vu de tout ce qui précède, il apparaît que le droit suisse connaît une obligation d'effectuer une analyse d'impact lors du traitement de données personnelles (AIPD), mais qui n'est pas suffisante au regard du cadre de gestion des risques et des impacts des systèmes d'IA tel que décrit à l'art. 16 de la convention sur l'IA. Une intervention du législateur serait donc nécessaire. Celle-ci devrait toutefois tenir compte des spécificités du système suisse, en particulier s'agissant de la portée des droits fondamentaux dans les relations entre privés.

On relève aussi qu'en droit suisse, l'art. 36 Cst. conditionne les restrictions aux droits fondamentaux et impose des conditions. Une atteinte provenant d'un système IA qui ne repose pas sur une base légale, ou qui toucherait l'essence même du droit en cause, serait injustifiée. Dans le secteur public, l'art. 16, par. 4, de la convention n'appelle donc pas de besoin d'agir. Une interdiction ou un moratoire concernant certaines utilisations des systèmes d'IA pourrait faire du sens dans le secteur privé, ou dans le secteur public en cas de décision politique d'intervenir activement pour interdire certaines pratiques. La question pourrait se poser par rapport à des outils comme la reconnaissance des émotions.

[204] Dans ce sens, DAVID ROSENTHAL/SAMIRA STUDER/ALEXANDRE LOMBARD (pour la traduction), La nouvelle loi sur la protection des données (n. 129), 58, N 149 ; CR LPD-PHILIPPE GILLIÉRON, art. 22 N 24.

L'introduction d'un cadre de gestion des risques et des impacts des systèmes d'IA en droit suisse devrait en outre tenir compte de l'AIPD déjà existante en droit de la protection des données.

À noter que le règlement sur l'IA de l'UE prévoit que si le fournisseur d'un système d'IA effectue déjà une analyse d'impact relative à la protection des données, l'analyse d'impact selon le règlement sur l'IA est menée conjointement avec celle-ci (cf. ch. 5.2.7.3.3). La mise en place d'un mécanisme de coordination serait nécessaire en droit suisse également en cas de ratification de la convention sur l'IA.

4.3.5 Chapitre VI : Mise en œuvre de la convention

4.3.5.1 Généralités

Le Chapitre VI contient des obligations à la charge des Parties strictement connexes à la phase de mise en œuvre de la convention en droit interne.

4.3.5.2 Article 17 – Non-discrimination

Cette disposition prévoit que la mise en œuvre des dispositions de la convention par les Parties doit être assurée sans discrimination fondée sur quelque motif que ce soit, conformément à leurs obligations internationales en matière de droits de l'homme.

L'art. 17 interdit donc toute discrimination *dans la mise en œuvre* de la convention par les Parties. La notion de discrimination est identique à celle énoncée dans le droit international applicable et couvre un large éventail de motifs de discrimination liés aux caractéristiques personnelles des individus, à leur situation ou à leur appartenance à un groupe.[205]

Dans l'ordre juridique suisse, l'art. 8, al. 2, Cst., interdit déjà toute discrimination. En particulier, les autorités d'application du droit doivent interpréter les dispositions légales à la lumière de l'interdiction de discriminer.

4.3.5.3 Article 18 – Droits des personnes handicapées et des enfants

Cette disposition impose aux Parties l'obligation de tenir dûment compte des besoins et des vulnérabilités spécifiques en rapport avec le respect des droits des personnes handicapées et des enfants, conformément à leur droit interne et aux obligations internationales applicables. Elle renvoie ainsi aux dispositions et au régime juridique de la convention relative aux droits des personnes handicapées[206] et de la convention relative aux droits de l'enfant[207], ainsi

[205] Rapport explicatif convention sur l'IA (n. 20), N 114.

[206] Convention relative aux droits des personnes handicapées (RS **0.109**).

[207] Convention relative aux droits de l'enfant (RS **0.107**).

qu'au droit interne applicable de chaque Partie en matière de droits des personnes handicapées et de droits de l'enfant.

L'objectif est de garantir le plus haut niveau de prise en compte de tous les besoins et vulnérabilités spécifiques en rapport avec le respect des droits des personnes handicapées et des enfants, y compris la formation à la maîtrise du numérique.[208]

La Suisse a ratifié aussi bien la convention relative aux droits des personnes handicapées que la convention relative aux droits de l'enfant précitées, et s'est ainsi engagée à mettre en œuvre les obligations prévues par ces conventions dans tous les domaines de la vie.

Dans l'hypothèse d'une ratification, lors de la mise en œuvre des obligations de la convention, la Suisse devrait intégrer les besoins spécifiques des personnes handicapées et des enfants.

4.3.5.4 Article 19 – Consultation publique

La disposition incite les Parties à impliquer toute la société dans le débat relatif aux questions importantes soulevées par les systèmes d'IA, à la lumière notamment de leurs incidences sociales, économiques, juridiques, éthiques et environnementales.

L'expression « le cas échéant » laisse aux Parties la liberté de déterminer les modalités de ces consultations.[209] La disposition recommande « un débat public et de[s] consultations multipartites ».

La Suisse connaît déjà des instruments permettant la participation du public aux questions importantes soulevées par les systèmes d'IA. Ainsi notamment, les projets de lois fédérales impliquent en règle générale une consultation externe (cf. la LCo). Une telle consultation externe serait aussi prévue pour des projets de loi en matière d'IA. La procédure de consultation vise à associer les cantons, les partis politiques et les milieux intéressés à la définition de la position de la Confédération et à l'élaboration de ses décisions (art. 2 LCo). Selon l'art. 4, al. 1, LCo, toute personne ou organisation peut participer à la consultation et exprimer un avis.

Dans le contexte spécifique des systèmes d'IA, il convient aussi de mentionner d'autres mécanismes permettant la participation des milieux intéressés en Suisse, notamment la Plateforme tripartite pour la gouvernance numérique et l'IA.[210] Il s'agit d'un réseau d'information national permettant l'échange sur des thèmes liés au numérique. La Plateforme Tripartite est gérée par l'OFCOM et est ouverte à toutes les personnes intéressées du secteur privé, de la

[208] Rapport explicatif convention sur l'IA (n. 20), N 117.

[209] Rapport explicatif convention sur l'IA (n. 20), N 121.

[210] Cf. www.bakom.admin.ch > L'OFCOM > Activités internationales > Société de l'information internationale / SMSI > La Plateforme tripartite suisse pour la gouvernance numérique et l'intelligence artificielle.

société civile, du monde scientifique et de tous les niveaux de l'administration. Elle permet de comprendre les besoins, les préoccupations et les attentes des différentes parties prenantes et d'en tenir compte dans l'élaboration de règles appropriées dans le domaine du numérique, y compris dans celui de l'IA.

Au vu de ce qui précède, il apparaît que la Suisse dispose déjà d'instruments adéquats permettant de respecter l'obligation découlant de l'art. 19 de la convention.

4.3.5.5 Article 20 – Maîtrise du numérique et compétences numériques

Cette disposition prévoit l'obligation pour les Parties d'encourager et promouvoir la maîtrise du numérique et les compétences numériques adéquates pour toutes les catégories de la population, notamment les compétences spécifiques de pointe pour les personnes chargées de l'identification, de l'évaluation, de la prévention et de l'atténuation des risques que présentent les systèmes d'IA.

Les termes « maîtrise du numérique et compétences numériques » se réfèrent à la capacité d'utiliser, de comprendre et d'interagir efficacement avec les technologies numériques, y compris celles d'IA. Ces capacités sont fondamentales pour permettre à toute la population de saisir les chances et les risques que posent les systèmes d'IA.[211]

La maîtrise du numérique et les compétences numériques sont particulièrement importantes pour ceux qui recourent à l'IA dans le cadre de l'exercice de leurs fonctions. On peut citer notamment les employés qui utilisent un système d'IA dans le cadre d'un processus de prise de décision, ou les responsables de l'achat de systèmes d'IA qui sont utilisés dans des administrations publiques. Les personnes impliquées doivent pouvoir saisir les enjeux que posent les systèmes d'IA. Ces compétences permettent par exemple de contrer les biais d'automatisation ou de confirmation, qui font que les êtres humains accordent une plus grande confiance aux machines et aux artefacts technologiques qu'à leur propre jugement potentiellement contradictoire, validant ainsi de manière aveugle les résultats algorithmiques sans les remettre en question.

L'art. 20 mentionne les compétences spécifiques de pointe pour les personnes impliquées dans le cadre de gestion des risques et des impacts selon l'art. 16 de la convention. Il va en effet de soi que ces personnes devront disposer des capacités spécifiques nécessaires au bon déroulement des analyses des risques et des impacts selon cette disposition.[212]

La Confédération pourrait prendre des mesures dans ses domaines de compétence, par exemple prévoir des exigences de formation pour le personnel de l'administration fédérale qui utilise des systèmes d'IA dans le cadre de l'exercice de ses fonctions.

[211] Rapport explicatif de la convention sur l'IA (n. 20), N 122.

[212] Rapport explicatif de la convention sur l'IA (n. 20), N 124.

4.3.5.6 Article 21 – Sauvegarde des droits de l'homme reconnus

Cette disposition vise à assurer la coexistence de la convention sur l'IA avec d'autres traités et instruments internationaux relatifs aux droits de l'homme. Elle prévoit qu'aucune disposition de la convention ne peut être interprétée comme limitant, dérogeant ou affectant d'une quelconque manière les droits de l'homme ou d'autres droits et obligations juridiques connexes qui peuvent être garantis en vertu du droit interne d'une Partie ou de tout autre accord international pertinent.

Dans ce contexte, toutes les références au droit interne dans la convention doivent être interprétées comme se limitant aux cas où le droit interne prévoit un niveau de protection des droits de l'homme plus élevé que le droit international applicable.[213]

Cet article n'appelle pas de remarques sous l'angle de la mise en œuvre en droit interne.

4.3.5.7 Article 22 – Protection plus étendue

Cette disposition protège les dispositions du droit interne et des instruments internationaux contraignants, existants ou futurs, qui prévoient une protection supplémentaire pour les activités menées dans le cadre du cycle de vie des systèmes d'IA, allant au-delà du niveau garanti par la convention sur l'IA. Cette protection ne saurait être restreinte par la convention.[214]

Cet article n'appelle pas de remarques sous l'angle de la mise en œuvre en droit interne.

4.4 Chapitre VII : Mécanisme de suivi et coopération

4.4.1 Généralités

Ce chapitre contient les dispositions formant le mécanisme de suivi auquel il est fait référence à l'art. 1, par. 3, de la convention. Le but du mécanisme est celui d'assurer la mise en œuvre effective de la convention par les Parties. Parmi les instruments prévus, les trois premiers relèvent du plan international (art. 23 à 25), et le quatrième du droit interne (art. 26).

4.4.2 Article 23 – Conférence des Parties

Cet article prévoit la création d'un organe dans le cadre de la convention, à savoir la Conférence des Parties, composé de représentants des Parties à la convention. La mise en place de cet organe garantira une participation égale de toutes les Parties au processus décisionnel et à la procédure de suivi de la convention et renforcera également la coopération entre les Parties pour assurer une mise en œuvre adéquate et efficace de la convention.[215]

---

[213] Rapport explicatif de la convention sur l'IA (n. 20), N 126.

[214] Rapport explicatif de la convention sur l'IA (n. 20), N 127.

[215] Rapport explicatif convention sur l'IA (n. 20), N 130.

Cet organe adoptera son propre règlement intérieur (art. 23, par. 4). Il est convoqué par la Secrétaire Générale ou le Secrétaire Général du Conseil de l'Europe chaque fois que nécessaire, et, dans tous les cas, lorsque la majorité des Parties ou le Comité des Ministres en demande la convocation (art. 23, par. 3).

La Conférence des Parties a les compétences usuelles en matière de suivi. Il convient de mentionner en particulier la possibilité de faire des amendements de la convention, en faisant des propositions conformément à l'art. 28 de la convention. En outre, la Conférence des Parties a un rôle consultatif général à l'égard de la convention. Elle peut formuler des recommandations spécifiques sur toute question relative à son interprétation ou son application, y compris, par exemple, en suggérant des interprétations de divers termes juridiques contenus dans la convention. Bien qu'elles ne soient pas juridiquement contraignantes, ces recommandations peuvent être considérées comme l'expression d'une opinion commune des Parties sur un sujet donné, qui devrait être prise en compte de bonne foi par les Parties dans leur mise en œuvre de la convention.[216]

Si la Suisse adhère à la convention, elle devra désigner un ou plusieurs représentants pour siéger au sein de la Conférence des Parties.

4.4.3 Article 24 – Obligation de rapport

Afin de permettre la coopération et d'informer régulièrement au sujet de la mise en œuvre de la convention, chaque Partie fournit à la Conférence des Parties, dans un délai de deux ans à compter de la date à laquelle elle devient Partie, puis de manière périodique par la suite, un rapport contenant les détails des activités qu'elle a entreprises pour donner effet à l'art. 3, par. 1, let. a et b (cf. ch. 4.2.3.1).

La Conférence des Parties déterminera le format et le processus pour le rapport en accord avec son règlement intérieur.

Si la Suisse adhère à la convention, elle devra faire un rapport conformément à cette disposition et selon les modalités déterminées par la Conférence des Parties.

4.4.4 Article 25 – Coopération internationale

Cette disposition énonce l'obligation pour les Parties de coopérer à la réalisation de l'objectif de la Convention.

Les Parties sont en outre encouragées à aider, le cas échéant, les États non Parties à agir conformément aux dispositions de la convention et à y adhérer (par. 1).

En outre, les Parties échangent les informations pertinentes et utiles sur les aspects liés à l'IA qui peuvent avoir un effet positif ou négatif significatif sur la jouissance des droits de l'homme, le fonctionnement de la démocratie et le respect de l'État de droit, notamment sur

[216] Rapport explicatif convention sur l'IA (n. 20), N 132, let. c.

les risques et les effets apparus dans le cadre de la recherche et en relation avec le secteur privé. Elles sont dans ce cadre encouragées à associer, le cas échéant, les parties prenantes pertinentes et les États qui ne sont pas Parties à la convention à cet échange d'information (par. 2).

Enfin, les Parties sont encouragées à renforcer la coopération, y compris, le cas échéant, avec les parties prenantes pertinentes (p. ex représentants d'organisations non gouvernementales), afin de prévenir et d'atténuer les risques et les impacts négatifs sur les droits de l'homme, la démocratie et l'État de droit dans le contexte des activités menées dans le cadre du cycle de vie des systèmes d'IA.

Cette disposition est relativement vague et laisse les Parties assez libres sur le type de coopération prévue. Il s'agit de prime abord plus d'échanges d'informations que de coopération opérationnelle entre autorités.

4.4.5 Article 26 – Mécanismes de contrôle effectifs

L'art. 26 exige des Parties qu'elles mettent en place ou désignent un ou plusieurs mécanismes effectifs de contrôle du respect des obligations nées de la convention. Cela signifie que les Parties doivent réexaminer les mécanismes existants dans leurs ordres juridiques respectifs et si nécessaire, redéfinir leurs fonctions, voire mettre en place des structures entièrement nouvelles.

Selon le rapport explicatif, ces mécanismes doivent être fonctionnellement indépendants des pouvoirs exécutif et législatif. Ce terme englobe divers types d'indépendance fonctionnelle. Il peut s'agir de fonctions de contrôle intégrées au sein d'organes gouvernementaux particuliers qui évaluent ou supervisent le développement et l'utilisation de systèmes d'IA.

Les organes en question doivent en outre disposer des pouvoirs, de l'expertise, y compris des connaissances et compétences techniques, ainsi que des autres ressources nécessaires pour s'acquitter efficacement de leurs tâches. En cas de surveillance partagée entre différents organes, la convention implique de mettre en place une coopération efficace.[217]

Des mécanismes de contrôle effectif devraient être prévus dans le secteur privé aussi, là où un effet horizontal direct ou indirect des droits fondamentaux entre privés est donné ou devait être reconnu à l'avenir (cf. ch. 4.2.3.1).

Cette disposition nécessite une intervention du législateur soit pour confier à une ou plusieurs autorités existantes la tâche de contrôler le respect de la convention, soit pour créer une nouvelle autorité. Il existe diverses autorités de surveillance indépendantes en droit suisse (par ex. PFPDT), mais il n'existe pas d'autorité de surveillance qui couvre tout le champ de la convention. Le législateur devrait régler l'organisation et les pouvoirs d'intervention de l'autorité ou des autorités de surveillance en lien avec la convention. En outre, il

[217] Rapport explicatif convention sur l'IA (n. 20), N 141 ss.

conviendrait de régler la coordination entre les autorités de surveillance, lorsque plusieurs d'entre elles sont compétentes.

4.5 Chapitre VIII : Clauses finales

À quelques exceptions près, les dispositions des art. 27 à 36 s'inspirent pour l'essentiel du « Modèle de clauses finales pour les conventions, protocoles additionnels et protocoles d'amendement conclus au sein du Conseil de l'Europe », comme adopté par le Comité des Ministres du Conseil de l'Europe lors de la 1291e réunion des délégués des Ministres le 5 juillet 2017.[218]

Il convient de signaler en particulier l'art. 30 relatif à la signature et l'entrée en vigueur de la convention. Le par. 1 énonce que la convention est ouverte à la signature des États membres du Conseil de l'Europe, des États non membres du Conseil de l'Europe qui ont participé à son élaboration (Argentine, Australie, Canada, Costa Rica, Saint-Siège, Israël, Japon, Mexique, Pérou, États-Unis d'Amérique et Uruguay), et de l'Union européenne.

Selon l'art. 30, par. 3, la convention entrera en vigueur le premier jour du mois qui suit l'expiration d'une période de trois mois après la date à laquelle cinq signataires, dont au moins trois États membres du Conseil de l'Europe, auront manifesté leur consentement à être liés par la convention (par le biais d'une ratification, acceptation ou approbation, cf. art. 30, paragraphe 2).

Une fois la convention entrée en vigueur, d'autres États non membres n'ayant pas participé à l'élaboration de la convention pourront être invités à y adhérer par le Comité des Ministres du Conseil de l'Europe, conformément aux modalités fixées à l'art. 31.

4.6 Conclusions intermédiaires

L'analyse de la convention sur l'IA a permis d'esquisser les besoins de légiférer en droit suisse dans l'hypothèse où la Suisse décidait de ratifier la convention.

Indépendamment de la question de la ratification, l'examen a aussi permis d'évaluer le niveau de protection offert par le droit suisse par rapport aux défis que posent actuellement les systèmes d'IA en lien avec les droits de l'homme, la démocratie et l'Etat de droit en général, puisque la convention sur l'IA cristallise les grands enjeux reconnus au niveau international. Elle a donc le mérite de fournir un fil rouge permettant de questionner l'état actuel du droit suisse par rapport à ces thématiques.

Il convient de préciser que l'analyse devra être approfondie sur certains aspects, afin de tenir compte de l'ensemble de la législation pertinente existante. La présente analyse n'est qu'une première étape en vue d'un examen plus extensif, qui interviendra le cas échéant dans le cadre de la suite des travaux, après que le Conseil fédéral aura pris une décision de principe sur l'approche de régulation en matière d'IA.

[218] Rapport explicatif de la convention sur l'IA (n. 20), N 145.

Les conclusions intermédiaires de l'analyse sont les suivantes :

- La convention sur l'IA a pour but de protéger les droits de l'homme, le fonctionnement de la démocratie et l'État de droit. D'une manière générale, elle s'adresse aux États et contient des normes qui ne sont pas directement applicables.

- Les Parties jouissent d'une grande marge de manœuvre quant à la manière dont elles souhaitent mettre en œuvre les dispositions de la convention, qui n'indique pas les mesures exactes et concrètes à prendre afin d'atteindre ses objectifs.

  En particulier, la convention sur l'IA ne prévoit pas par exemple à quelles conditions des systèmes d'IA peuvent être autorisés, mais reste à un niveau plus général. Sur les relations entre la convention sur l'IA et le règlement sur l'IA, il est renvoyé au ch. 5.3.4.

- S'agissant du besoin de légiférer dans l'hypothèse d'une ratification, l'analyse a permis de constater trois cas de figure :

  - Au regard de certaines dispositions de la convention, le droit suisse paraît offrir un niveau de protection suffisant par rapport aux exigences de la convention, si bien qu'une intervention du législateur ne semble pas nécessaire à ce stade. Tel est le cas par exemple le cas de l'art. 5 (« Intégrité des processus démocratiques et respect de l'État de droit ») et de l'art. 19 (« Consultation publique »).

  - Au regard d'autres dispositions de la convention, le droit suisse contient déjà des règles pertinentes, mais celles-ci ne vont pas assez loin par rapport aux obligations du traité. Des adaptations seraient donc nécessaires afin de mettre en œuvre les dispositions de la convention. On peut citer notamment l'art. 8 (« Transparence et contrôle »), l'art. 13 (« Innovation sûre »), l'art. 14 (« Recours »), et l'art. 15 (« Garanties procédurales »).

    La réalisation de certains principes clés, tel celui prévu à l'art. 8 (« Transparence et contrôle »), permettent par ailleurs une meilleure efficacité du cadre légal déjà en vigueur en droit suisse, par exemple en matière de responsabilité, d'égalité et non-discrimination et de protection des données.

  - Au regard d'autres dispositions encore, le droit suisse ne prévoit pas de normes correspondantes. Il s'agit notamment de l'art. 16 (« Cadre de gestion des risques et des impacts ») et de l'art. 26 (« Mécanismes de contrôle effectifs »). Le législateur pourrait être amené à intervenir avec des mesures nouvelles, puisque le droit suisse ne règle que des aspects ponctuels.

  En conclusion, le cadre légal actuel contient donc des dispositions pertinentes, mais dans bien des cas il devrait être complété. Il convient donc de conclure que, dans l'hypothèse où la Suisse devait ratifier la convention sur l'IA, des adaptations seraient nécessaires.

- Le cas échéant, il appartiendra au législateur de décider du type et de l'étendue des mesures à adopter. Dans le cadre de l'élaboration d'un projet législatif proprement dit,

il conviendra aussi de déterminer, sur la base des principes de technique législative et en fonction des choix politiques qui seraient faits, comment les nouvelles normes devraient s’insérer dans le droit actuel. Par exemple, il conviendrait d’examiner si les mesures nouvelles pourraient être prévues dans une ou plusieurs lois existantes ou alors si elles devraient figurer dans une loi nouvelle. S’agissant de la densité normative, on se référera notamment aux exigences de l’art. 164 Cst.

Il apparaît cependant d’ores et déjà que, vu l’étendue des domaines juridiques qui seraient éventuellement touchés, les modifications concerneraient probablement plusieurs lois. Un besoin important de coordination apparaît déjà, compte tenu des domaines du droit concernés. En particulier il y a un important besoin de coordination avec la LPD.

- L’analyse a aussi permis de démontrer qu’un aspect important de la mise en œuvre serait celui de la proportionnalité. La portée des obligations de la convention devrait être adaptée en fonction de la gravité des atteintes potentielles. Encore une fois, la question de savoir comment concrètement il conviendrait d’établir cette gradation est avant tout de nature légistique et politique.

- S’agissant du secteur privé, l’analyse a démontré que la portée de la convention se limiterait aux cas où un effet horizontal direct ou indirect des droits fondamentaux dans les relations entre privés existe ou devait être reconnu à l’avenir (cf. ch. 4.2.3.1). La convention laisse une certaine marge de manœuvre afin de tenir compte des spécificités du droit suisse.

5 Règlement de l'Union européenne établissant des règles harmonisées concernant l'intelligence artificielle

5.1 Structure du chapitre et méthode

Ce chapitre expose tout d'abord le contenu du règlement sur l'IA (cf. ch. 5.2).

Il présente ensuite une appréciation juridique du règlement sur l'IA en analysant certains de ses aspects pertinents du point de vue du droit suisse (cf. ch. 5.3).

Sont dans ce cadre abordés les effets juridiques du règlement sur l'IA sur les opérateurs suisses (cf. ch. 5.3.1), les relations entre le règlement sur l'IA et l'Accord entre la Confédération suisse et l'Union européenne relatif à la reconnaissance mutuelle en matière d'évaluation de la conformité (ARM) (cf. ch. 5.3.2), et les relations avec la décision d'adéquation en matière de protection des données de la Commission européenne (cf. ch. 5.3.3). Les relations avec la convention sur l'IA sont aussi abordées (cf. ch. 5.3.4). Le chapitre appréhende ensuite d'autres éléments choisis (cf. ch. 5.3.5), et se termine par des conclusions intermédiaires (cf. ch. 5.4).

D'un point de vue méthodologique, la présentation diffère de l'analyse ci-dessus relative à la convention sur l'IA du Conseil de l'Europe (ch. 4). En effet, le règlement sur l'IA est une législation interne à l'UE qui ne lie pas la Suisse. D'une manière générale, ce règlement ne contient en l'état pas de pendant en droit suisse. Une approche d'analyse consistant en l'examen de chacune de ses dispositions par rapport au droit suisse, telle que celle choisie pour la convention du CAI, ne paraît donc pas utile à ce stade. Il convient en revanche de présenter le contenu du règlement sur l'IA et de se concentrer sur certains aspects pertinents permettant de dresser une appréciation générale du point de vue du droit suisse. Dans la mesure où le texte vient d'être adopté au sein de l'UE et que de nombreuses questions devront encore être clarifiées s'agissant de son application, seuls les principaux enjeux appréciables à ce stade pour la Suisse ont été mis en lumière.

5.2 Contenu du règlement

5.2.1 Contexte

Le 21 avril 2021, la Commission européenne a présenté le paquet législatif sur l'IA, comprenant une proposition de règlement établissant des règles harmonisées concernant l'intelligence artificielle et une proposition de règlement sur les machines et équipements (nouvelles procédures de conformité réglant entre autres l'Internet des objets).[219]

[219] Proposition de règlement du Parlement européen et du Conseil du 21 avril 2021 établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union, COM/2021/206 final, disponible sous https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A52021PC0206 (consulté le 26 août 2024) ; Proposition de règlement du Parlement européen et du Conseil du 21 avril 2021 sur les machines et produits connexes, COM/2021/202 final, disponible sous https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52021PC0202 (consulté le 26 août 2024).

Du point de vue de la Commission européenne, avec cette législation l'UE souhaite cimenter son rôle de pionnière dans la régulation de l'économie numérique et de leader dans l'établissement de standards qui ont vocation à influer sur le débat au niveau global concernant la manière d'aborder l'IA.

Le règlement sur l'IA constitue la première réglementation horizontale et directement contraignante de l'IA au niveau régional. Il couvre en particulier le développement, la mise sur le marché, la mise en service et l'utilisation des systèmes d'IA dans l'UE. Il a été adopté par le Parlement européen le 13 mars 2024 et par le Conseil le 21 mai 2024, signé le 12 juin, et publié au Journal officiel de l'UE le 12 juillet 2024. Il est entré en vigueur le 1er août 2024.[220]

Ce règlement ne lie pas la Suisse.

5.2.2 Objectifs de la réglementation

Selon l'art. 1, par. 1, l'objectif du règlement est d'améliorer le fonctionnement du marché intérieur et de promouvoir l'adoption d'une IA axée sur l'humain et digne de confiance, tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux consacrés dans la Charte des droits fondamentaux de l'UE, notamment la démocratie, l'état de droit et la protection de l'environnement, contre les effets néfastes des systèmes d'IA dans l'UE, tout en soutenant l'innovation.

L'objectif est donc d'établir un cadre juridique uniforme, en particulier pour le développement, la mise sur le marché, la mise en service et l'utilisation de systèmes d'IA dans l'UE.[221] Le règlement vise à garantir que les systèmes d'IA mis sur le marché de l'UE sont sûrs et respectent les normes européennes en vigueur, notamment s'agissant de la sécurité des produits.[222] Il vise également à permettre la libre circulation transfrontalière des biens et services fondés sur l'IA et à prévenir la fragmentation du marché.

L'expression « *tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux* » (cf. art. 1, par. 1) démontre que par le biais de cette approche de régulation fondée sur la sécurité des produits, le législateur européen a également voulu protéger d'autres intérêts, notamment les droits fondamentaux des personnes concernées. Il en résulte un règlement avec des objectifs hybrides (cf. ch. 5.3.4).

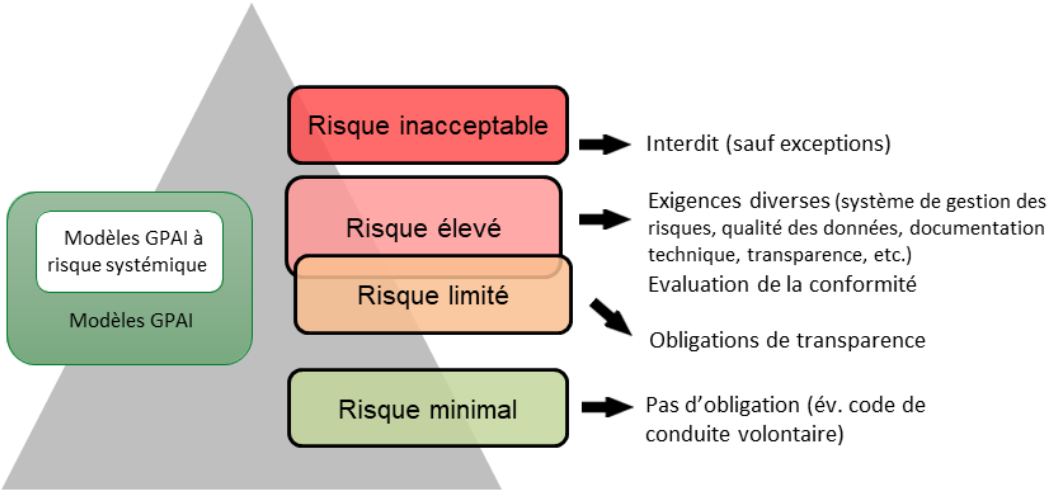
[220] Règlement (UE) 2024/1689 établissant des règles harmonisées concernant l'intelligence artificielle (n. 3).

[221] DAVID ROSENTHAL, Der EU AI Act – Verordnung über künstliche Intelligenz, Jusletter 5 août 2024, 5 N 6.

[222] MARTINA ARIOLI, Risikomanagement nach der EU-Verordnung über Künstliche Intelligenz, Jusletter IT 4 juillet 2024, 4 N 8 s.

5.2.3 Approche fondée sur les risques

Le règlement sur l'IA a une approche fondée sur les risques. En effet, il établit une classification des systèmes d'IA selon leur niveau de risque pour la santé, la sécurité et les droits fondamentaux. En fonction de cette classification, le règlement prévoit diverses obligations pour l'accès au marché.[223]



Source : OFJ

Selon cette approche, les systèmes d'IA présentant des risques inacceptables sont interdits, sauf exceptions (cf. ch. 5.2.6).

Les systèmes d'IA considérés comme étant à haut risque sont autorisés, mais sont soumis à de nombreuses exigences pour être mis sur le marché de l'UE (cf. ch. 5.2.7). En particulier, ils doivent faire l'objet d'une évaluation de la conformité (cf. ch. 5.2.7.4). Dans son évaluation d'impact de la réglementation[224], la Commission européenne estime que les applications d'IA à haut risque ne représentent que 5 à 15 % de toutes les applications sur le marché.

Les systèmes d'IA ne présentant qu'un risque limité sont seulement soumis à des obligations de transparence (cf. ch. 5.2.8). Toutefois, si ces systèmes remplissent les critères pour être

[223] Cf. ANGELA MÜLLER, Der Artificial intelligence Act der EU : Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz, Zeitschrift für Europarecht 1/2022, 1 ss, 7 ss et 15 s. (en lien avec une précédente version du règlement sur l'IA).

[224] Commission staff working document, Impact assessment – Accompanying the Proposal for a Regulation of the European Parliament and of the Council, Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative act, SWD(2021) 84 final, 68 (disponible sous https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0084 (consulté le 26 août 2024).

considérés à haut risque, les obligations prévues pour ces derniers systèmes s'appliqueront également.

Si un système d'IA ne tombe sous aucune des catégories mentionnées ci-devant, il est considéré en tant que système d'IA à risque minimal. Selon la Commission européenne, la grande majorité des systèmes d'IA présentent un risque minimal. Le règlement ne prévoit donc pas d'intervention étatique pour ces systèmes. Il autorise l'utilisation libre d'applications telles que les jeux vidéo intégrant des systèmes d'IA ou les filtres anti-spam reposant sur l'IA. Dans ce cas, les acteurs concernés sont encouragés à établir des codes de conduite.

La législation européenne réglemente également les modèles d'IA à usage général (en anglais *general purpose AI model* ; ci-après : modèles GPAI), qui ne figuraient pas dans la proposition législative initiale. Le règlement prévoit désormais des obligations pour tous les modèles GPAI et des obligations supplémentaires pour les modèles GPAI qui comportent des risques systémiques (cf. ch. 5.2.10).

5.2.4 Définitions

L'art. 3 du règlement sur l'IA est consacré à la définition de diverses notions (68 au total). La présente analyse se concentre sur les définitions de « système d'IA » et de « modèle d'IA à usage général ».

5.2.4.1 Système d'intelligence artificielle

L'art. 3, point 1, du règlement sur l'IA prévoit que l'on entend par système d'IA « un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels ».

La réglementation a pour vocation d'être à l'épreuve du temps et de couvrir les développements technologiques actuels et futurs en matière d'IA. Ainsi, la notion de système d'IA de l'UE est étroitement alignée sur les travaux des organisations internationales œuvrant dans le domaine de l'IA, telles que l'OCDE, afin d'assurer la sécurité juridique, faciliter la convergence internationale et l'acceptation générale.

Au vu des similitudes entre cette définition de système d'IA et celle prévue dans la convention sur l'IA, qui s'inspire elle aussi des travaux de l'OCDE, il est renvoyé aux développements en lien avec celle-ci pour plus d'explications (cf. ch. 4.2.2.1).

Il convient de relever également que la Commission élaborera des lignes directrices sur l'application de la définition d'un système d'IA (cf. art. 96, par. 1, point f).

5.2.4.2 Modèle d'intelligence artificielle à usage général

Le règlement fait la distinction entre les modèles d'IA à usage général et les systèmes d'IA.

Selon l'art. 3, point 63, un modèle d'IA à usage général est « un modèle d'IA, y compris lorsque ce modèle d'IA est entraîné à l'aide d'un grand nombre de données utilisant l'auto-supervision à grande échelle, qui présente une généralité significative et est capable d'exécuter de manière compétente un large éventail de tâches distinctes, indépendamment de la manière

dont le modèle est mis sur le marché, et qui peut être intégré dans une variété de systèmes ou d'applications en aval, à l'exception des modèles d'IA utilisés pour des activités de recherche, de développement ou de prototypage avant leur mise sur le marché ».

La définition de modèle d'IA à usage général se fonde sur les principales caractéristiques fonctionnelles de ce type de modèle, en particulier la généralité et la capacité d'exécuter avec compétence un large éventail de tâches distinctes. En sont des exemples « GPT » et « Llama ».[225]

Il résulte de cette définition que les modèles d'IA à vocation générale sont des composants essentiels des systèmes d'IA, mais ne constituent pas des systèmes d'IA à eux seuls. Les modèles d'IA à usage général nécessitent l'ajout d'autres composants, comme une interface utilisateur, pour devenir des systèmes d'IA. Les modèles d'IA à usage général sont ainsi généralement intégrés dans les systèmes d'IA et en font partie.

Lorsqu'un système d'IA est fondé sur un modèle d'IA à usage général, on parle de système d'IA à usage général, à savoir un système d'IA qui a la capacité de répondre à diverses finalités, tant pour une utilisation directe que pour une intégration dans d'autres systèmes d'IA (cf. art. 3, point 66).

5.2.5 Champ d'application

5.2.5.1 Étendue du champ d'application

Le règlement sur l'IA a un champ d'application large. Il s'applique aussi bien aux personnes privés qu'aux autorités publiques.[226]

Pour expliquer le champ d'application, il convient tout d'abord de préciser que le règlement fait la distinction entre les rôles suivants :

- Fournisseur : il s'agit de toute personne physique ou morale, autorité publique, agence ou tout autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général *et* le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit. Il peut être établi dans l'UE ou dans un pays tiers (cf. art. 3, point 3).

  À noter que par « mise sur le marché », on entend la première mise à disposition d'un système d'IA ou d'un modèle d'IA à usage général sur le marché de l'UE (cf. art. 3, point 9).

[225] DAVID ROSENTHAL, Der EU AI Act (n. 221), 3 N 5.

[226] DAVID ROSENTHAL, Der EU AI Act (n. 221), 2 N 2 ; 6 N 11.

Par « mise en service », on entend la fourniture d'un système d'IA en vue d'une première utilisation directement au déployeur ou pour usage propre dans l'Union, conformément à la destination du système d'IA (art. 3, point 11).

- Déployeur : il s'agit de toute personne physique ou morale, autorité publique, agence ou autre entité qui utilise un système d'IA sous son autorité, sauf si cette utilisation se fait dans le cadre d'une activité personnelle à caractère non professionnel (cf. art. 3, point 4).

- Importateur : toute personne physique ou morale située ou établie dans l'UE qui met sur le marché un système d'IA qui porte le nom ou la marque d'une personne physique ou morale établie dans un pays tiers (cf. art. 3, point 6).

- Distributeur : toute personne physique ou morale faisant partie de la chaîne d'approvisionnement, autre que le fournisseur ou l'importateur, qui met un système d'IA à disposition sur le marché de l'UE (cf. art. 3, point 7).

- Opérateur : un fournisseur, fabriquant de produits, déployeur, mandataire, importateur ou distributeur (cf. art. 3, point 8).

Le règlement sur l'IA s'applique avant tout aux fournisseurs établis ou situés dans l'UE ou dans un pays tiers qui mettent sur le marché ou mettent en service des systèmes d'IA, ou qui mettent sur le marché des modèles d'IA à usage général dans l'UE (cf. art. 2, par. 1, point a).

Il s'applique aussi aux déployeurs de systèmes d'IA qui ont leur lieu d'établissement ou sont situés dans l'UE (cf. art. 2, par. 1, point b).

La législation s'applique également aux fournisseurs et aux déployeurs de systèmes d'IA qui ont leur lieu d'établissement ou sont situés dans un pays tiers, lorsque les sorties produites par le système d'IA sont utilisées dans l'UE (art. 2, par. 1, point c). Cette disposition vise à éviter les mesures de contournement de la législation qui pourraient être mises en œuvre. La question de l'effet du règlement sur l'IA sur les opérateurs suisses sera approfondie ci-dessous (cf. ch. 5.3.1).

En outre, le règlement s'applique aux importateurs et aux distributeurs de systèmes d'IA (art. 2, par. 1, point d), aux fabricants de produits qui mettent sur le marché ou mettent en service un système d'IA en même temps que leur produit et sous leur propre nom ou leur propre marque (art. 2, par. 1, point e) et aux mandataires des fournisseurs qui ne sont pas établis dans l'UE (art. 2, par. 1, point f).

Il s'applique également aux personnes concernées qui sont situées dans l'UE (art. 2, par. 1, point g). Cela concerne par exemple les voies de recours prévues aux art. 85 ss du règlement.

5.2.5.2 Exceptions

Le règlement prévoit plusieurs exceptions à son application, notamment les suivantes :

- Les systèmes d'IA mis sur le marché, mis en service ou utilisés avec ou sans modifications exclusivement à des fins militaires, de défense ou de sécurité nationale, quel que

soit le type d'entité exerçant ces activités, sont exclus du champ d'application du règlement (art. 2, par. 3). Le domaine de la sécurité nationale ne fait en effet pas partie du droit de l'UE et relève de la responsabilité exclusive des États membres (art. 4, par. 2, TUE). L'exclusion du domaine militaire et de la défense est justifiée par l'art. 4, par. 2, TUE ainsi que par les spécificités de la politique de défense des États membres et de la politique de défense commune de l'UE relevant du titre V, ch. 2, TUE, qui sont soumises au droit international public. À noter que si un système d'IA est, temporairement ou définitivement, utilisé en dehors des domaines précités à d'autres fins (p. ex. à des fins civiles ou humanitaires, à des fins répressives ou de sécurité publique), il tombera dans le champ d'application du règlement (cf. consid. 24 du règlement).

- Le règlement ne s'applique pas aux autorités publiques d'un pays tiers, ni aux organisations internationales relevant du champ d'application du règlement selon l'art. 2, par. 1, lorsque ces autorités ou organisations utilisent des systèmes d'IA dans le cadre d'accords internationaux de coopération des services répressifs et judiciaires avec l'UE ou avec un ou plusieurs États membres, pour autant que des garanties en ce qui concerne la protection des droits fondamentaux et des libertés des personnes soient fournies (art. 2, par. 4).

- S'agissant des systèmes d'IA classés à haut risque selon l'art. 6, par. 1, qui sont liés aux produits couverts par les actes législatifs d'harmonisation énumérés à l'annexe I, section B (p. ex. aviation civile, véhicules agricoles et forestiers, équipements marins), seuls les art. 6, par. 1, 102 à 109 et l'art. 112 du règlement s'appliquent (cf. art. 2, par. 2).

- Le règlement ne s'applique pas aux systèmes ou aux modèles d'IA spécifiquement développés et mis en service uniquement à des fins de recherche et de développement scientifiques, ni à leurs sorties (art. 2, par. 6). En ce qui concerne les activités de recherche, d'essai et de développement, celles-ci ne sont pas incluses dans le champ d'application du règlement, mais elles doivent être menées conformément au droit de l'UE applicable. Les essais en conditions réelles ne sont pas couverts par cette exclusion (art. 2, par. 8 et art. 60).

- Le règlement ne s'applique pas aux obligations incombant aux déployeurs qui sont des personnes physiques utilisant des systèmes d'IA dans le cadre d'une activité strictement personnelle à caractère non professionnel (art. 2, par. 10).

- Finalement, il ne s'applique pas non plus aux systèmes d'IA publiés dans le cadre de licences libres et ouvertes, sauf s'ils sont mis sur le marché ou mis en service en tant que systèmes d'IA à haut risque ou en tant que système d'IA qui relèvent de l'art. 5 (systèmes interdits) ou de l'art. 50 (systèmes d'IA soumis aux obligations de transparence) (cf. art. 2, par. 12).

5.2.6 Pratiques interdites

Suivant l'approche fondée sur les risques décrite ci-dessus (cf. ch. 5.2.3), le Chapitre II du règlement sur l'IA interdit huit pratiques considérées comme présentant un risque inacceptable (cf. art. 5, par. 1) :[227]

- Systèmes d'IA utilisant des techniques subliminales : est interdite la mise sur le marché, la mise en service ou l'utilisation des systèmes d'IA qui ont recours à des techniques subliminales, ou à des techniques délibérément manipulatrices ou trompeuses, avec pour objectif ou effet d'altérer substantiellement le comportement d'une personne ou d'un groupe de personnes en portant considérablement atteinte à leur capacité à prendre une décision éclairée, amenant ainsi la personne à prendre une décision qu'elle n'aurait pas prise autrement, d'une manière qui cause ou est raisonnablement susceptible de causer un préjudice important à cette personne, à une autre personne ou à un groupe de personnes (art. 5, par. 1, point a).

- Systèmes d'IA qui exploitent des vulnérabilités : sont interdits les systèmes qui exploitent les vulnérabilités dues à l'âge, au handicap ou à la situation sociale ou économique d'une personne physique ou d'un groupe de personnes avec pour objectif ou effet d'altérer substantiellement le comportement de cette personne ou d'un membre de ce groupe d'une manière qui cause ou est raisonnablement susceptible de causer un préjudice important à cette personne ou à un tiers (art. 5, par. 1, point b).

- Notation sociale par des acteurs publics et privés : sont interdits les systèmes qui évaluent ou classent des personnes physiques ou des groupes de personnes en fonction de leur comportement social ou de caractéristiques personnelles ou de personnalité, connues, déduites ou prédites, *et* que la note sociale conduit à l'une ou l'autre des situations suivantes (art. 5, par. 1, point c) :

  - un traitement préjudiciable ou défavorable dans des contextes sociaux dissociés du contexte dans lequel les données ont été générées ou collectées à l'origine ;

  - un traitement préjudiciable ou défavorable qui est injustifié ou disproportionné par rapport à leur comportement social ou à la gravité de celui-ci.

- Police prédictive : sont interdits les systèmes d'IA visant à évaluer ou prédire le risque qu'une personne physique commette une infraction pénale, uniquement sur la base de son profilage ou de l'évaluation de ses traits de personnalité ou caractéristiques. L'interdiction ne vaut pas si le système est utilisé pour étayer une évaluation humaine (art. 5, par. 1, point d).

[227] Voir pour des considérations critiques DAVID ROSENTHAL, Der EU AI Act (n. 221), 18 N 42 ss.

- Collecte en masse d'images à des fins de reconnaissance faciale : sont interdits les systèmes d'IA qui créent ou développent des bases de données de reconnaissance faciale en récupérant de manière non ciblée des images faciales sur l'Internet ou des images de vidéosurveillance (art. 5, par. 1, point e).

- Reconnaissance émotionnelle : sont interdits les systèmes d'IA qui permettent d'inférer les émotions d'une personne sur le lieu de travail et dans les établissements d'enseignement, sauf si l'utilisation est due à des raisons médicales ou de sécurité (art. 5, par. 1, point f).

- Systèmes de catégorisation biométrique : les systèmes de catégorisation biométrique basés sur les données biométriques des individus, telles que le visage ou les empreintes digitales, pour déduire les opinions politiques, l'appartenance à un syndicat, les croyances religieuses ou philosophiques, la race, la vie sexuelle ou l'orientation sexuelle d'un individu, sont interdits (art. 5, par. 1, point g).

- Identification biométrique à distance, en temps réel, dans des espaces accessibles au public et à des fins répressives : une telle pratique est en principe interdite (art. 5, par. 1, point. h). L'expression « en temps réel » signifie que l'acquisition des données biométriques, la comparaison et l'identification se déroulent sans décalage temporel significatif (cf. art. 3, point 42).

  L'utilisation de cette technique est considérée comme étant particulièrement attentatoire aux droits et libertés des personnes, dans la mesure où elle peut affecter la vie privée de la population, susciter un sentiment de surveillance constante et dissuader indirectement l'exercice de la liberté de réunion et d'autres droits fondamentaux. En outre, les imprécisions techniques des systèmes d'IA destinés à l'identification biométrique à distance des personnes physiques peuvent conduire à des résultats biaisés et entraîner des effets discriminatoires (consid. 32 du règlement).

  Le recours à ces systèmes est autorisé dans trois situations exhaustives (cf. art. 5, par. 1, point h) : i) la recherche ciblée de victimes spécifiques d'enlèvement, de la traite ou de l'exploitation sexuelle d'êtres humains, ainsi que la recherche de personnes disparues ; ii) la prévention d'une menace spécifique, substantielle et imminente pour la vie ou la sécurité de personnes, ou la menace réelle et actuelle ou réelle et prévisible d'attaque terroriste ; iii) la localisation ou l'identification d'auteurs ou de suspects d'infractions pénales listées à l'annexe II et punissables dans l'État membre concerné d'une peine ou d'une mesure de sûreté privatives de liberté d'une durée maximale d'au moins quatre ans.

  L'art. 5, par. 2 et 3, fixe les conditions auxquelles l'identification biométrique à distance peut être utilisée dans ces situations exceptionnelles. En particulier, chaque utilisation doit être soumise à l'autorisation d'une autorité judiciaire ou d'une autorité administrative indépendante. Cette autorisation doit en principe être obtenue avant l'utilisation du système en vue d'identifier une ou plusieurs personnes, sauf dans des cas d'urgence justifiés. Dans ces derniers cas, il est possible de commencer à utiliser le système sans autorisation à condition que cette dernière soit demandée sans retard injustifié, au plus tard dans les 24 heures. L'autorisation n'est accordée que si l'utilisation est nécessaire et proportionnée à la réalisation de l'un des objectifs énumérés à l'art. 5, par. 1, point h et, en particulier, que cette utilisation reste limitée au strict nécessaire.

En outre, l'utilisation de ces systèmes n'est autorisée que si l'autorité répressive a réalisé une analyse d'impact sur les droits fondamentaux selon l'art. 27 du règlement, et a enregistré le système dans la base de données prévue à l'art. 49, sauf cas exceptionnels (cf. art. 5, par. 2, *in fine*) (cf. ch. 5.2.7.3.3).

L'autorité de surveillance du marché nationale (cf. ch. 5.2.14) et l'autorité nationale de protection des données doivent être informées de chaque utilisation d'un système d'identification biométrique à distance en temps réel et doivent soumettre à la Commission européenne un rapport annuel sur l'utilisation de ces systèmes. Sur la base des rapports des États membres, la Commission publie annuellement un rapport global (art. 5, par. 6 et 7).

Dans les limites des art. 5, par. 1, point h, et 5, par. 2 et 3, les États membres peuvent prévoir dans leur droit national des règles détaillées en ce qui concerne notamment la demande et la délivrance des autorisations (cf. art. 5, par. 5).

5.2.7 Systèmes d'intelligence artificielle à haut risque

5.2.7.1 Classification

Le règlement sur l'IA prévoit que les systèmes d'IA à haut risque ne doivent être mis sur le marché, mis en service ou utilisés que s'ils sont conformes à certaines exigences.

Selon le règlement, il existe deux catégories de systèmes d'IA à haut risque :

- Les systèmes d'IA utilisés comme composants de sécurité d'un produit couvert par la législation d'harmonisation de l'UE selon l'annexe I (p. ex. jouets, dispositifs médicaux), ainsi que les systèmes d'IA constituant eux-mêmes de tels produits, à la condition que ces produits soient soumis à une évaluation de la conformité par un tiers en vue de leur mise sur le marché ou leur mise en service, conformément à la législation d'harmonisation de l'UE (art. 6, par. 1).

- Les systèmes d'IA visés à l'annexe III (art. 6, par. 2), à savoir des systèmes d'IA déployés dans huit domaines spécifiques[228] :

  - Biométrie[229] ;
  - Infrastructures critiques ;

[228] Voir pour des considérations critiques DAVID ROSENTHAL, Der EU AI Act (n. 221), 22 N 51 ss.

[229] Sont exclus de cette catégorie les systèmes d'IA destinés à être utilisés à des fins de vérification biométrique, parmi lesquelles l'authentification, dont la seule finalité est de confirmer qu'une personne physique donnée est bien celle qu'elle prétend être et de confirmer l'identité d'une personne physique dans le seul but d'avoir accès à un service, de déverrouiller un dispositif ou de disposer d'un accès sécurisé à des locaux. Les systèmes biométriques destinés à être utilisés uniquement dans le but de permettre la cybersécurité et les mesures de protection des données à caractère personnel ne devraient pas être considérés comme des systèmes d'IA à haut risque (cf. annexe III, point 1, let. a et consid. 54).

- Éducation et formation professionnelle ;
- Emploi, gestion de la main-d'œuvre et accès à l'emploi indépendant ;
- Accès et droit aux services privés essentiels et aux services publics et prestations sociales essentiels ;
- Répression ;
- Migration, asile et gestion des contrôles aux frontières ;
- Administration de la justice et processus démocratiques.

La Commission est habilitée à modifier cette liste (cf. art. 7).

Cela étant, bien que visé à l'annexe III, un système d'IA n'est pas considéré comme étant à haut risque dans certains cas où le système est considéré comme n'ayant pas d'incidence significative sur la prise de décision ou ne causant pas de préjudice important aux intérêts juridiques à protéger (cf. art. 6, par. 3 et consid. 53) :

- Le système d'IA est destiné à exécuter une tâche procédurale étroite. Il peut par exemple s'agir d'un système qui transforme des données non structurées en données structurées ou un système d'IA qui est utilisé pour détecter les doublons parmi un grand nombre d'applications (consid. 53 du règlement).

- Le système d'IA est destiné à améliorer le résultat d'une activité humaine précédemment réalisée, par exemple si un système d'IA est destiné à améliorer la langue utilisée dans des documents déjà rédigés (consid. 53 du règlement).

- Le système d'IA est destiné à détecter les constantes en matière de prise de décision ou les écarts par rapport aux constantes habituelles antérieures et n'est pas destiné à se substituer à l'évaluation humaine préalablement réalisée ni à influencer celle-ci, sans examen humain approprié. Le risque est considéré comme étant réduit parce que l'utilisation du système d'IA fait suite à une évaluation humaine préalable, qu'il n'est pas censé remplacer ou influencer, sans examen humain approprié (consid. 53).

- Le système d'IA est destiné à effectuer une tâche qui n'est que préparatoire en vue d'une évaluation pertinente aux fins des cas d'utilisation visés à l'annexe III (consid. 53).

Nonobstant ces quatre cas de figure, un système d'IA visé à l'annexe III sera toujours considéré comme étant à haut risque s'il effectue un profilage de personnes physiques (cf. art. 6, par. 3 *in fine*).

Pour assurer la traçabilité et une transparence des systèmes qui sont exemptés en vertu des critères susmentionnés, le fournisseur qui estime qu'un système d'IA visé à l'annexe III ne présente pas de risque doit documenter son évaluation avant que ce système ne soit mis sur le marché ou mis en service. Ce fournisseur est également tenu d'enregistrer le système dans la base de données de l'UE. Il doit fournir cette documentation aux autorités nationales compétentes qui en font la demande (art. 6, par. 4).

La Commission européenne devra fournir des orientations supplémentaires pour la mise en œuvre pratique de l'art. 6 du règlement, après avoir consulté le Comité européen de l'IA (ch.

5.2.13). La Commission élaborera des lignes directrices précisant la mise en œuvre, et donnera une liste exhaustive d'exemples de cas d'utilisation de systèmes d'IA présentant ou non un risque élevé (cf. art. 6, par. 5).

5.2.7.2 Exigences applicables aux systèmes d'IA à haut risque

Les systèmes d'IA à haut risque doivent respecter les exigences énoncées à la Section 2 du Chapitre III, soit les art. 8 à 15 du règlement. La question de savoir quelles obligations incombent à quel acteur (fournisseur, déployeur ou autre) en vue de respecter les exigences de base est tranchée aux art. 16 ss du règlement (cf. ch. 5.2.7.3).

Les exigences à respecter sont les suivantes :

- Mise en place d'un système de gestion des risques (art. 9) : le système de gestion des risques vise à identifier et à atténuer les risques connus et raisonnablement prévisibles que le système d'IA à haut risque peut poser pour la santé, la sécurité ou les droits fondamentaux, lorsqu'il est utilisé conformément à sa destination (art. 9, par. 2, point a). Il vise aussi à estimer et évaluer les risques susceptibles d'apparaître en cas de mauvaise utilisation raisonnablement prévisible (art. 9, par. 2, point b). Il s'agit d'un processus continu et itératif, planifié et mis en œuvre tout au long du cycle de vie d'un système d'IA à haut risque. Le système de gestion des risques implique l'adoption de mesures appropriées et ciblées de gestion des risques, conçues pour répondre aux risques identifiés (art. 9, par. 2, point d).[230]

- Données et gouvernance des données (art. 10) : les systèmes d'IA à haut risque faisant appel à des techniques qui impliquent l'entraînement de modèles d'IA au moyen de données, doivent respecter des conditions particulières en matière de données et gouvernance de données. En particulier, les jeux de données d'entraînement, de validation et de test doivent satisfaire aux exigences des art. 10, par. 2 à 5 du règlement.

  Les données utilisées doivent notamment être de haute qualité afin de garantir que le système fonctionne comme prévu et qu'il ne devienne pas une source de discriminations. Cela implique que les ensembles de données utilisés soient pertinents, suffisamment représentatifs et, dans toute la mesure du possible, exempts d'erreurs et complets au regard de la finalité. Les ensembles de données doivent en outre présenter les propriétés statistiques appropriées, notamment en ce qui concerne les personnes ou les groupes de personnes pour lesquels le système d'IA est destiné à être utilisé, et accorder une attention particulière à l'atténuation des biais possibles dans les ensembles de données. Il faut en outre prévoir des mesures visant à garantir que les données traitées soient protégées et soumises à des garanties appropriées. Des règles spécifiques concernant l'utilisation de données personnelles, y compris de données sensibles, sont aussi prévues.

- Documentation technique (art. 11) : pour permettre la traçabilité des systèmes d'IA à haut risque, vérifier le respect des exigences prévues par le règlement, ainsi que le

[230] Voir pour des approfondissements MARTINA ARIOLI, Risikomanagement (n. 222), 6 N 19 ss.

contrôle de leur fonctionnement et la surveillance postérieure à la mise sur le marché, il est nécessaire de disposer d'informations compréhensibles sur la manière dont ces systèmes ont été développés et sur leur fonctionnement tout au long de leur durée de vie. L'art. 11 prévoit donc l'obligation de disposer d'une documentation technique sous une forme claire et intelligible pour évaluer la conformité du système d'IA, comprenant au minimum les éléments énoncés à l'annexe IV. Il s'agit notamment des caractéristiques générales, les capacités et les limites du système, les algorithmes, les données, l'entraînement, les essais et les processus de validation utilisés, ainsi que la documentation relative au système de gestion des risques.

- Enregistrement (art. 12) : les systèmes d'IA à haut risque doivent permettre, d'un point de vue technique, l'enregistrement automatique des événements (journaux) tout au long de la durée de vie du système. Un des buts est de faciliter la surveillance après commercialisation visée à l'art. 72, ainsi que la surveillance du fonctionnement du système comme prévu à l'art. 26, par. 5, du règlement (cf. art. 12, par. 2).

- Transparence et fourniture d'informations aux déployeurs (art. 13) : les systèmes d'IA à haut risque doivent être conçus et développés de manière à permettre aux déployeurs d'interpréter les sorties d'un système et de les utiliser de manière appropriée. Les systèmes d'IA à haut risque doivent être accompagnés d'informations appropriées sous la forme d'un mode d'emploi.

- Contrôle humain (art. 14) : les systèmes d'IA à haut risque doivent être conçus et développés de manière à pouvoir être effectivement contrôlés par des personnes physiques pendant la période d'utilisation du système d'IA. Ce contrôle effectif vise à prévenir ou à réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux susceptibles d'apparaître lorsqu'un système d'IA à haut risque est utilisé conformément à sa destination ou dans des conditions de mauvaise utilisation raisonnablement prévisible. Des mesures proportionnées aux risques doivent être prises. Elles peuvent soit être directement intégrées dans le système par le fournisseur, soit être identifiées par celui-ci, mais mises en œuvre par le déployeur (cf. art. 14, par. 3). Il peut s'agir par exemple de la possibilité d'interrompre le fonctionnement d'un système d'IA à l'aide d'un bouton d'arrêt (cf. art. 14, par. 4, point e).

  Les systèmes d'identification biométrique visés à l'annexe III, point 1 a), sont soumis à un contrôle humain renforcé. En effet, aucune mesure ou décision ne peut être prise par le déployeur sur la base de l'identification résultant du système, à moins qu'elle n'ait été vérifiée et confirmée séparément par au moins *deux* personnes physiques disposant des compétences, de la formation et de l'autorité nécessaires. Cette obligation ne s'applique toutefois pas aux systèmes d'IA utilisés à des fins répressives ou dans les domaines de la migration, des contrôles aux frontières ou de l'asile, lorsque le droit de l'Union ou le droit national considère que l'application de cette exigence est disproportionnée (art. 14, par. 5 *in fine*).

- Exactitude, robustesse et cybersécurité (art. 15) : des mesures techniques doivent être prises pour garantir un niveau approprié d'exactitude, de robustesse et de cybersécurité des systèmes d'IA à haut risque.

5.2.7.3 Obligations incombant aux fournisseurs et à d'autres parties

5.2.7.3.1 Obligations incombant aux fournisseurs

Le règlement met l'accent sur le rôle clé du fournisseur, à qui il attribue la responsabilité de la conformité des systèmes d'IA à haut risque.[231]

Les fournisseurs de systèmes d'IA à haut risque assument ainsi la plupart des obligations en lien avec ces systèmes. Ce sous-chapitre se concentre sur les obligations principales :

- Selon l'art. 16, point a, les fournisseurs veillent à ce que leurs systèmes soient conformes aux exigences du Chapitre III, Section 2 (art. 8 à 15 du règlement, cf. ch. 5.2.7.2). À la demande motivée d'une autorité nationale compétente, il leur incombe de prouver la conformité du système d'IA à ces exigences (cf. art. 16, point k).

- Les fournisseurs doivent être identifiables (art. 16, point b).

- Selon les art. 16, point c, et 17, les fournisseurs de systèmes d'IA à haut risque doivent mettre en place un *système de gestion de la qualité*, incluant le système de gestion des risques prévu à l'art. 9, et prévoyant divers autres aspects, comme la stratégie du respect de la réglementation, les systèmes de conservation des documents, la gestion des ressources, un cadre de responsabilisation interne, etc. Si un mécanisme de gestion de la qualité est déjà prévu dans la législation sectorielle, ces aspects peuvent être intégrés dans le système de gestion de la qualité conformément à ladite législation (cf. art. 17, par. 3).

- Les art. 18 et 19 prévoient des obligations en matière de conservation des documents et de tenue des journaux générés automatiquement.

- L'art. 20 prévoit l'obligation d'intervenir en cas de soupçon qu'un système d'IA mis sur le marché n'est pas conforme au règlement.

  Les fournisseurs de systèmes d'IA à haut risque doivent ainsi également disposer d'un « système de surveillance après commercialisation » (cf. art. 3, point 25, et 72) pour garantir que les risques éventuels liés aux systèmes d'IA qui continuent à « apprendre » après leur mise sur le marché ou leur mise en service puissent être traités efficacement (consid. 155).

- Conformément à l'art. 21, le fournisseur a en outre une obligation de coopérer avec les autorités compétentes.

- Les fournisseurs doivent veiller à ce que le système d'IA à haut risque soit soumis à la procédure d'évaluation de la conformité selon l'art. 43, avant sa mise sur le marché ou sa mise en service (art. 16, point f). Il leur incombe également d'élaborer la déclaration

[231] Cf. pour une critique de cette approche ANGELA MÜLLER, Der Artificial intelligence Act der EU (n. 223), 18.

de conformité selon l'art. 47, d'apposer le marquage CE (conformité européenne) sur le système (art. 48) et de dûment enregistrer le système (cf. art. 49).

Il est renvoyé sur ces aspects aux développements ci-dessous (cf. ch. 5.2.7.4).

Les obligations prévues aux art. 16 ss passent à tout distributeur, importateur ou déployeur ou autre tiers en présence de circonstances particulières, à savoir par exemple lorsqu'un de ces acteurs modifie la destination d'un système d'IA de telle manière qu'il devient un système d'IA à haut risque (par exemple un système de type ChatGPT est utilisé pour analyser les CV de candidats à l'emploi).[232] Dans ces cas, le fournisseur « originaire » n'est plus considéré en tant que fournisseur au sens du règlement. Pour les autres cas de figure, il est renvoyé à l'art. 25 du règlement.

5.2.7.3.2 Obligations incombant aux importateurs et aux distributeurs

Les importateurs ont l'obligation de s'assurer, avant de mettre sur le marché un système d'IA à haut risque, que celui-ci est conforme au règlement (cf. art. 23). Ils devront par exemple vérifier que le fournisseur a suivi la procédure d'évaluation de la conformité selon l'art. 43, et qu'il a établi la documentation technique conformément à l'art. 11 et à l'annexe IV. Ils devront également indiquer leur nom ainsi que l'adresse à laquelle ils peuvent être contactés (cf. art. 23, par. 3).

Les distributeurs devront quant à eux s'assurer que le système porte le marquage CE requis, qu'il est accompagné d'une copie de la déclaration UE de conformité et de la notice d'utilisation, et que le fournisseur et l'importateur dudit système, selon le cas, ont respecté leurs obligations respectives en vertu de l'art. 16, points b) et c), et de l'art. 23, par. 3 (cf. art. 24).

5.2.7.3.3 Obligations incombant aux déployeurs

Selon le règlement, après les fournisseurs, les déployeurs constituent le deuxième groupe d'acteurs auxquels il incombe le plus grand nombre d'obligations en matière de systèmes d'IA à haut risque.

Le règlement établit tout d'abord des devoirs généraux, à savoir notamment :

- L'obligation de prendre des mesures techniques et organisationnelles appropriées afin de garantir une utilisation conforme des systèmes d'IA à haut risque (art. 26, par. 1).

- Les déployeurs doivent en outre s'assurer que le contrôle humain soit confié à des personnes physiques disposant des compétences, de la formation et de l'autorité et du soutien nécessaire (art. 26, par. 2).

[232] DAVID ROSENTHAL, Der EU AI Act (n. 221), 10 N 25.

- Si le déployeur a la possibilité de contrôler les données d'entrée, il devra s'assurer que celles-ci soient pertinentes et suffisamment représentatives au regard de la destination du système d'IA à haut risque (art. 26, par. 4).

- Les déployeurs ont l'obligation de surveiller le fonctionnement du système à haut risque. Ils ont une obligation d'informer en cas de risque selon l'art. 79, ou lorsqu'un incident grave est détecté (art. 26, par. 5).

- Ils assurent la tenue des journaux générés automatiquement par les systèmes d'IA à haut risque, dans la mesure où ces journaux se trouvent sous leur contrôle (art. 26, par. 6).

Le règlement établit ensuite des devoirs spécifiques en fonction du déployeur en cause, notamment les suivants :

- Les déployeurs qui sont des employeurs informent les représentants des travailleurs et les travailleurs concernés qu'ils seront soumis à l'utilisation d'un système d'IA à haut risque, et ce avant sa mise en service ou son utilisation (art. 26, par. 7).

- Les déployeurs qui sont des autorités publiques ou des institutions, organes ou organismes de l'UE, doivent respecter les obligations d'enregistrement prévues à l'art. 49 (art. 26, par. 8). Ils doivent notamment enregistrer l'usage des systèmes énumérés à l'annexe III, à l'exception de ceux énumérés au point 2 (ces derniers étant enregistrés au niveau national) (cf. art. 49, par. 3 et 5). L'enregistrement des systèmes à haut risque utilisés visés à l'annexe III, points 1, 6, et 7, dans les domaines des activités répressives, de la migration, de l'asile et de la gestion des contrôles aux frontières, se fait dans une section non publique de la base de données, à laquelle seules la Commission et certaines autorités nationales ont accès (art. 49, par. 4).

- Les déployeurs d'un système d'IA à haut risque pour l'identification biométrique à distance a posteriori[233] dans le cadre d'une enquête en vue de la recherche ciblée d'une personne soupçonnée d'avoir commis une infraction pénale ou condamnée pour avoir commis une infraction pénale, doivent demander une autorisation pour l'utilisation du système, ex ante ou sans retard injustifié, au plus tard dans les 48 heures, de la part d'une autorité judiciaire ou administrative (cf. art. 26, par. 10).

  Le règlement précise qu'en aucun cas, ces systèmes ne peuvent être utilisés à des fins répressives de manière non ciblée, sans aucun lien avec une infraction pénale, une procédure pénale, une menace réelle et actuelle ou réelle et prévisible d'une infraction pénale ou la recherche d'une personne disparue spécifique. Il convient en outre d'assurer qu'aucune décision produisant des effets juridiques défavorables à l'égard d'une

[233] L'expression « a posteriori » signifie qu'il y a un décalage temporel entre l'acquisition de données biométriques, la comparaison et l'identification (cf. art. 3, points 42 et 43).

personne ne puisse être prise par les autorités répressives sur la seule base des sorties de tels systèmes.

Une autorisation n'est pas nécessaire si le système est utilisé pour l'identification initiale d'un suspect potentiel sur la base de faits objectifs et vérifiables directement liés à l'infraction.

Finalement, les déployeurs de ces systèmes doivent soumettre aux autorités de surveillance du marché concernées et aux autorités nationales chargées de la protection des données des rapports annuels sur leur utilisation, à l'exclusion de la divulgation de données opérationnelles sensibles liées aux services répressifs (art. 26, par. 10).

- Si le déployeur utilise un système d'IA à haut risque visé à l'annexe III pour prendre des décisions ou faciliter la prise de décision concernant des personnes physiques, il devra informer les personnes concernées (art. 26, par. 11). Cette obligation s'applique sans préjudice de l'art. 50 (obligation de transparence, cf. ch. 5.2.8). Pour les systèmes utilisés à des fins répressives, l'art. 13 de la directive UE 2016/680[234] s'applique.

  Le droit de la personne concernée à obtenir des explications en lien avec des décisions individuelles automatisées s'applique en sus (cf. art. 86, ch. 5.2.15).

- Finalement, certains déployeurs ont l'obligation d'effectuer une analyse d'impact des systèmes d'IA à haut risque sur les droits fondamentaux.[235] Cette obligation est réglée à l'art. 27 :

  - Destinataires de l'obligation : doivent effectuer une analyse d'impact 1) les organismes de droit public ou les entités privées fournissant des services publics[236], 2) les

[234] Directive (UE) 2016/680 du Parlement Européen et du Conseil du 27 avril 2016, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil, JO L 119/89, 4 mai 2016.

[235] Sur la différence avec le système de gestion des risques (art. 9), cf. MARTINA ARIOLI, Risikomanagement (n. 222), 14 N 48 s.

[236] Il n'est pas clair si le terme « services publics » désigne également des activités relevant purement du droit privé, mais tel semble être le cas au vu des considérants (cf. consid. 96). Sont mentionnés les services fournis dans l'intérêt public, notamment dans les domaines de l'éducation, des soins de santé, des services sociaux, du logement et de l'administration de la justice, cf. DAVID ROSENTHAL, Der EU AI Act (n. 221), 30 N 58.

déployeurs (publiques ou privés) de systèmes d'IA à haut risque visés à l'annexe III, par. 5, points b) et c)[237].

- Champ d'application : l'analyse d'impact sur les droits fondamentaux doit être effectuée avant le déploiement des systèmes d'IA à haut risque visés à l'art. 6, par. 2, du règlement, à savoir les systèmes visés à l'annexe III. Ne sont pas touchés par l'obligation les systèmes d'IA destinés à être utilisés dans le domaine des infrastructures critiques (cf. annexe III, point 2).

- Contenu : l'analyse d'impact sur les droits fondamentaux doit permettre au déployeur d'identifier les risques spécifiques pour les droits fondamentaux des individus susceptibles d'être affectés, et d'identifier les mesures à prendre si ces risques se matérialisent (cf. art. 27, par. 1, point a à f). Afin d'aider les déployeurs à effectuer l'analyse, le Bureau de l'IA élabore un modèle de questionnaire, y compris au moyen d'un outil automatisé (art. 27, par. 5).

  Si certains aspects de l'analyse sont déjà couverts au moyen de l'analyse d'impact relative à la protection des données (art. 35 RGPD ou art. 27 de la directive UE 2016/680[238]), l'analyse d'impact sur les droits fondamentaux pourra compléter l'analyse d'impact relative à la protection des données (cf. art. 27, par. 4).

- Surveillance : une fois l'analyse effectuée, le déployeur notifie le résultat à l'autorité de surveillance du marché (cf. ch. 5.2.14). Des exceptions sont possibles dans certains cas (cf. art. 27, par. 3).

5.2.7.4 Évaluation de la conformité

5.2.7.4.1 Procédures

Selon l'art. 16, point f), du règlement, les fournisseurs de systèmes d'IA à haut risque veillent à ce que ces systèmes soient soumis à la procédure d'évaluation de la conformité visée à l'art. 43, avant sa mise sur le marché ou sa mise en service.

---

[237] À savoir : les déployeurs de systèmes d'IA destinés à être utilisés pour évaluer la solvabilité des personnes physiques ou pour établir leur note de crédit, à l'exception des systèmes d'IA utilisés à des fins de détection de fraudes financières (annexe III, point 5, let. b) ; les déployeurs de systèmes d'IA destinés à être utilisés pour l'évaluation des risques et la tarification en ce qui concerne les personnes physiques en matière d'assurance-vie et d'assurance maladie (annexe III, point 5, let. c).

[238] Directive (UE) 2016/680 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel dans le domaine répressif (n. 234).

L'art. 43 traite de l'évaluation de la conformité. Les procédures varient en fonction du système[239] :

- Pour les systèmes d'IA à haut risque couverts par la législation d'harmonisation de l'UE (cf. annexe I, section A), le respect des exigences relatives aux systèmes d'IA à haut risque (art. 8 ss, cf. ch. 5.2.7.2) s'intégrera aux procédures d'évaluation de conformité prévues par ces législations, afin de réduire la charge pesant sur les opérateurs et d'éviter tout doublon. Les organismes notifiés selon ces règles sectorielles sont habilités à évaluer la conformité des systèmes d'IA à haut risque (cf. art. 43, par. 3).

  En effet, l'applicabilité des exigences du règlement sur l'IA ne doit pas affecter la logique spécifique, la méthodologie ou la structure générale de l'évaluation de la conformité prévues dans la législation sectorielle.

- Pour les systèmes d'IA à haut risque qui ne tombent pas sous le coup d'une réglementation d'harmonisation (cf. annexe III), le règlement sur l'IA prévoit deux façons de procéder à l'évaluation de la conformité, à savoir une évaluation de la conformité *basée sur le contrôle interne* (selon l'annexe VI) et, dans certains cas, une évaluation de la conformité par une *tierce partie* (organisme notifié[240]) (selon l'annexe VII).

  Cette approche devrait contribuer à créer et à consolider une culture de la conformité chez les fournisseurs qui ne sont pas actuellement soumis à des obligations liées aux produits, sans imposer dès le départ des procédures plus lourdes qui pourraient décourager l'innovation de manière disproportionnée.

  Par conséquent, l'évaluation de la conformité doit être effectuée en règle générale par le fournisseur sous sa propre responsabilité (self-assessment). Dans ce sens, l'art. 43, par. 2, du règlement prévoit que les systèmes d'IA à haut risque visés à l'annexe III, points 2 à 8, sont soumis à une procédure d'évaluation de la conformité fondée sur le contrôle interne, qui ne prévoit donc pas l'intervention d'un organisme notifié.

  Si nécessaire, afin notamment de prévenir ou réduire les risques pour la sécurité et les droits fondamentaux, la Commission est habilitée à intervenir et à soumettre ces systèmes à tout ou partie de la procédure d'évaluation par un organisme notifié (art. 43, par. 6).

---

[239] Les procédures sont basées sur le nouveau cadre législatif (New Legislative Framework, NLF), voir à ce propos MARTINA ARIOLI, Risikomanagement (n. 222), 4 N 8 s.

[240] Pour effectuer des évaluations de la conformité par des tiers lorsque cela est requis, les organismes notifiés doivent être notifiés au titre de la législation par les autorités nationales compétentes, à condition qu'ils satisfassent à un ensemble d'exigences, notamment en matière d'indépendance, de compétence, d'absence de conflits d'intérêts et d'exigences appropriées en matière de cybersécurité. Les autorités nationales compétentes doivent envoyer la notification de ces organismes à la Commission et aux autres États membres (cf. art. 28 ss du règlement).

S'agissant des systèmes d'IA destinés à être utilisés à des fins biométriques (cf. annexe III, point 1), l'art. 43, par. 1, du règlement, prévoit un mécanisme plus détaillé.

Le règlement sur l'IA accorde dans les deux cas un rôle important aux normes harmonisées (il est renvoyé sur ce point au ch. 5.2.11).

L'art. 46 du règlement prévoit les conditions auxquelles il est possible d'obtenir une autorisation, de la part de l'autorité de surveillance du marché, permettant de mettre sur le marché ou mettre en service des systèmes d'IA à haut risque tout en dérogeant à la procédure d'évaluation de la conformité. Des dérogations sont possibles par exemple pour des raisons exceptionnelles de sécurité publique. Pour les systèmes d'IA à haut risque liés à des produits couverts par la législation d'harmonisation de l'Union dont la liste figure à l'annexe I, section A, seules les dérogations à l'évaluation de la conformité établies dans ladite législation d'harmonisation de l'Union s'appliquent (cf. art. 46, par. 7).

5.2.7.4.2 Déclaration de conformité, marquage CE et enregistrement

Une fois qu'il a été évalué si le système d'IA à haut risque respecte les exigences du règlement, soit par le biais d'un contrôle interne, soit via une évaluation par un organisme notifié – conduisant à la livraison d'un certificat de conformité (cf. art. 44) – le fournisseur doit établir une déclaration UE de conformité écrite, contenant les informations prévues à l'annexe V. Cette dernière atteste que le système d'IA à haut risque satisfait aux exigences réglementaires. Le fournisseur assume ainsi la responsabilité du respect de ces exigences (cf. art. 47, par. 4).

Toujours dans le but d'éviter des procédures doubles, si une déclaration UE de conformité est déjà prévue par des actes législatifs sectoriels (actes d'harmonisation), une seule déclaration est établie (art. 47, par. 3).

Les systèmes d'IA à haut risque doivent porter le marquage CE (conformité européenne) indiquant leur conformité au règlement pour pouvoir circuler librement dans le marché intérieur (cf. art. 48). Lorsque des systèmes d'IA à haut risque sont soumis à d'autres actes législatifs de l'Union qui prévoient aussi l'apposition du marquage CE, ce marquage indique que les systèmes d'IA à haut risque satisfont également aux exigences de ces autres actes législatifs.

Finalement, avant de mettre sur le marché, mettre en service ou utiliser un système d'IA à haut risque visé à l'annexe III (sauf le point 2) les fournisseurs et certains déployeurs devront s'enregistrer et enregistrer leurs systèmes conformément à leurs obligations en la matière (cf. art. 49), ce dans une base de données gérée par la Commission, en collaboration avec les États membres (cf. art. 71) (les systèmes d'IA à haut risque visés à l'annexe III, point 2 [infrastructures] sont en revanche enregistrés au niveau national). Les informations à fournir figurent à l'annexe VIII.

5.2.8 Obligations de transparence pour les fournisseurs et les déployeurs de certains systèmes d'IA

Au Chapitre IV, le règlement prévoit des obligations de transparence pour les fournisseurs et les déployeurs dans des cas spécifiques (art. 50 ss). Ces obligations s'appliquent à tous les systèmes d'IA, qu'ils soient des systèmes d'IA à haut risque ou non. Ainsi, pour les systèmes

d'IA à haut risque, ces obligations s'ajoutent aux autres obligations de transparence déjà prévues pour ces systèmes (art. 50, par. 6) :

- Systèmes d'IA destinés à interagir directement avec des personnes physiques : les *fournisseurs* doivent veiller à ce que les systèmes d'IA soient conçus et développés de manière à ce que les personnes physiques concernées soient informées qu'elles interagissent avec un système d'IA, par exemple avec un robot logiciel (chatbot), à moins que cela ne soit évident en raison des circonstances et du contexte de l'utilisation. L'obligation ne s'applique pas en présence par exemple de systèmes utilisés à des fins de détection d'infractions pénales (art. 50, par. 1).

- Systèmes d'IA, y compris les systèmes d'IA à usage général, qui génèrent des contenus de type audio, image, vidéo ou texte : les *fournisseurs* sont tenus de faire en sorte que les sorties du système d'IA soient marquées dans un format lisible par une machine et identifiables comme ayant été générées ou manipulées par une IA. Le règlement prévoit une exception pour les systèmes d'IA qui remplissent une fonction d'assistance pour la mise en forme standard ou ne modifient pas de manière substantielle les données d'entrées fournies par le déployeur, ainsi que pour les systèmes dont l'utilisation est autorisée par la loi à des fins de par exemple de détection des infractions pénales (art. 50, par. 2).[241]

- Systèmes de reconnaissance des émotions ou systèmes de catégorisation biométrique : les *déployeurs* doivent informer les personnes physiques exposées au système de son fonctionnement et du traitement de leurs données personnelles. Les systèmes d'IA qui sont légalement utilisés pour détecter, prévenir, enquêter et poursuivre des infractions pénales sont exclus de cette obligation (art. 50, par. 3).

- Hypertrucages (deepfakes) : les *déployeurs* de systèmes d'IA qui génèrent des hypertrucages doivent indiquer que le contenu a été généré ou manipulé par une IA. L'utilisation des techniques dans le cadre de la loi pour détecter, prévenir, enquêter et poursuivre une infraction pénale est toutefois permise. Les hypertrucages utilisés dans le contexte artistique, créatif, satirique, fictif ou analogue, peuvent être indiqués de manière à ne pas entraver la présentation ou la jouissance de l'œuvre (art. 50, par. 4).

- Les textes publiés dans le but d'informer le public sur des questions d'intérêt public indiquent qu'ils ont été générés ou manipulés par l'IA, à moins que le texte n'ait été vérifié par un être humain ou un contrôle éditorial, et qu'une personne physique ou morale n'ait assumé la responsabilité éditoriale de sa publication (art. 50, par. 4).

Le règlement prévoit que les informations visées à l'art. 50, par. 1 à 4, doivent être fournies aux personnes physiques de manière claire et reconnaissable et au plus tard au moment de la première interaction ou de l'exposition.

---

[241] Cf. Pour des considérations pratiques, cf. DAVID ROSENTHAL, Der EU AI Act (n. 221), 34 N 69.

Le Bureau de l'IA (cf. ch. 5.2.13) encourage l'élaboration de codes de bonne pratique au niveau de l'UE afin de faciliter la mise en œuvre de ces obligations relatives à la détection et à l'étiquetage des contenus générés ou manipulés par l'IA. La Commission peut approuver des codes de bonnes pratiques par le biais d'un acte d'exécution, ou adopter un acte d'exécution précisant les règles communes pour la mise en œuvre de ces obligations si elle considère le code de bonnes pratiques inadéquat (art. 50, par. 7).

5.2.9 Autres systèmes d'intelligence artificielle

Si un système d'IA ne constitue ni une pratique interdite, ni un système d'IA à haut risque, ni ne tombe sous le coup des dispositions relatives à la transparence (art. 50), le règlement sur l'IA n'impose en principe aucune obligation aux fournisseurs, aux déployeurs et aux autres entités impliquées.

Toutefois, le règlement prévoit que le Bureau de l'IA et les États membres encouragent et facilitent l'élaboration de codes de conduite destinés à favoriser l'application volontaire à ces autres systèmes de tout ou partie des exigences énoncées au Chapitre III, Section 2, soit les règles applicables aux systèmes d'IA à haut risque (cf. art. 95, par. 1). Les codes de conduite doivent aussi encourager d'autres aspects, tels que des principes éthiques, une gestion respectueuse des ressources environnementales, la maîtrise de l'IA, et la prévention des impacts négatifs sur les personnes ou groupes de personnes vulnérables (art. 95, par. 2).

En outre, l'art. 4 du règlement prévoit une sorte d'obligation générale à charge de tous les fournisseurs et déployeurs de systèmes d'IA de prendre des mesures pour garantir, dans toute la mesure du possible, un niveau suffisant de maîtrise de l'IA pour leur personnel et les autres personnes s'occupant du fonctionnement et de l'utilisation des systèmes d'IA pour leur compte (« AI literacy »).[242]

5.2.10 Modèles d'intelligence artificielle à usage général

5.2.10.1 Remarques générales

Comme relevé ci-dessus (cf. ch. 5.2.4), le règlement sur l'IA fait la distinction entre les systèmes d'IA et les modèles GPAI. Le Chapitre V du règlement est spécialement dédié aux modèles GPAI.

Les modèles GPAI n'étaient pas inclus dans la proposition de règlement sur l'IA de la Commission européenne du 4 avril 2021. L'émergence de ces modèles, avec le lancement de ChatGPT en novembre 2022, a toutefois substantiellement imprégné les négociations entre le Parlement européen et le Conseil. L'approche réglementaire finale sur laquelle les colégislateurs se sont accordés inclut ainsi de nouvelles obligations pour les fournisseurs de modèles GPAI « simples » et ceux qui comportent des risques systémiques.

[242] DAVID ROSENTHAL, Der EU AI Act (n. 221), 36 N 72.

5.2.10.2 Obligations pour tous les modèles d'intelligence artificielle à usage général

Les fournisseurs[243] de modèles GPAI – indépendamment du risque que ces derniers représentent – ont notamment les obligations suivantes (cf. art. 53) :

- Ils doivent créer et mettre à jour régulièrement la documentation technique de leur modèle. Les éléments minimaux de la documentation sont décrits aux annexes XI et XII.

  La documentation doit notamment inclure les détails des processus d'entraînement et d'essai, ainsi que les résultats de l'évaluation. Elle doit être mise à la disposition, sur demande, du Bureau de l'IA et des autorités nationales compétentes (cf. art. 53, par. 1, point a).

- Les fournisseurs de modèles GPAI doivent également préparer et mettre à jour des informations et la documentation à l'intention des fournisseurs de systèmes d'IA qui envisagent d'intégrer le modèle GPAI dans leurs systèmes, afin notamment de les aider à comprendre les capacités et les limites du modèle GPAI et à se conformer au règlement (cf. art. 53, par. 1, point b).

- Les fournisseurs de modèles GPAI sont également tenus de mettre en place des règles internes pour se conformer au droit d'auteur de l'UE (art. 53, par. 1, point c).[244]

- Les fournisseurs sont en outre tenus d'établir et de mettre à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour l'entraînement du modèle GPAI. Ce résumé doit permettre aux parties ayant des intérêts légitimes, y compris les titulaires de droits d'auteur, d'exercer et de faire respecter leurs droits en vertu du droit de l'UE. Le Bureau de l'IA établira un modèle permettant aux prestataires de fournir le résumé requis sous forme narrative (art. 53, par. 1, point d).

- Les fournisseurs de modèles GPAI situés en dehors de l'UE, mais qui mettent des modèles GPAI sur le marché de l'UE doivent désigner un représentant dans l'UE (cf. art. 54).

Les modèles GPAI sous une licence libre et ouverte sont dispensés des obligations figurant aux art. 53, par. 1, point a et b, et à l'art. 54. Cette exemption s'applique lorsque les paramètres des modèles d'IA, y compris les poids, les informations sur l'architecture du modèle, et les informations sur l'utilisation du modèle, sont accessibles au public. Elle ne vaut toutefois pas pour les modèles GPAI présentant un risque systémique (cf. ch. 5.2.10.3 ; art. 53, par. 2 et 54, par. 6).

Les fournisseurs de modèles GPAI peuvent s'appuyer sur des codes de bonne pratique au sens de l'art. 56 du règlement, pour démontrer qu'ils se conforment à leurs obligations, ce

---

[243] Pour les modèles GPAI, le rôle du déployeur n'existe pas, cf. DAVID ROSENTHAL, Der EU AI Act (n. 221), 31 N 62.

[244] Sur l'interprétation de cette obligation, cf. DAVID ROSENTHAL, Der EU AI Act (n. 221), 33 N 65.

jusqu'à ce qu'une norme harmonisée soit publiée. Le respect des normes harmonisées conférera au fournisseur une présomption de conformité (art. 53, par. 4). Le Bureau de l'IA est chargé d'encourager et de faciliter la création de codes de bonne pratique, en tenant compte des approches internationales.

5.2.10.3 Obligations pour les modèles d'intelligence artificielle à usage général présentant un risque systémique

5.2.10.3.1 Définition et procédure

Les modèles GPAI présentant des risques systémiques sont soumis à des obligations supplémentaires, d'après le principe selon lequel l'augmentation des capacités accroît également les risques.

Un modèle GPAI est en principe considéré comme présentant des risques systémiques s'il possède des « capacités à fort impact » (cf. art. 51). La législation prend comme critère la quantité cumulée de calcul utilisée pour son entraînement, mesurée en opérations en virgule flottante (floating point operations, FLOP) (cf. consid. 111 du règlement). Le seuil de 10^25 FLOPs est fixé pour déterminer s'il agit d'un modèle GPAI présentant des risques systémiques (cf. art. 51, par. 2).

Ce seuil et ce critère pourront être ajustés au fil du temps pour refléter les changements technologiques et industriels, tels que les améliorations algorithmiques ou l'efficacité accrue du matériel. Le groupe scientifique d'experts (cf. ch. 5.2.13) peut également fournir une alerte qualifiée au Bureau de l'IA lorsqu'il a des raisons de soupçonner qu'un modèle GPAI pose un risque systémique au niveau de l'UE (art. 90).

Les règles pour la classification d'un modèle GPAI comme modèle à risque systémique sont prévues aux art. 51 et 52 du règlement. Ainsi, lorsqu'un modèle d'IA à usage général remplit le critère visé à l'art. 51, par. 1, point a) (il dispose de capacités à fort impact), le fournisseur concerné en informe la Commission européenne sans tarder. Il peut dans le même temps tenter de prouver que le modèle ne présente malgré tout pas de risques systémiques (cf. art. 52, par. 2).

La Commission peut aussi elle-même désigner un modèle d'IA à usage général comme présentant un risque systémique (art. 52, par. 4). Dans ce cas, le fournisseur peut solliciter une réévaluation.

La Commission publie et tient à jour une liste des modèles d'IA à usage général présentant un risque systémique, dans le respect des règles sur la propriété intellectuelle et les secrets d'affaires (art. 52, par. 6).

5.2.10.3.2 Obligations supplémentaires

En sus des obligations auxquelles sont soumis tous les modèles GPAI (cf. art. 53 et 54, ch. 5.2.10.2), les fournisseurs de modèles GPAI représentant des risques systémiques ont notamment les obligations suivantes (cf. art. 55) :

- Tests contradictoires (adversarial testing) : les fournisseurs doivent effectuer les évaluations nécessaires des modèles, en particulier avant leur première mise sur le marché, y compris la réalisation et la documentation de tests contradictoires des modèles,

le cas échéant par le biais de tests internes ou externes indépendants, afin d'identifier et atténuer les risques systémiques.

- Atténuation des risques systémiques : les fournisseurs doivent évaluer et atténuer les éventuels risques systémiques au niveau de l'UE, qui peuvent résulter du développement, de la mise sur le marché ou de l'utilisation du modèle.

- Notification des incidents : si, malgré les efforts déployés pour identifier et prévenir les risques liés à un modèle GPAI susceptible de présenter des risques systémiques, le développement ou l'utilisation du modèle provoque un incident grave, le fournisseur du modèle doit sans retard injustifié assurer le suivi de l'incident et communiquer au Bureau de l'IA ainsi qu'aux autorités nationales compétentes toute information pertinente ainsi que les éventuelles mesures correctives prises pour y remédier.

- Cybersécurité : les fournisseurs doivent assurer un niveau adéquat de protection de la cybersécurité pour le modèle et son infrastructure physique, le cas échéant, tout au long du cycle de vie du modèle.

Les fournisseurs peuvent démontrer leur conformité par le biais de codes de bonne pratique qui seront facilités par le Bureau de l'IA jusqu'à ce qu'une norme harmonisée soit publiée. Le respect de normes harmonisées conférera au fournisseur une présomption de conformité (art. 55, par. 2 ; ch. 5.2.11).

5.2.11 Les normes harmonisées et leur rôle dans le règlement sur l'intelligence artificielle[245]

La normalisation joue un rôle essentiel pour la mise en œuvre du règlement sur l'IA.[246] Le respect des normes harmonisées permet en effet aux fournisseurs des systèmes d'IA et modèles GPAI concernés de démontrer qu'ils se conforment aux exigences du règlement.

Pour comprendre de quelle manière ces normes s'articulent avec le règlement sur l'IA, sont présentés ci-dessous la notion de normes harmonisées (cf. ch. 5.2.11.1), le système de normalisation de l'UE (cf. ch. 5.2.11.2) et le mécanisme de présomption de conformité (cf. ch. 5.2.11.3). Les travaux de normalisation dans le domaine de l'IA sont ensuite exposés (cf. ch. 5.2.11.4).

5.2.11.1 Les normes harmonisées

On entend par « norme harmonisée » une norme européenne adoptée sur la base d'une demande formulée par la Commission pour l'application de la législation d'harmonisation de

[245] Ce chapitre a été rédigé sur la base de contributions de la DDIP.

[246] Cf. consid. 121 du règlement sur l'IA ; voir aussi MARTINA ARIOLI, Risikomanagement (n. 222), 12 N 38 ss.

l'UE (cf. art. 2, par. 1, point c du règlement UE 1025/2012[247]). Ces normes sont des normes techniques qui ont pour but de soutenir la législation européenne. Il s'agit de règles, lignes directrices ou caractéristiques non contraignantes, élaborées par des experts et expertes au sein des organisations européennes de normalisation. Elles s'appliquent à presque tous les domaines de la vie économique et quotidienne moderne. Les normes techniques régissent de nombreux objets matériels et immatériels tels que les produits, les procédés, les méthodes de mesure, les processus et les services et sont utilisées dans presque tous les secteurs et domaines spécialisés.[248]

5.2.11.2 Le système de normalisation de l'UE

Le système des organismes européens de normalisation est constitué du Comité européen de normalisation (CEN), du Comité européen de normalisation électrotechnique (CENELEC) et de l'Institut européen des normes de télécommunications (ETSI). Le CEN est responsable des normes européennes (EN) dans tous les domaines techniques, à l'exception de l'électrotechnique et des télécommunications. Le CENELEC est responsable de la normalisation européenne dans le domaine de l'électrotechnique (marquage ENEC) et l'ETSI dans le domaine des télécommunications.

Le concept de « nouvelle approche » introduit par l'UE en 1985 pour la réglementation des produits a établi un lien entre la législation et la normalisation. Le législateur se contente de formuler des exigences essentielles et des objectifs de protection qui sont ensuite concrétisés techniquement par des normes. Cette approche vise d'une part à alléger l'appareil législatif de l'UE et d'autre part à permettre de s'adapter à l'état de la technique et des connaissances.

Ces normes harmonisées sont élaborées par les organismes européens de normalisation (CEN, CENELEC et ETSI) sur la base d'un mandat délivré par la Commission européenne et publiées au Journal officiel de l'UE. Une fois qu'elles sont publiées au Journal officiel de l'UE, elles déploient une présomption de conformité lors de leur application. Ainsi, la normalisation joue un rôle essentiel pour la mise en œuvre du règlement de l'UE.

---

[247] Règlement (UE) 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision no 1673/2006/CE du Parlement européen et du Conseil, JO L 316/12, 14 novembre 2012.

[248] SECO, Förderung der Normungsorganisationen im Bereich der Digitalisierung : Akteure und Erkenntnisse in ausgewählten Themenbereichen, Bericht an den Bundesrat, 16 août 2022, 6, disponible sous www.admin.ch > Documentation > Communiqués > Avis de publication > Rapport « Soutien aux organismes de normalisation dans le domaine de la numérisation : acteurs et résultats dans les domaines sélectionnés (consulté le 26 août 2024) ; pour un résumé des normes existantes au niveau international et européen, voir MÉLANIE GORNET/WINSTON MAXWELL, Normes techniques et éthique de l'IA. CNIA 2023 – Conférence Nationale en Intelligence Artificielle, Strasbourg 2023, disponible sous https://hal.science/hal-04121843 (consulté le 26 août 2024).

5.2.11.3 La présomption de conformité

5.2.11.3.1 En général

Les produits fabriqués conformément aux normes harmonisées bénéficient d'une présomption de conformité aux exigences essentielles correspondantes visées dans la réglementation applicable. Cette procédure est également appelée « procédure d'évaluation de la conformité » et conduit, si le résultat est positif, à l'autorisation de signer une « déclaration de conformité » et d'apposer le marquage CE sur le produit.

Conformément à la directive ou au règlement de l'UE applicable en l'espèce, le responsable de la mise sur le marché peut effectuer la procédure d'évaluation de la conformité dans le cadre de l'autocontrôle ou bien faire appel à des organismes de contrôle accrédités (Certified Bodies) et à des organismes notifiés (Notified Bodies). En cas de besoin, d'autres normes peuvent également être prises en compte ; dans ce cas, il faut toutefois apporter la preuve que ces normes permettent également de satisfaire aux « exigences essentielles de sécurité et de santé ». Pour les autres solutions, l'état de la technique et des connaissances doit être démontré.

5.2.11.3.2 Dans le règlement sur l'intelligence artificielle

Selon l'art. 40, par. 1, du règlement sur l'IA, les systèmes d'IA à haut risque ou les modèles GPAI conformes à des normes harmonisées ou à des parties de normes harmonisées dont les références ont été publiées au Journal officiel de l'UE conformément au règlement 1025/2012[249] sont présumés conformes aux exigences visées à la Section 2 du Chapitre III, en lien avec les systèmes d'IA à haut risque (art. 8 ss du règlement sur l'IA) ou, le cas échéant, aux obligations énoncées au Chapitre V, Sections 2 et 3, du règlement, en lien avec les modèles GPAI, dans la mesure où ces exigences ou obligations sont couvertes par ces normes.

Ainsi, cette disposition établit une présomption de conformité aux exigences du règlement lorsque les normes harmonisées sont respectées.

Afin que ces normes harmonisées soient élaborées, la Commission a présenté une demande de normalisation aux organisations européennes de normalisation (cf. ch. 5.2.11.4).

5.2.11.4 Les travaux de normalisation dans le domaine de l'intelligence artificielle

Le 22 mai 2023, la Commission a adressé au CEN et au CENELEC une demande[250] visant l'élaboration des normes techniques nécessaires pour mettre en œuvre le règlement sur l'IA.

---

[249] Règlement (UE) 1025/2012 relatif à la normalisation européenne (n. 247).

[250] Décision d'exécution de la Commission du 22 mai 2023 relative à une demande de normalisation adressée au Comité européen de normalisation et au Comité européen de normalisation électrotechnique à l'appui de la politique de l'Union en matière d'intelligence artificielle, C(2023) 3215 final, disponible sous https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en (consulté le 26 août 2024).

Concrètement, dans cette demande, la Commission européenne a demandé à ces deux organismes d'élaborer dix nouvelles normes européennes dans le domaine de l'IA jusqu'au 30 avril 2025. Les normes concernent les dix domaines suivants :

- Gestion des risques des systèmes d'IA
- Gouvernance et qualité des jeux de données utilisés pour élaborer des systèmes d'IA
- Tenue de registres et capacité de journalisation des systèmes d'IA
- Transparence et fourniture d'informations aux utilisateurs des systèmes d'IA
- Contrôle humain des systèmes d'IA
- Spécifications d'exactitude applicables aux systèmes d'IA
- Spécifications de robustesse applicables aux systèmes d'IA
- Spécifications de cybersécurité applicables aux systèmes d'IA
- Systèmes de gestion de la qualité mis en place par les fournisseurs de systèmes d'IA, y compris les processus de surveillance après commercialisation
- Évaluation de la conformité des systèmes d'IA

Les dix normes techniques évoquées ci-dessus sont élaborées par des experts et des expertes au sein d'un comité technique mixte : le *Joint CEN-CENELEC Technical Committee 21 'Artificial Intelligence' (CEN-CLC/JTC 21)*.

Si ce Comité ne parvient pas à édicter les dix normes demandées par la Commission européenne dans le délai échéant le 30 avril 2025, la Commission pourra adopter des spécifications communes, conformément à l'art. 41 du règlement. Ce mécanisme permet ainsi à la Commission de maintenir une certaine pression sur les organismes de normalisation.

5.2.12 Mesures de soutien à l'innovation

Le règlement sur l'IA s'efforce de concilier l'innovation avec la sécurité des systèmes d'IA, en mettant sur pied un régime de bacs à sable réglementaires de l'IA (regulatory sandboxes), soit des cadres juridiques qui permettent des essais limités d'innovations sous contrôle réglementaire.[251]

Les États membres sont tenus d'établir au moins un bac à sable réglementaire pour l'IA au niveau national qui devait être opérationnel deux ans après l'entrée en vigueur du règlement. Les États membres peuvent aussi s'acquitter de cette obligation en participant à un bac à

[251] Pour une critique sur l'efficacité de ces mesures, cf. DAVID ROSENTHAL, Der EU AI Act (n. 221), 4 N 5.

sable existant dans la mesure où cette participation assure un niveau équivalent de couverture nationale pour les États membres participants (cf. art. 57).

L'objectif de ces bacs à sable est de fournir un environnement contrôlé qui favorise l'innovation et facilite le développement, l'entraînement, la mise à l'essai et la validation de systèmes d'IA innovants pendant une période limitée avant leur mise sur le marché ou leur mise en service (cf. art. 57, par. 5). Les autorités nationales compétentes doivent informer le Bureau de l'IA et le Comité (cf. ch. 5.2.13) de l'établissement d'un bac à sable réglementaire. Le Bureau de l'IA publie une liste des bacs à sable prévus ou existants (art. 57, par. 15).

Le rôle des autorités nationales compétentes est de fournir des orientations, une supervision et un soutien au sein du bac à sable visant à identifier les risques pour les droits fondamentaux, la santé et la sécurité, les mesures d'atténuation et leur efficacité par rapport aux obligations et exigences du règlement. À la demande des fournisseurs ou fournisseurs potentiels, l'autorité compétente peut fournir un rapport de sortie détaillant les activités menées dans le bac à sable et les résultats correspondants. Les fournisseurs peuvent utiliser cette documentation pour démontrer le respect du règlement dans le cadre du processus d'évaluation de la conformité (cf. art. 57, par. 7).

Les fournisseurs et les fournisseurs potentiels dans le bac à sable réglementaire restent responsables, en vertu de la législation communautaire et des États membres, pour tout préjudice causé à des tiers du fait de l'expérimentation menée dans le bac à sable. Toutefois, si le fournisseur a respecté le plan et les conditions de participation et a suivi de bonne foi les orientations données par l'autorité nationale compétente, aucune amende administrative n'est imposée par les autorités en cas d'infraction au règlement sur l'IA (art. 57, par. 12). Le règlement donne ainsi une incitation forte à participer à un bac à sable réglementaire.

Enfin, le règlement prévoit la possibilité, à certaines conditions, de tester des systèmes d'IA à haut risque dans des conditions réelles, dans et en dehors d'un bac à sable réglementaire (art. 57, 58 et 60 s. du règlement). L'autorisation de mener un test dans des conditions réelles doit être donnée par l'autorité de surveillance du marché (cf. art. 60).

5.2.13 Gouvernance

Le règlement établit un cadre de gouvernance à plusieurs niveaux qui vise à coordonner et à soutenir l'application du règlement au niveau national, à renforcer les capacités au niveau de l'UE, et à intégrer les parties prenantes dans le domaine de l'IA.

Au niveau de l'UE, le règlement désigne les organes de gouvernance suivants :

- Bureau de l'IA (*AI Office*, art. 64) : établi au sein de la Commission européenne, ce bureau a pour fonction de contribuer à la mise en œuvre, au suivi et à la surveillance des systèmes d'IA et de modèles GPAI, et de la gouvernance de l'IA (cf. art. 3, point 47). Il met en particulier à disposition des modèles qui aident dans le cadre de l'application du règlement (cf. par exemple l'art. 27 par. 5, qui prévoit que le Bureau de l'IA élabore un modèle de questionnaire, y compris au moyen d'un outil automatisé, afin d'aider les déployeurs à se conformer à l'obligation de conduire une analyse d'impact).

- Comité européen de l'intelligence artificielle (*European AI Board*, art. 65) : composé de représentants des États membres, il a pour tâche de conseiller et assister la Commission et les États membres afin de faciliter l'application cohérente et efficace du règlement. Le Contrôleur européen de la protection des données participe en qualité d'observateur.

- Forum consultatif (*Advisory Forum*, art. 67) : il fournit une expertise technique et conseille le Comité de l'IA et le Bureau de l'IA. Il a pour fonction de garantir la participation des parties prenantes à la mise en œuvre et à l'application du règlement. Le forum est constitué, entre autres, de parties prenantes de l'industrie, des universités, de la société civile, ainsi que de l'Agence des droits fondamentaux, de l'Agence de l'UE pour la cybersécurité, du CEN, du CENELEC et de l'ETSI.

- Groupe scientifique d'experts indépendants (*Scientific Panel of Independent Experts,* art. 68) : réunit des experts issus de la communauté scientifique afin de soutenir la mise en œuvre et l'application du règlement, en particulier les activités de contrôle du Bureau de l'IA en ce qui concerne les modèles GPAI (cf. ch. 5.2.10). Les États membres peuvent demander l'aide du groupe scientifique pour leurs activités de contrôle de l'application du règlement (art. 69).

Au niveau national, les États membres établissent ou désignent en tant qu'autorités compétentes au moins une autorité notifiante et au moins une autorité de surveillance du marché (art. 70) :

- Autorité notifiante : il s'agit d'une autorité nationale chargée de mettre en place et d'accomplir les procédures nécessaires à l'évaluation, à la désignation et à la notification des organismes d'évaluation de la conformité et à leur contrôle, à savoir les organismes chargés de l'évaluation de la conformité des systèmes d'IA à haut risque (cf. art. 3, points 19 à 22 ; ch. 5.2.7.4).

- Autorité de surveillance du marché : il s'agit de l'autorité nationale assurant la mission et prenant les mesures prévues par le règlement UE 2019/1020 sur la surveillance du marché et la conformité des produits[252] (cf. art. 3, point 26) (cf. ch. 5.2.14).

  Chaque État membre devra désigner une autorité de surveillance du marché qui fera office de point de contact unique pour le règlement (art. 70, par. 2).

[252] Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011, JO L 169/1, 25.6.2019.

5.2.14 Surveillance et contrôle de l'application

Les États membres jouent un rôle clé dans l'application et le contrôle du respect du règlement. À cet égard, chaque État membre désigne une ou plusieurs autorités nationales compétentes pour surveiller l'application et la mise en œuvre des règles, ainsi que pour mener des activités de surveillance du marché.

Le règlement sur l'IA renvoie au règlement UE 2019/1020[253] en ce qui concerne la surveillance (cf. art. 74). Il contient toutefois des précisions importantes, notamment en matière d'organisation et de coordination des autorités chargées de surveillance du marché. Par exemple :

- Pour les systèmes d'IA à haut risque liés à des produits couverts par la législation d'harmonisation de l'UE mentionnés à l'annexe I, section A, l'autorité de surveillance du marché selon le règlement sur l'IA est l'autorité responsable des activités de surveillance désignée en vertu de ces actes juridiques d'harmonisation (art. 74, par. 3).

- Pour les systèmes d'IA utilisés dans des établissements financiers, l'autorité de surveillance du marché est l'autorité nationale responsable de la surveillance financière de ces établissements en vertu de la législation de l'Union sur les services financiers, dans la mesure où il y a un lien avec la fourniture de ces services (art. 74, par. 6).

- S'agissant des systèmes d'IA à haut risque énumérés au point 1 de l'annexe III (biométrie), dans la mesure où ils sont utilisés à des fins répressives, de gestion des frontières et de justice et démocratie, et pour les systèmes d'IA à haut risque énumérés à l'annexe III, point 6, 7 et 8, les États membres désignent en tant qu'autorité de surveillance les autorités compétentes en matière de protection des données (art. 74, par. 8).[254]

En principe, les fournisseurs accordent aux autorités de surveillance du marché un accès complet à la documentation ainsi qu'aux jeux de données d'entraînement, de validation et de test utilisés pour le développement des systèmes d'IA à haut risque (art. 74, par. 12). À certaines conditions, les autorités de surveillance peuvent se voir accorder l'accès au code source d'un système d'IA à haut risque (cf. art. 74, par. 13).

Le règlement sur l'IA précise les pouvoirs d'intervention et les procédures que les autorités de surveillance pourront suivre en fonction des circonstances (art. 79 ss). Par exemple,

---

[253] Règlement (UE) 2019/1020 sur la surveillance du marché et la conformité des produits (n. 252).

[254] Cf. DAVID ROSENTHAL, Der EU AI Act (n. 221), 6 N 9, qui estime qu'au-delà des situations visées par l'art. 74, par. 8, du règlement sur l'IA, les Etats membres désigneront, dans environ la moitié des cas, les autorités de protection des données en tant qu'autorités de surveillance. Voir aussi la déclaration du Comité européen de la protection des données du 16 juillet 2024, par laquelle les autorités de protection des données souhaitent être chargées de l'application du règlement sur l'IA en ce qui concerne les systèmes d'IA à haut risque, disponible sous https://www.edpb.europa.eu/news/news/2024/edpb-adopts-statement-dpas-role-ai-act-framework-eu-us-data-privacy-framework-faq_en (consulté le 26 août 2024).

lorsqu'un système d'IA présente un risque pour la santé, la sécurité ou pour les droits fondamentaux des personnes, l'autorité de surveillance procède elle-même à une évaluation de la conformité avec l'ensemble des exigences énoncées dans le règlement. Si elle constate un non-respect des conditions, elle invite l'opérateur concerné à prendre toutes les mesures correctives appropriées (art. 79).

Dans une sorte de clause générale, le règlement donne en outre à l'autorité de surveillance du marché la compétence d'ordonner des mesures particulières à tout système d'IA qui, bien que conforme à la réglementation, présente un risque pour la santé ou la sécurité des personnes, pour les droits fondamentaux ou pour d'autres aspects relatifs à la protection de l'intérêt public (cf. art. 82, par. 1).[255]

Pour ce qui est des institutions, agences et organes de l'Union relevant du champ d'application du présent règlement, le Contrôleur européen de la protection des données est désigné en tant qu'autorité compétente pour la surveillance du marché (art. 74, par. 9 et consid. 156).

S'agissant des modèles GPAI, la Commission, via le Bureau de l'IA, dispose de pouvoirs exclusifs pour surveiller et contrôler le respect du règlement (cf. art. 88).

Finalement, le règlement sur l'IA accorde certains pouvoirs également aux autorités de protection des droits fondamentaux. Selon l'art. 77, par. 1, lorsque cela est nécessaire à l'accomplissement effectif de leur mandat dans les limites de leurs compétences, les autorités ou organismes publics nationaux qui supervisent ou font respecter les obligations du droit de l'UE visant à protéger les droits fondamentaux sont habilitées à demander toute documentation en lien avec l'utilisation des systèmes d'IA à haut risque visés à l'annexe III. Les États membres établissent une liste de ces autorités ou organismes (art. 77, par. 2). Lorsque cela est nécessaire pour déterminer s'il y a eu violation des obligations, l'autorité ou organisme peut requérir l'organisation de tests du système d'IA à haut risque (cf. art. 77, par. 3).

5.2.15 Droits individuels

Dans une section séparée intitulée « Voies de recours » (art. 85 s.), le règlement sur l'IA contient deux dispositions accordant des droits individuels :

- Droit à la réclamation : toute personne physique ou morale ayant des motifs de considérer qu'il y a eu violation des dispositions du règlement peut déposer une réclamation auprès de l'autorité de surveillance du marché concernée. Ces réclamations sont prises en compte conformément au règlement UE 2019/1020[256] et à l'exercice des activités de surveillance. Ce droit est sans préjudice d'autres recours administratifs ou judiciaires (art. 85).

- Droit à l'explication des décisions individuelles : toute personne concernée a le droit d'obtenir du déployeur d'un système d'IA des explications claires et pertinentes sur le

[255] DAVID ROSENTHAL, Der EU AI Act (n. 221), 30 N 59.

[256] Règlement (UE) 2019/1020 sur la surveillance du marché et la conformité des produits (n. 252).

rôle du système d'IA dans la procédure décisionnelle et sur les principaux éléments de la décision prise, aux conditions suivantes (cf. art. 86, par. 1) :

- il s'agit d'un système d'IA à haut risque mentionné à l'annexe III, à l'exception de ceux énumérés au point 2 (infrastructures critiques) ;
- la décision a été prise par le déployeur sur la base des sorties de ce système ;
- la décision produit des effets juridiques ou affecte significativement la personne concernée de façon similaire d'une manière qu'elle considère comme ayant des conséquences négatives sur sa santé, sa sécurité ou ses droits fondamentaux.

Ce droit ne peut pas être invoqué si des exceptions ou restrictions sont prévues par le droit de l'UE ou dans le droit national. Finalement, l'art. 86 ne s'applique que dans la mesure où le droit visé au par. 1 n'est pas prévu ailleurs dans le droit de l'UE.

5.2.16 Sanctions

Le règlement prévoit des sanctions financières importantes en cas de violation des obligations prévues par le règlement IA.

Ainsi, notamment, les infractions concernant le non-respect des interdictions des pratiques d'IA sont passibles d'amendes administratives pouvant aller jusqu'à 35 millions d'euros ou, si l'auteur est une entreprise, jusqu'à 7 % du chiffre d'affaires annuel mondial total réalisé au cours de l'exercice précédent, le montant le plus élevé étant retenu (cf. art. 99, par. 3).

La non-conformité avec les dispositions relatives aux opérateurs ou aux organismes notifiés (cf. p. ex. les obligations découlant de l'art. 16 pour les fournisseurs) est passible d'amendes administratives pouvant aller jusqu'à 15 millions d'euros ou, si l'auteur de l'infraction est une entreprise, jusqu'à 3 % du chiffre d'affaires annuel mondial total réalisé au cours de l'exercice précédent, le montant le plus élevé étant retenu (cf. art. 99, par. 4). Les fournisseurs de modèles d'IA à usage général peuvent se voir infliger des amendes de même ampleur (cf. art. 101).

La fourniture d'informations inexactes, incomplètes ou trompeuses aux organismes notifiés ou aux autorités nationales compétentes en réponse à une demande est passible d'amendes administratives pouvant aller jusqu'à 7,5 millions d'euros ou, si le contrevenant est une entreprise, jusqu'à 1 % du chiffre d'affaires mondial pour l'exercice précédent, le montant le plus élevé étant retenu (cf. art. 99, par. 5).

Dans le cas des PME, y compris les start-up, les amendes précitées s'élèvent au maximum aux pourcentages ou montants visés à l'art. 99, par. 3, 4 et 5, le chiffre le plus faible étant retenu (cf. art. 99, par. 6).

5.2.17 Entrée en vigueur et application

Le règlement sur l'IA sera applicable deux ans après son entrée en vigueur, qui a eu lieu le 1er août 2024 (art. 113). Le règlement s'appliquera directement à tous ses destinataires (fournisseurs, déployeurs, personnes concernées, etc.) sans avoir à être transposé au préalable

dans la législation des États membres (art. 288 TFUE).[257] Ce délai de deux ans souffre d'exceptions. Par exemple :

- Les Chapitres I (Dispositions générales) et II (Pratiques interdites) seront applicables déjà 6 mois après l'entrée en vigueur, à savoir à partir du 2 février 2025.

- Les règles relatives aux modèles GPAI, à l'exception de l'art. 101, seront applicables déjà 12 mois après l'entrée en vigueur, à savoir à partir du 2 août 2025. Il en va de même des règles relatives aux organismes pouvant délivrer des déclarations de conformité, de celles sur la création des autorités et des organismes prévus dans le règlement (par exemple le Bureau de l'IA) et des dispositions relatives aux sanctions.

- L'art. 6, par. 1, et les obligations correspondantes du règlement (systèmes à haut risque dans le cadre de la législation harmonisée de l'UE sur les produits) sera applicable seulement 36 mois après l'entrée en vigueur, à savoir dès le 2 août 2027.

5.3 Appréciation

5.3.1 Effets juridiques sur les opérateurs suisses

5.3.1.1 Opérateurs concernés

Comme mentionné ci-devant (cf. ch. 5.2.5), le règlement sur l'IA s'applique aux fournisseurs établis ou situés dans l'UE *ou dans un pays tiers*, qui mettent sur le marché ou mettent en service un système d'IA, ou qui mettent sur le marché des modèles GPAI dans l'UE (art. 2, par. 1, point a). La mise sur le marché et la mise en service ne concernent, d'après les définitions de ces termes (cf. art. 3, points 9 et 11), que la première mise à disposition, respectivement la première utilisation dans le marché de l'UE.

Selon l'art. 2, par. 1, point c, le règlement sur l'IA s'applique aussi aux fournisseurs et déployeurs de systèmes d'IA qui ont leur lieu d'établissement ou sont situés *dans un pays tiers*, lorsque les sorties produites par le système d'IA sont utilisées dans l'UE.

Ces dispositions visent à empêcher que la réglementation soit contournée en développant des systèmes d'IA dans un pays tiers, alors que le résultat généré par le système d'IA est finalement utilisé dans l'UE.

Une partie de la doctrine relève cependant que l'interprétation des art. 2, par. 1, points a et c n'est pas sans poser quelques difficultés : dans la mesure où la notion même de fournisseur implique la mise sur le marché dans l'UE (cf. art. 3, points 3 et 9), cela signifie qu'un fournisseur qui ne met pas le système d'IA sur le marché de l'UE n'est pas considéré comme fournisseur au sens du règlement, même si son système d'IA trouve finalement son chemin dans l'UE ou y produit des effets. Dès lors, l'art. 2, par. 1, point c, qui soumet au règlement tout fournisseur basé dans un pays tiers lorsque les résultats produits par son système d'IA sont

[257] DAVID ROSENTHAL, Der EU AI Act (n. 221), 2 N 2.

utilisés dans l'UE, élargirait *de facto* la notion de fournisseur. Dans ce dernier cas, l'application du règlement est donnée du simple fait que le résultat est utilisé dans l'UE.[258] Il s'agira d'observer comment ces deux dispositions vont s'articuler à l'avenir. Il apparaît en tout état de cause que la volonté du législateur européen tend plutôt vers une application large du règlement.

Concernant les déployeurs situés dans des pays tiers, le critère est l'utilisation du résultat dans l'UE (cf. art. 2, par. 1, point c). Ainsi, on peut par exemple citer le cas d'une entreprise suisse qui envoie des textes générés par un système d'IA à ses clients dans l'UE.[259]

Les considérants du règlement sur l'IA citent également le cas de figure dans lequel un opérateur établi dans l'UE confie à un opérateur externe établi dans un pays tiers la tâche d'exécuter certains services ayant trait à une activité devant être réalisée par un système d'IA qui serait considéré comme étant à haut risque. Dans ces circonstances, le système d'IA utilisé dans un pays tiers par l'opérateur pourrait traiter des données légalement collectées et transférées depuis l'UE, et fournir à l'opérateur contractant établi dans l'UE les sorties dudit système d'IA provenant de ce traitement. Le règlement sur l'IA devrait donc s'appliquer dans ce cas de figure (cf. consid. 22).

5.3.1.2 Effets

Une fois concernés par l'application du règlement, les opérateurs suisses devront se conformer à ses obligations en fonction du système d'IA ou modèle GPAI en cause.

Si le système d'IA est à risque minimal, il n'y a pas d'obligations, mais tout au plus une application volontaire des codes de conduite (cf. ch. 5.2.9). Si le système d'IA tombe dans un des cas visés à l'art. 50, les obligations de transparence correspondantes devront être respectées (cf. ch. 5.2.8).

Pour les systèmes d'IA à haut risque, les obligations de la Section 2, Chapitre III, doivent être observées (cf. ch. 5.2.7). En particulier, les fournisseurs suisses de systèmes d'IA à haut risque devront effectuer l'évaluation de la conformité de leurs produits. Il est renvoyé aux développements ci-dessous en lien avec l'ARM pour d'autres développements sur ce point (cf. ch. 5.3.2).

Les systèmes d'IA à risque inacceptable sont tout simplement interdits (sauf exceptions) (cf. ch. 5.2.6). S'agissant des modèles GPAI, les fournisseurs se réfèreront aux obligations découlant des art. 51 ss du règlement sur l'IA (cf. ch. 5.2.10).

À noter qu'au surplus, pour les systèmes d'IA à haut risque et les modèles GPAI, les fournisseurs établis en Suisse devront, avant de mettre ces systèmes et modèles sur le marché de l'UE, désigner, par mandat écrit, un mandataire établi dans l'UE (cf. art. 22 et 54 du règle-

---

[258] DAVID ROSENTHAL, Der EU AI Act (n. 221), 9 N 24 ; voir aussi MARTINA ARIOLI, Risikomanagement (n. 222), 15 N 50 s.

[259] Voir pour d'autres exemples DAVID ROSENTHAL, Der EU AI Act (n. 221), 14 N 32.

ment sur l'IA). Pour les systèmes d'IA à haut risque, le mandataire devra en particulier s'assurer que le système d'IA est enregistré sur la base de données de l'UE conformément à l'art. 49.

Sur demande motivée des autorités compétentes, respectivement du Bureau de l'IA, le mandataire doit fournir toutes les informations nécessaires pour démontrer la conformité d'un système d'IA à haut risque ou d'un modèle GPAI (cf. art. 22, par. 3, point c et d, et 54, par. 3, point c et d du règlement sur l'IA). En outre, il devra coopérer avec elles à toute mesure qui serait prise. À noter que si le mandataire d'un fournisseur établi en Suisse recevait des demandes de production d'informations sises en Suisse, et transmettait ces informations aux autorités européennes compétentes sans y être autorisé, alors que cela relève de la compétence des pouvoirs publics suisses, il se poserait la question d'une éventuelle violation de l'art. 271 CP. Cette disposition réprime les actes exécutés sans droit pour un État étranger.

Le règlement prévoit en outre que le mandataire met fin au mandat s'il considère ou a des raisons de considérer que le fournisseur agit de manière contraire aux obligations qui lui incombent en vertu du règlement.

Les fournisseurs de systèmes d'IA à haut risque indiquent en outre sur le système d'IA ou, lorsque cela n'est pas possible, sur son emballage ou dans la documentation l'accompagnant, selon le cas, leur nom, raison sociale ou marque déposée, l'adresse à laquelle ils peuvent être contactés (cf. art. 16, point b).

5.3.2 Relations avec l'Accord entre la Suisse et l'UE relatif à la reconnaissance mutuelle en matière d'évaluation de la conformité[260]

5.3.2.1 Fonctionnement de l'accord

L'Accord entre la Suisse et l'UE relatif à la reconnaissance mutuelle en matière d'évaluation de la conformité (ARM)[261] vise à assurer la reconnaissance mutuelle des évaluations de conformité pour la commercialisation de produits industriels (ex. machines, dispositifs médicaux, ascenseurs) dans des secteurs pour lesquels il existe des prescriptions harmonisées en droit de l'UE et pour lesquels il existe une évaluation de conformité obligatoire.[262] L'accord permet un accès facilité au marché intérieur de l'UE dans les secteurs de produits qu'il couvre et réduit les délais et les coûts liés à la commercialisation des produits. Plus concrètement, l'ARM contribue à réduire les entraves techniques au commerce, d'une part grâce à l'harmonisation entre les prescriptions suisses et européennes, et d'autre part grâce à la mise en place d'une seule évaluation de conformité pour accéder au marché de l'UE, respectivement de la Suisse, effectuée sur la base des prescriptions techniques suisses ou européennes par l'un des organismes d'évaluation de la conformité reconnus dans l'accord. Ce dernier est basé

[260] Ce chapitre a été rédigé sur la base de contributions de l'OFCOM.

[261] Accord entre la Confédération suisse et la Communauté européenne relatif à la reconnaissance mutuelle en matière d'évaluation de la conformité, RS **0.946.526.81**.

[262] Message relatif à l'approbation des accords sectoriels entre la Suisse et la CE, FF **1999** 5440, 5521.

sur l'équivalence des législations de la Suisse et de l'UE. Lors de chaque révision substantielle de la législation technique de l'UE dans un secteur couvert par l'accord, il faut adapter la législation suisse de manière à maintenir l'équivalence avec celle de l'UE et mettre à jour le chapitre correspondant de l'accord.

Néanmoins, depuis quelques années, l'UE ne consent plus à une mise à jour de l'ARM, faisant référence aux questions institutionnelles dans les relations entre la Suisse et l'UE. En 2017, l'UE a édicté une nouvelle réglementation sur les dispositifs médicaux qui est entrée en application en mai 2021. La Suisse a adopté une législation équivalente à celle de l'UE. Toutefois l'UE refuse depuis le 26 mai 2021 d'actualiser l'ARM pour les dispositifs médicaux faute de progrès concernant les négociations sur les questions institutionnelles. En conséquence, la Suisse ne bénéficie plus de la reconnaissance mutuelle en matière d'évaluation de la conformité et les fournisseurs suisses de dispositifs médicaux rencontrent de multiples entraves pour accéder au marché de l'UE, parmi lesquelles la nécessité de faire procéder à l'évaluation de la conformité des produits par un organisme d'évaluation de la conformité de l'UE. À l'avenir, ce blocage pourrait affecter d'autres secteurs de l'ARM qui ont fait ou font actuellement l'objet d'une révision majeure dans l'UE (machines, produits de construction, jouets). Cela dépendra toutefois de l'issue des négociations actuelles en cours entre la Suisse et l'UE.

5.3.2.2 Impact du règlement sur l'IA sur les opérateurs suisses concernés par l'accord

Le règlement sur l'IA règle spécifiquement les aspects liés à l'IA, notamment dans la majorité des secteurs de produits couverts par l'ARM. Il s'applique en plus des législations sectorielles sur les produits, en introduisant des exigences spécifiques supplémentaires liées à l'IA.

En effet, le règlement définit comme « systèmes d'IA à haut risque » les systèmes d'IA destinés à être utilisés en tant que composants de sécurité des produits ou constituant eux-mêmes des produits couverts par les législations listées à son annexe I – dont par exemple les machines ou composants de sécurité des machines – et qui font l'objet d'une évaluation de la conformité par un organisme tiers (cf. art. 6, al. 1).[263]

Les secteurs de l'ARM correspondant aux domaines couverts à l'annexe I, section A du règlement sur l'IA, sont les suivants :

- Machines (chapitre 1 de l'ARM)

[263] Il convient de souligner que la législation fait une distinction entre les législations d'harmonisation relevant du nouveau cadre législatif (New Legislative Framework, NFL, voir sur celui-ci MARTINA ARIOLI, Risikomanagement [n. 222)], 4 N 8 s.), dans la section A de l'annexe I du règlement, et des législations relevant de l'ancienne approche, dans la section B de l'annexe I du règlement. Le règlement sur l'IA s'applique directement aux domaines couverts dans la section A de l'annexe I, tandis que pour les domaines relevant de l'ancienne approche, lorsque la Commission adopte des actes délégués ou des actes d'exécution en ce qui concerne des systèmes d'IA qui sont des composants de sécurité au sens du règlement sur l'IA, elle devra tenir compte des exigences applicables aux systèmes à haut risque. Une grande partie des domaines couverts par l'ARM tombent dans la section A, tandis que deux domaines couverts par l'ARM ainsi que l'aviation civile et l'interopérabilité du système ferroviaire relevant des accords bilatéraux sur les transports terrestres et aérien sont dans la section B de l'annexe I.

- Équipements de protection individuelle (chapitre 2 de l'ARM)
- Jouets (chapitre 3 de l'ARM)
- Dispositifs médicaux (chapitre 4 de l'ARM)
- Appareils à gaz (chapitre 5 de l'ARM)
- Appareils à pression (chapitre 6 de l'ARM)
- Équipements de télécommunications (chapitre 7 de l'ARM)
- Appareils et systèmes de protection destinés à être utilisés en atmosphère explosible (ATEX) (chapitre 8 de l'ARM)
- Ascenseurs (chapitre 17 de l'ARM)
- Installations à câbles (chapitre 19 de l'ARM)

Les secteurs de l'ARM correspondant aux domaines couverts à l'annexe I, section B du règlement sur l'IA, sont les suivants :

- Véhicules à moteur (chapitre 12 de l'ARM)
- Tracteurs agricoles ou forestiers (chapitre 13 de l'ARM)

Le règlement sur l'IA s'applique ainsi à la majorité des secteurs couverts par l'ARM qui sont ainsi soumis aux dispositions applicables aux systèmes d'IA à haut risque, et ce, trois ans après l'entrée en vigueur du règlement (date à laquelle l'art. 6, par. 1, du règlement sur l'IA et les obligations correspondantes trouveront application, cf. ch. 5.2.17). En effet, dans les domaines relevant de la législation d'harmonisation de l'UE, listés à l'annexe I du règlement, le respect des exigences relatives aux systèmes d'IA à haut risque s'intégrera aux procédures d'évaluation de conformité prévues par ces législations (cf. ch. 5.2.7.1 et 5.2.7.4.1). Les obligations en vertu du règlement sur l'IA s'ajouteront ainsi aux obligations sectorielles existantes.

Pour les fournisseurs suisses de produits d'IA couverts à la fois par l'ARM et la législation d'harmonisation de l'UE visée à l'annexe I du règlement sur l'IA (systèmes d'IA à haut risque selon l'art. 6, par. 1, du règlement sur l'IA), le processus d'évaluation de la conformité se fera selon la procédure prévue par l'ARM, mais devra aussi inclure l'évaluation de la conformité avec les exigences pour les systèmes à haut risque prévues par le règlement sur l'IA. Dans la situation actuelle, en l'absence d'une législation suisse en matière d'IA et tant que les dispositions du règlement sur l'IA relatives aux produits visées ci-dessus n'auront pas été incluses dans l'ARM, l'évaluation de la conformité quant au respect des exigences du règlement sur l'IA devra être effectuée par un organisme d'évaluation de la conformité de l'UE et selon le droit de l'UE. Pour mettre un tel produit ou composant de produit sur le marché de l'UE, les opérateurs suisses devront en outre désigner un mandataire dans l'UE qui aura notamment pour tâche de vérifier que le processus d'évaluation de la conformité a été effectué (cf. art. 22 et 54 du règlement sur l'IA). En outre, les importateurs (qui, selon la définition du règlement sur l'IA, sont établis dans l'UE) doivent indiquer leurs coordonnées sur l'emballage (art. 23). Il en résulte donc de nouvelles entraves techniques au commerce.

5.3.2.3 Extension possible de l'ARM

Deux étapes seraient nécessaires pour éliminer complètement ces obstacles techniques au commerce dans le cadre de l'ARM. Premièrement, la Suisse devrait harmoniser sa législation

sur les produits avec celle de l'UE, c'est-à-dire en l'occurrence avec la législation pertinente du règlement sur l'IA. Deuxièmement, l'ARM devrait être élargi en conséquence.

L'UE et la Suisse pourraient donc décider d'élargir l'ARM afin d'y incorporer la législation en matière d'IA relative aux produits. Cela permettrait une reconnaissance mutuelle des futures évaluations de conformité des produits comprenant des systèmes d'IA. Cette option est néanmoins tributaire de l'issue des négociations entre la Suisse et l'UE actuellement.

Même si la Suisse devait introduire une législation en matière d'IA se rapprochant de celle de l'UE, tant que l'ARM n'est pas mis à jour, il n'y aura pas de reconnaissance des procédures d'évaluation de la conformité relative aux aspects d'IA réalisés en Suisse.

5.3.3 Relations avec la décision d'adéquation de la Commission européenne en matière de protection des données

Le transfert de données personnelles d'un État de l'UE vers un État tiers est, sauf exception, subordonné à la condition que les données soient protégées de manière similaire dans l'État de destination. Ce niveau de protection peut être assuré par un traité, des garanties ad hoc (p. ex des clauses contractuelles types) ou encore une décision d'adéquation (voir art. 45 RGPD et art. 36 directive UE 2016/680[264]).[265] Cette décision atteste que l'État de destination fournit un niveau de protection considéré comme équivalent. Dans ce cadre, la Commission européenne évalue en détail la situation juridique dans l'État concerné, en particulier la législation pertinente et sa mise en œuvre, à l'aune des exigences du droit de l'UE.

La Suisse est au bénéfice d'une décision d'adéquation de la Commission européenne depuis 2000.[266] Cette dernière a été confirmée en janvier 2024, avec celles de 10 autres pays[267], après un réexamen de plusieurs années. Cette décision concerne le transfert de données dans les secteurs privé et public, à l'exclusion des transferts dans un cadre répressif. Cette décision est très importante, car elle autorise le transfert de données personnelles sans garanties supplémentaires. Elle facilite les échanges économiques.

---

[264] Directive (UE) 2016/680 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel dans le domaine répressif (n. 234).

[265] La liste des pays au bénéfice d'une décision d'adéquation est disponible ici : https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en (consulté le 26 août 2024).

[266] Décision de la Commission du 26 juillet 2000 relative à la constatation, conformément à la directive 95/46/CE du Parlement européen et du Conseil, du caractère adéquat de la protection des données à caractère personnel en Suisse, JO L 215, 25 août 2000, 001 – 003.

[267] Rapport de la Commission au Parlement européen et au Conseil sur le premier réexamen du fonctionnement des décisions d'adéquation adoptées sur la base de l'art. 25, par. 6, de la directive 95/46/CE du 15 janvier 2024.

L'utilisation de systèmes d'IA dans l'UE est susceptible d'impliquer d'une part des transferts de données personnelles vers la Suisse, et d'autre part le traitement de données personnelles de ressortissants de l'UE transférées en Suisse. Le niveau de protection offert en doit suisse dans ce cadre pourrait donc être pertinent.

5.3.4 Relation avec la convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit

La convention sur l'IA est un traité international contraignant auquel peuvent adhérer les États membres du Conseil de l'Europe, les États non membres du Conseil de l'Europe, et l'UE. Le règlement sur l'IA est une législation interne à l'UE. La convention sur l'IA s'adresse aux États et contient des règles et principes généraux, lesquels devront en principe être mis en œuvre en droit interne en cas de ratification. Le règlement sur l'IA est en revanche directement applicable et contient des règles plus détaillées pour les systèmes d'IA dans l'UE. Il prévoit des obligations à la charge des opérateurs dans le domaine de l'IA. La nature et la portée de ces textes sont donc différentes.

L'objectif de la convention sur l'IA, à savoir la protection des droits de l'homme, la démocratie et l'État de droit, s'inscrit dans le mandat du Conseil de l'Europe de protection des droits fondamentaux. Le règlement sur l'IA est avant tout une réglementation sur les produits d'IA.[268] Il a but d'améliorer le fonctionnement du marché intérieur en établissant un cadre juridique uniforme applicable aux systèmes d'IA dans l'UE, tout en protégeant les droits fondamentaux (cf. art. 1, par. 1 et consid. 1 du règlement). Les objectifs sont donc partiellement différents. La convention sur l'IA n'a pas pour but de garantir des règles uniformes sur la circulation des produits d'IA. Dans ce sens, le règlement sur l'IA se distingue de la convention sur l'IA en particulier de par ses contenus techniques. Il semble dès lors que ces deux textes sont appelés à se compléter mutuellement.

Afin de comparer, lorsque cela est possible, la convention sur l'IA et le règlement sur l'IA, l'analyse contient une proposition de tableau mettant en évidence des corrélations possibles entre les dispositions des deux textes (cf. annexe 1).

5.3.5 Autres éléments choisis

Ce chapitre présente quelques autres caractéristiques du règlement sur l'IA du point de vue du droit suisse, qui paraissent importantes tant en lien avec la question de ses effets sur les opérateurs suisses, que s'agissant d'un éventuel rapprochement de la législation suisse avec ce règlement. L'analyse ne se veut pas exhaustive :

- Comme déjà mentionné, le règlement sur l'IA a pour objectif d'améliorer le fonctionnement du marché intérieur en lien avec la circulation des produits d'IA au sein de l'UE, tout en protégeant les droits fondamentaux (art. 1, par. 1, du règlement sur l'IA). La Suisse ne connaît pas ce même besoin d'unifier son marché intérieur.

[268] DAVID ROSENTHAL, Der EU AI Act (n. 221), 5 N 7 ; MARTINA ARIOLI, Risikomanagement (n. 222), 4 N 8 s.

- Le règlement sur l'IA renvoie abondamment à la législation européenne d'harmonisation, qui réglemente les produits circulant dans le marché intérieur. Par exemple, sont des systèmes d'IA à haut risque les systèmes d'IA qui sont des produits couverts par la législation d'harmonisation de l'UE selon l'annexe I, ou sont destinés à être utilisés comme composants de sécurité de ces produits (cf. art. 6, al. 1, du règlement). En outre, le règlement sur l'IA s'insère dans les mécanismes d'évaluation de la conformité existants et y ajoute des prescriptions (cf. notamment l'art. 43 sur le règlement).

  Du point de vue du droit suisse, ce lien étroit avec la législation européenne d'harmonisation accroît la complexité du fonctionnement du règlement.

- Le règlement sur l'IA contient en outre des mécanismes permettant au législateur européen d'intervenir rapidement afin de tenir compte des développements technologiques. Par exemple, l'art. 7, par. 1, du règlement prévoit que la Commission européenne peut adopter des actes délégués afin de modifier l'annexe III en y ajoutant des cas d'utilisation de systèmes d'IA à haut risque. Cette faculté a pour conséquence que le contenu du règlement sur l'IA peut potentiellement évoluer rapidement. Il s'agit d'un enjeu important pour les opérateurs suisses, qui devront observer de près les possibles évolutions pour s'y adapter dans la mesure du possible. En outre, dans l'hypothèse d'un éventuel rapprochement de la législation suisse à celle de l'UE, il conviendrait également de tenir compte de cet aspect afin de pouvoir si besoin réagir rapidement au développement du droit européen.

- Le règlement sur l'IA contient des règles transversales applicables à tous les systèmes d'IA, dans le secteur privé et dans le secteur public. Dans un État fédéral comme la Suisse, la répartition des compétences entre la Confédération et les cantons ne permettrait pas d'avoir une législation fédérale avec un champ d'application aussi large en droit public. Les compétences cantonales doivent être préservées.

  Par exemple, le règlement sur l'IA prévoit à son annexe III, point 3, let. c, que sont à haut risque les systèmes d'IA « destinés à être utilisés pour évaluer le niveau d'enseignement approprié qu'une personne recevra ou sera en mesure d'atteindre, dans le contexte ou au sein d'établissements d'enseignement et de formation professionnelle à tous les niveaux ». Or, dans la mesure où, en matière d'éducation, la Suisse connaît un système fédéraliste dans lequel les compétences cantonales revêtent une importance particulière, une telle règle pourrait tout au plus être prévue au niveau fédéral là où la Confédération est compétente.

  Ainsi, dans l'hypothèse d'un rapprochement de la législation suisse avec celle de l'UE en matière d'IA, il faudrait dûment tenir compte de la répartition des compétences Confédération-cantons.

- Toujours dans cette dernière hypothèse, le législateur suisse devrait également être conscient que la réglementation européenne délègue aux organismes de normalisation l'élaboration de normes d'harmonisation qui, dans le domaine de l'IA, sont d'une grande importance (cf. ch. 5.2.11).

- En tant que tel, le règlement sur l'IA contient peu de droits individuels (cf. les art. 85 et 86 du règlement).[269] Toutefois, le règlement sur l'IA n'est pas le seul instrument pertinent en matière d'IA dans l'UE. D'autres législations européennes s'appliquent aussi dans ce domaine.[270]

  En particulier, s'agissant de la protection des droits fondamentaux, le règlement sur l'IA constitue un élément additionnel qui contribue à la mise en œuvre d'autres législations de base protégeant les droits fondamentaux. Par exemple, au sein de l'UE et des États membres, les droits fondamentaux sont déjà protégés par la législation en vigueur (p. ex. en matière de protection des données et de non-discrimination). Le règlement sur l'IA veille à ce que les systèmes d'IA soient conformes dès la conception à ces autres législations. En cas de violation, les exigences du règlement sur l'IA permettront en outre aux autorités d'avoir accès aux informations nécessaires pour condamner d'éventuelles violations.[271] L'étendue de la protection effective des droits des individus dépend ainsi fortement des garanties existantes au niveau d'autres législations de fond.

  Un autre exemple est celui de la responsabilité. Le règlement sur l'IA ne règle par exemple pas les questions de responsabilité civile. Pour obtenir la réparation de leur dommage découlant de l'utilisation de systèmes d'IA, les personnes lésées devront se référer à d'autres législations matérielles (voir par exemple le projet de directive sur la responsabilité du fait des produits défectueux, cf. ch. 6.3.2). Le règlement sur l'IA crée les conditions pour faciliter la mise en œuvre de ces règles de fond.

  Un éventuel rapprochement de la Suisse avec le règlement sur l'IA devrait tenir compte du fait que ce dernier ne règle pas exhaustivement toutes les questions relatives à l'IA.

Ces quelques considérations permettent ainsi de constater que, du point de vue du droit suisse, le règlement sur l'IA présente des spécificités qui ne sont pas sans poser des enjeux juridiques, que ce soit en lien avec ses effets sur les opérateurs suisses, ou s'agissant d'un éventuel rapprochement de la législation suisse avec ce règlement.

---

[269] Cf. pour une critique de cette approche cf. ANGELA MÜLLER, Der Artificial intelligence Act der EU (n. 223), 19 s. ; MARTINA ARIOLI, Risikomanagement (n. 222), 16 N 54.

[270] Cf. dans ce sens DAVID ROSENTHAL, Der EU AI Act (n. 221), 5 N 9, qui relève que le règlement sur l'IA mentionne à plusieurs reprises qu'il s'applique en sus du droit en vigueur, cf. p. ex. art. 2, par. 7, et consid. 10.

[271] Cf. Commission européenne, Intelligence artificielle – Questions et réponses, *Comment les règles protègeront-elles les droits fondamentaux ?*, décembre 2023, disponible sous https://ec.europa.eu/commission/presscorner/detail/fr/QANDA_21_1683, consulté le 26 août 2024).

5.4 Conclusions intermédiaires

L'examen du règlement sur l'IA a permis de donner un aperçu de son contenu et d'identifier les principaux enjeux de la réglementation du point de vue du droit suisse. Les conclusions suivantes peuvent être formulées :

- Le cœur de la réglementation concerne les obligations des fournisseurs de systèmes d'IA à haut risque. En conséquence, les opérateurs suisses entrant dans cette catégorie sont les plus fortement impactés par le règlement sur l'IA.

- Le règlement sur l'IA vient compléter la réglementation harmonisée de l'UE, en y renvoyant et en y ajoutant des prescriptions. Cela contribue à la complexité du règlement.

- En cas de volonté politique de se rapprocher du règlement sur l'IA, il faudra tenir compte des spécificités respectives des ordres juridiques suisse et européen, telles que la répartition des compétences entre la Confédération et les cantons.

- Le règlement sur l'IA vise à harmoniser le fonctionnement du marché intérieur de l'UE en lien avec les systèmes d'IA, tout en protégeant les droits fondamentaux. Il contient cependant peu de droits individuels. La protection des intérêts individuels en matière d'IA est principalement garantie par le biais d'autres normes européennes, par exemple en matière de protection des données et de non-discrimination. Le règlement sur l'IA vient renforcer ces droits et leur mise en œuvre dans le cadre de l'IA.

- S'agissant du rapport avec la convention sur l'IA du Conseil de l'Europe, l'analyse a montré que ces deux textes ont une nature, une portée et des objectifs partiellement différents. On peut considérer qu'ils se complètent.

- L'adoption en Suisse d'une législation équivalente au règlement sur l'IA ainsi que l'incorporation dans l'ARM de ladite législation et du règlement sur l'IA adopté par l'UE pourrait permettre à la Suisse de bénéficier d'une reconnaissance mutuelle des futures évaluations de conformité des produits comprenant des systèmes d'IA.

  À défaut de reconnaissance mutuelle, les opérateurs suisses concernés qui exportent dans le marché de l'UE devront se référer aux exigences du règlement sur l'IA pour l'évaluation de la conformité des aspects d'IA de leurs produits.

Le règlement sur l'IA vient d'être adopté au sein de l'UE. De nombreuses questions devront encore être clarifiées s'agissant de son application, d'un point de vue tant juridique que pratique. L'analyse a mis en lumière les principaux enjeux appréciables à ce stade pour la Suisse. L'analyse devra se poursuivre en cas de volonté politique de se rapprocher de ce règlement.

6 Autres domaines du droit spécifiques

6.1 Introduction

La convention sur l'IA du Conseil de l'Europe et le règlement de l'UE sur l'IA sont des développements internationaux importants. Il convient cependant de constater que ces textes ne couvrent pas tous les domaines du droit pertinents en matière d'IA. Parfois, ils se limitent à renvoyer au droit existant dans d'autres matières. Tel est le cas par exemple du droit de la propriété intellectuelle et du droit de la responsabilité civile et pénale.

Afin de donner un aperçu un peu plus large de la situation du droit suisse par rapport aux défis de l'IA, nous abordons dans ce chapitre une série d'autres domaines du droit spécifiques et esquissons l'état actuel de la législation.

Dans ces domaines également, des approfondissements sont en tout état de cause nécessaires afin de réaliser un état des lieux complet de la situation.

6.2 Propriété intellectuelle[272]

L'IA transforme en permanence la façon de réaliser des inventions ou de créer des textes, de la musique ou des images. De ce fait, les principes établis dans le domaine de la propriété intellectuelle doivent éventuellement être repensés. Le droit d'auteur et le droit des brevets, notamment, sont confrontés à de nouveaux défis. Il convient donc d'étudier dans quelle mesure une adaptation au contexte technologique actuel s'impose dans ces deux domaines du droit.

6.2.1 Droit d'auteur et intelligence artificielle

6.2.1.1 Généralités

La LDA règle avant tout la protection des auteurs d'œuvres littéraires et artistiques (art. 1, al. 1, let. a, LDA).

L'art. 2, al. 1, LDA dispose que « par œuvre, quelles qu'en soient la valeur ou la destination, on entend toute création de l'esprit, littéraire ou artistique, qui a un caractère individuel ». La notion de création de l'esprit implique qu'une idée soit exprimée. L'auteur ne peut être qu'une personne physique (art. 6 LDA). Ne sauraient donc être des œuvres les objets qui n'ont pas été fabriqués par un humain.[273] Les termes littérature et art sont à entendre au sens large. L'art. 2, al. 2, LDA contient une liste non exhaustive de ce qui relève de l'art et de la littérature (p. ex. les œuvres recourant à la langue, les œuvres musicales, les beaux-arts, etc.). En principe, seul ce qui a un caractère individuel est protégé. Une certaine unicité est requise, c'est-

[272] Ce chapitre a été rédigé par l'IPI.

[273] FF **1989** III 465, 506.

à-dire des caractéristiques qui distinguent une création d'autres créations, existantes ou possibles.[274] Selon la jurisprudence du TF, l'originalité dans le sens du caractère personnel apporté par l'auteur n'est pas requise ; le critère décisif réside dans l'individualité, qui doit s'exprimer dans l'œuvre elle-même. Le caractère individuel exigé dépend de la marge de manœuvre de l'auteur. Lorsque celle-ci est restreinte, une activité indépendante réduite suffira à fonder la protection.[275]

L'auteur a le droit exclusif de décider si, quand et de quelle manière son œuvre sera utilisée. (art. 10, al. 1, LDA). On distingue généralement les droits d'utilisation des droits moraux de l'auteur. Les droits d'utilisation ont le plus souvent une composante patrimoniale, ainsi p. ex. l'autorisation de reproduire une œuvre est souvent accordée contre une rémunération. Les droits moraux ont une composante idéelle qui souligne le lien indissociable entre l'auteur et son œuvre. Le droit de faire reconnaître sa qualité d'auteur (art. 9 LDA) fait partie des droits moraux de l'auteur.

Une caractéristique majeure de la loi sur le droit d'auteur est sa neutralité technologique. La loi ne fait pas de distinction entre les techniques employées pour utiliser une œuvre. Une reproduction analogique est traitée de la même manière qu'une reproduction numérique. L'avantage de ce principe est que la loi n'a pas besoin d'être adaptée à chaque progrès technologique, mais cela ne signifie pas pour autant que les changements technologiques n'ont jamais d'influence sur la LDA. Les évolutions titanesques dans le contexte numérique (pensons à Internet) et les nouvelles formes d'utilisation auxquelles elles donnent naissance ont entraîné plusieurs révisions partielles de la LDA.[276]

6.2.1.2 Problématiques et nécessité de légiférer du fait de l'intelligence artificielle

L'importance croissante des applications d'IA générative met le droit d'auteur face à de nouveaux défis, voire remet en cause des principes bien établis. Eu égard aux systèmes d'IA générative, les points suivants doivent être examinés en détail : premièrement l'entraînement de ces systèmes, deuxièmement les résultats qu'ils produisent et troisièmement les instructions (*prompts*) qui leur sont transmises.

L'entraînement des systèmes d'IA nécessite un grand volume de données.[277] Les corpus d'entraînement sont par exemple composés d'images et de descriptions d'images disponibles

---

[274] FF **1989** III 465, 506.

[275] ATF **143** III 373, consid. 2.1 p. 377.

[276] Par exemple en 2008 et en 2020.

[277] Par exemple ADITYA RAMESH/PRAFULLA DHARIWAL/ALEX NICHOL/CASEY CHU/MARK CHEN, Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022, 23 (disponible sous : https://arxiv.org/abs/2204.061, consulté le 6 mai 2024).

sur Internet qui sont téléchargées de manière automatisée et systématique.[278] L'IA apprend ensuite à partir de ces corpus de données d'entraînement. Les contenus utilisés à cette fin peuvent être protégés par le droit d'auteur. La question se pose donc de savoir si l'utilisation de contenus protégés par le droit d'auteur pour l'entraînement des systèmes d'IA nécessite l'autorisation du titulaire des droits d'auteurs (ou du moins l'application de restrictions au droit d'auteur). La réponse à cette question dépend de plusieurs facteurs, à commencer par le droit applicable à l'entraînement de l'IA.[279] Concernant le droit suisse, qui nous intéresse ici, il s'agira de déterminer si l'assemblage de corpus d'entraînement et l'entraînement de systèmes d'IA constitue une « utilisation » d'œuvres au sens du droit d'auteur. Si l'utilisation relève du droit d'auteur, elle porte atteinte au droit exclusif du titulaire et nécessite une autorisation. L'art. 10, al. 2, LDA énonce diverses utilisations soumises à autorisation, parmi lesquelles la reproduction (copie). L'utilisation relève du droit d'auteur lorsque l'œuvre n'est pas simplement consommée, mais reproduite, transmise à un tiers et rendue perceptible.[280] La reproduction est souvent un acte préalable à la jouissance de l'œuvre[281] (qui en soi ne porte pas atteinte au droit exclusif du titulaire).

Tous les auteurs de doctrine ne s'accordent pas pour dire si seules les reproductions visant spécifiquement la jouissance de l'œuvre sont pertinentes pour le droit d'auteur, ou si c'est également le cas de celles qui ne la permettent que potentiellement. La doctrine majoritaire semble soutenir le deuxième avis.[282] Si l'on suit ce raisonnement, les reproductions réalisées dans le but de constituer un corpus d'entraînement et celles éventuellement utilisées pour l'entraînement proprement dit relèvent du droit d'auteur (puisqu'elles permettent potentiellement la jouissance de l'œuvre). Le cas échéant, la doctrine considère qu'il y a atteinte au droit de reproduction.[283] Toutefois, les avis divergent. Certains auteurs considèrent que la constitution du corpus est certes une reproduction relevant du droit d'auteur, mais que l'entraînement proprement dit constitue une perception de l'œuvre sans conséquence.[284] D'autres auteurs encore se fondent sur l'approche étasunienne et adoptent un point de vue différent. Ils considèrent que la reproduction est uniquement pertinente pour le droit d'auteur si elle aboutit à une exploitation économique de l'œuvre. À notre connaissance, les tribunaux suisses n'ont jamais eu à trancher cette question. Si la jurisprudence devait établir que la

---

[278] PAULINA JO PESCH/RAINER BÖHME, Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern, GRUR 2023/14, 997 ss, 998.

[279] Le droit international privé répond à cette question.

[280] DENIS BARRELET/WILLI EGLOFF, Das neue Urheberrecht. Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, Art. 10 N 8.

[281] MATHIS BERGER, Künstliche Intelligenz und Immaterialgüterrecht, Jusletter IT du 4 juillet 2024, 4.

[282] SANDRA MARMY-BRÄNDLI/ISABELLE OEHRI, Das Training künstlicher Intelligenz, sic! 2023, 655 ss, 658 (et les réf. citées).

[283] SANDRA MARMY-BRÄNDLI/ISABELLE OEHRI, Das Training (n. 282), 657 s. (et les réf. citées).

[284] Sur cette controverse, voir SANDRA MARMY-BRÄNDLI/ISABELLE OEHRI, Das Training (n. 282), 658 s.

création de corpus d'entraînement ou l'entraînement de systèmes d'IA constitue une reproduction pertinente pour le droit d'auteur, une autorisation du titulaire des droits d'auteur serait requise, à moins qu'une restriction ne s'applique.

Les restrictions suivantes pourraient être pertinentes ici : l'utilisation à des fins privées (art. 19, al. 1, let. c, LDA), les reproductions provisoires (art. 24*a* LDA) ou encore l'utilisation d'œuvres à des fins de recherche scientifique (art. 24*d* LDA). Pour chacune de ces restrictions, il existe cependant des critères qui, en fonction du cas d'espèce, pourraient les rendre inapplicables : le critère de la finalité d'information interne ou de documentation pour l'utilisation à des fins privées, le critère de la finalité transitoire ou accessoire pour les reproductions provisoires et le critère de la finalité scientifique pour la dernière restriction citée. Il conviendrait d'examiner dans le cas d'espèce si une de ces restrictions au droit d'auteur pourrait s'appliquer à un système d'IA particulier.[285]

Quelle que soit la réponse à ces questions, il peut s'avérer nécessaire de légiférer : si la création ou l'entraînement de systèmes d'IA ont des implications en termes de droit d'auteur et, en l'absence de restrictions applicables, nécessitent une autorisation de chaque titulaire des droits d'auteur, il faudrait s'assurer que le développement de systèmes d'IA reste possible de manière appropriée. Si, à l'inverse, ces activités devaient être considérées comme sans incidences en termes de droit d'auteur, il faudrait clarifier les conséquences pour les titulaires des droits d'auteur et les éventuelles mesures à prendre.

En ce qui concerne les résultats d'un système d'IA (*output*), on peut se demander d'une part s'ils peuvent porter atteinte au droit d'auteur et d'autre part s'ils peuvent eux-mêmes bénéficier de la protection offerte par le droit d'auteur. Concernant la première question, il convient d'abord d'examiner si une œuvre existante est reconnaissable dans le résultat, voire est reproduite à l'identique. Si un tel résultat est utilisé sans autorisation du titulaire des droits ou en l'absence de restriction applicable, il y a atteinte au droit d'auteur.

Quant à savoir si le résultat d'une IA peut bénéficier de la protection du droit d'auteur, il faut déterminer s'il constitue une « création de l'esprit » (en plus de satisfaire au critère de l'individualité).[286] Les exigences du Tribunal fédéral pour déterminer ce qui constitue une création de l'esprit sont loin d'être insurmontables. Une activité intellectuelle très modeste suffit pour accorder la protection du droit d'auteur.[287] Si les créations d'animaux ou les produits du hasard, par exemple, ne peuvent constituer des œuvre protégées, il y a création de l'esprit lorsqu'un humain se sert délibérément des lois du hasard (p. ex. l'*action painting*) ou de la technologie (p. ex. un ordinateur).[288] Dès que la volonté humaine décide du résultat, il y a

---

[285] Voir également MATHIS BERGER, Künstliche Intelligenz und Immaterialgüterrecht (n. 281), 5 ss.

[286] Voir le ch. 6.2.1.1.

[287] ATF **59** II 401, p. 405.

[288] DENIS BARRELET/WILLI EGLOFF, Das neue Urheberrecht. Kommentar (n. 280), Art. 2 N 8.

création de l'esprit.[289] On peut donc partir du principe que le résultat d'un système d'IA est une création de l'esprit lorsque l'utilisateur influence des éléments importants du résultat, et ce même lorsque le recours à l'IA comporte un certain degré de hasard. Un résultat produit en autonomie par un système d'IA ou seulement avec une faible influence de l'utilisateur ne serait toutefois pas une création de l'esprit et ne bénéficierait donc pas de la protection du droit d'auteur.[290]

En ce qui concerne la protection des instructions (*prompts*) soumises par les utilisateurs, nous renverrons aux considérations qui précèdent. Une série d'instructions rédigée par un humain remplit les critères de la création de l'esprit, littéraire ou artistique. Le critère déterminant pour le droit d'auteur sera donc celui du « caractère individuel » des instructions, qu'il conviendra d'examiner au cas par cas.

Les systèmes d'IA générative sont souvent entraînés avec des œuvres protégées par le droit d'auteur. En parallèle, de plus en plus de personnes utilisent ces systèmes d'IA pour créer elles-mêmes des images, des textes ou de la musique. Cela soulève des questions de droit d'auteur qui n'ont pas encore reçu de réponse définitive. Par exemple, les résultats des IA sont-ils protégés par le droit d'auteur et l'entraînement des IA est-il soumis au droit d'auteur ou non ? On peut supposer que la réponse à cette dernière question, en particulier, mettra en évidence la nécessité d'une réglementation. Si l'entraînement des IA est soumis au droit d'auteur, il convient d'examiner comment le développement (ultérieur) des IA peut être garanti. Dans le cas contraire, il faudra examiner dans quelle mesure il convient de tenir compte des intérêts des titulaires des droits d'auteur.

[289] FF **1989** III 465, 507.

[290] Il n'existe encore aucune jurisprudence contraignante sur ce point en Suisse. C'est toutefois la tendance qui se dessine à l'étranger pour le droit de la propriété immatérielle, p. ex. aux Etats-Unis en matière de droit d'auteur (U.S. Copyright Office, Robert J. Kasunic, Zarya of the Dawn [Registration # VAu001480196]) ou en Allemagne en matière de droit des brevets (Beschluss des Bundespatentgerichtes X ZB 5/22 du 11 juin 2024).

6.2.2 Intelligence artificielle et droit des brevets

6.2.2.1 Généralités

Le droit international des brevets est composé de nombreuses conventions.[291] La Convention sur le brevet européen[292] joue un rôle déterminant pour la Suisse.[293] Sur le plan national, la LBI et l'ordonnance relative aux brevets d'invention[294] régissent le droit des brevets.

Le droit des brevets protège des solutions techniques à des problèmes techniques, à condition que lesdites solutions soient nouvelles, inventives et reproductibles industriellement (art. 1 LBI). Les idées abstraites sans étapes techniques concrètes, les méthodes mathématiques, les algorithmes ou les méthodes d'apprentissage notamment, ne sont pas brevetables.

Le droit des brevets s'applique à tous les domaines techniques.[295] Il s'applique notamment aux inventions mises en œuvre par ordinateur[296] et aux inventions basées sur de l'IA. L'Organisation Mondiale de la Propriété Intellectuelle a enregistré une croissance de 718 % des demandes de brevets dans le domaine de l'IA entre 2016 et 2022.[297] À la suite de l'augmentation du nombre de demandes dans ce domaine, l'Office européen des brevets (OEB) a précisé ses directives d'examen pour les inventions basées sur l'IA dès 2018.[298]

L'IA est de plus en plus fréquemment utilisée comme outil de recherche et développement (R&D) pour développer des inventions dans tous les domaines techniques. Dans ce cadre, l'IA permet d'accélérer, d'étendre et de diversifier la R&D.

6.2.2.2 Défis

L'IA générative pourrait accroître drastiquement le volume de la littérature scientifique et technique. Celle-ci constitue l'état de l'art, à la lumière duquel la nouveauté et l'inventivité d'une demande de brevet sont analysées. Il pourrait ainsi devenir plus difficile d'obtenir un brevet.[299]

---

[291] En particulier, la Suisse a ratifié la Convention de Paris pour la protection de la propriété industrielle (RS **0.232.04**), le Traité de coopération en matière de brevets (RS **0.232.141.1**), le Traité sur le droit des brevets (RS **0.232.141.2**) et l'Accord sur les aspects des droits de propriété intellectuelle qui touchent au commerce (Annexe 1C de l'Accord instituant l'Organisation mondiale du commerce, **RS 0.632.20**).

[292] RS **0.232.142.2**.

[293] L'Office européen des brevets (OEB) est une organisation internationale indépendante de l'UE qui propose une procédure d'examen des brevets centralisée pour 39 pays. La majorité des demandes de brevets concernant la Suisse sont déposées directement auprès de l'OEB. Une fois délivrés par l'OEB, ils sont valables en Suisse.

[294] RS **232.14** et RS **232.141**.

[295] La LBI connaît quelques exceptions aux art. 1*a*. 1*b* et 2 LBI. Il s'agit par exemple du corps humain et ses éléments, les séquences géniques, les méthodes de traitement chirurgical, thérapeutique ou de diagnostic, les variétés végétales et les races animales. Par ailleurs, une invention dont la mise en œuvre porterait atteinte à la dignité humaine ou à l'intégrité des organismes vivants, ou serait d'une autre manière contraire à l'ordre public ou aux bonnes mœurs, ne peuvent être brevetées (par exemple instruments de torture).

[296] Les programmes d'ordinateur ne sont pas brevetables en tant que tels. Mais ils peuvent faire partie d'une invention si celle-ci a produit un effet au-delà du simple fonctionnement du programme.

[297] https://www.wipo.int > Understand and Learn > IP in > Frontier Technologies > AI Inventions factsheet, 1 (consulté le 15 mai 2024).

[298] Directives relatives à l'examen pratiqué à l'Office européen des brevets, Partie G, 3.3.1, « Intelligence artificielle et apprentissage automatique », version 03.2024, (disponible sous https://www.epo.org/fr/legal/guidelines-epc/2024/g_ii_3_3_1.html, consulté le 15 mai 2024) ; voir aussi https://www.epo.org/fr/news-events/in-focus/ict/artificial-intelligence (consulté le 15 mai 2024).

[299] Voir notamment All Prior Art, un générateur aléatoire de texte scientifique et technique, https://allpriorart.com (consulté le 15 mai 2024). L'Office américain des brevets a d'ailleurs organisé une consultation publique à ce sujet le 30 avril 2024, Request for Comments Regarding the Impact

L'utilisation de l'IA dans la R&D devrait faire évoluer l'évaluation du critère d'inventivité de l'invention. Toute solution technique ou toute amélioration d'une solution technique n'est pas brevetable. Il faut que la solution présentée dans la demande de brevet soit inventive, c'est-à-dire, qu'elle ne soit pas évidente pour la personne du métier. L'IA possède des capacités d'analyse et de génération de solutions complexes qui dépassent les limites de la compréhension humaine traditionnelle. Cela pourrait étendre le champ des idées novatrices en identifiant des combinaisons non évidentes pour les spécialistes humains, forçant une évolution du critère d'inventivité.

La divulgation des données d'entraînement d'une invention basée sur l'IA est discutée. Un brevet est délivré lorsque l'invention est décrite dans le fascicule du brevet de telle manière qu'une personne du métier puisse la reproduire. Pour les inventions basées sur l'IA, les données d'entraînement jouent un rôle essentiel pour la reproduction de l'invention. La manière et la mesure de la divulgation de ces données d'entraînement sont donc débattues. Il convient ici de tenir compte d'autres obligations ou intérêts, comme la protection des données et la protection des secrets d'affaires ou d'obligations de confidentialité. Lors de la dernière mise à jour de ses Directives d'examen, l'Office européen des brevets a précisé ce point : *« Si l'effet technique dépend de caractéristiques particulières de l'ensemble de données d'entraînement utilisé, ces caractéristiques qui sont nécessaires à la reproduction de l'effet technique doivent être exposées, à moins que l'homme du métier puisse les déterminer sans effort excessif à l'aide des connaissances générales. Cependant, en général, il n'est pas nécessaire d'exposer l'ensemble de données d'entraînement spécifique lui-même. »*[300] Le dernier défi concerne l'exigence d'indiquer une personne physique comme inventeur dans toute demande de brevet. La question est cristallisée par *The Artificial Inventor Project*[301] qui cherche à obtenir l'enregistrement de deux brevets au nom de *« DABUS, un système d'intelligence artificielle »*, sans indiquer de personne physique comme inventeur. Cette démarche questionne cette exigence du droit des brevets et renvoie à des questions fondamentales sur le but du droit des brevets.

6.2.2.3 Besoin de légiférer

Il ne semble pas y avoir de besoin de légiférer. L'augmentation spectaculaire des dépôts pour les inventions basées sur l'IA entre 2016 et 2022 en fait la démonstration. Les conditions de protection de ces inventions sont claires et stables et des brevets sont délivrés pour ce type

---

of the Proliferation of Artificial Intelligence on Prior Art, the Knowledge of a Person Having Ordinary Skill in the Art, and Determinations of Patentability Made in View of the Foregoing, disponible sous https://www.federalregister.gov/documents/2024/04/30/2024-08969/request-for-comments-regarding-the-impact-of-the-proliferation-of-artificial-intelligence-on-prior (consulté le 15 mai 2024).

[300] Cf. n. 298.

[301] The Artificial Inventor Project, https://artificialinventor.com/ (consulté le 15 mai 2024). Il s'agit d'un test du système de brevets grandeur nature. Le déposant affirme avoir créé un système d'IA capable de générer de manière autonome des inventions. Des demandes de brevets ont été déposées dans le monde entier. L'affaire est montée jusqu'à la Cour suprême des Etats-Unis et jusqu'à celle du Royaume-Uni. Elles se sont toutes deux prononcées contre l'admission d'un inventeur artificiel. L'OEB a rejeté la demande d'enregistrement. En Suisse, l'IPI a rejeté la demande d'enregistrement. Un recours est pendant au Tribunal administratif fédéral.

d'inventions. Le système dispose de plus d'une certaine flexibilité. Les notions d'état de l'art, de nouveauté, d'inventivité et l'exigence de divulgation de l'invention peuvent – dans une certaine mesure – être interprétées par les offices de propriété intellectuelle et les autorités judiciaires, en fonction des évolutions technologiques.

Concernant l'exigence de la mention d'une personne physique comme inventeur, le débat suscité par *The Artificial Inventor Project* semble prématuré. L'être humain conserve un rôle déterminant dans la R&D et il demeure possible d'indiquer au moins une personne physique comme inventeur.

- L'utilisation de l'IA dans la R&D pourrait faire évoluer les notions de nouveauté et d'inventivité.
- Il convient d'établir une pratique claire et stable en ce qui concerne la divulgation des données d'entraînement pour les inventions basées sur l'IA.
- Les développements jurisprudentiels mondiaux concernant l'exigence de mentionner une personne physique comme inventeur doivent être suivis avec attention.
- Il ne semble pas y avoir besoin de légiférer. Le système du droit des brevets est suffisamment flexible pour s'adapter aux évolutions technologiques sans être modifié. L'augmentation spectaculaire des dépôts de brevets pour des inventions basées sur l'IA depuis 2016 démontre que le système fonctionne à satisfaction.

6.3 Responsabilité extracontractuelle

6.3.1 Droit de la responsabilité civile extracontractuelle

6.3.1.1 Proposition de directive européenne sur la responsabilité en matière d'intelligence artificielle

La proposition de directive de la Commission européenne du 28 septembre 2022 relative à l'adaptation des règles de la responsabilité civile extracontractuelle à l'intelligence artificielle (proposition de directive sur la responsabilité en matière d'IA)[302] concerne la *responsabilité extracontractuelle pour faute* et vise à relever les défis que pose l'IA pour l'application des règles de responsabilité existantes en recourant à la divulgation et aux présomptions réfragables. L'autonomie, la complexité et l'opacité de certains systèmes d'IA rendent difficile, sinon impossible, de prouver le lien de causalité entre l'acte ou l'omission des responsables. La directive vise à corriger cette situation en allégeant la charge de la preuve. La proposition

[302] Proposition de directive du Parlement et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle, COM/2022/496 final, disponible sous https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52022PC0496 (consulté le 28 août 2024).

s'appuie sur les définitions et les catégories du règlement sur l'IA[303] et prévoit des conséquences juridiques particulières pour les *systèmes d'IA à haut risque*, ce qui devrait également contribuer au respect des exigences de la loi sur l'IA.[304] Selon l'exposé des motifs de la proposition de directive, la présomption réfragable est l'instrument le moins restrictif.[305] L'approche prévue dans la résolution[306] du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle, à savoir une responsabilité pour risque[307] pour les opérateurs de systèmes à haut risque, n'a pas été poursuivie. L'application de la directive doit être examinée cinq ans après l'expiration du délai de transposition. La nécessité d'une responsabilité pour risque sans faute assortie d'une assurance obligatoire devrait donc être examinée à cette occasion.[308] Plus précisément, le contenu de la proposition est le suivant :

- **Champ d'application :** la proposition de directive contient des règles sur la responsabilité pour faute extracontractuelle en cas de dommages causés par des systèmes d'IA. La directive ne s'applique pas à la responsabilité contractuelle. Elle ne s'applique pas non plus dans le domaine de la responsabilité pénale (art. 1(1)+(2) de la proposition de directive).

- **Production de moyens de preuve pertinents concernant des systèmes d'IA à haut risque :** la proposition de directive vise à fournir aux personnes qui introduisent une action en réparation un moyen d'*obtenir des informations sur les systèmes d'IA à haut risque* qui doivent être enregistrées ou documentées conformément au règlement sur l'IA (art. 3 de la proposition de directive).[309] Sur demande, les tribunaux seraient habilités à ordonner aux fournisseurs et aux utilisateurs de systèmes d'IA à haut risque soupçonnés d'avoir causé un dommage de produire des moyens de preuve. Le demandeur potentiel doit étayer la plausibilité de sa demande en réparation et démontrer que ses propres efforts n'ont pas abouti. Par souci de proportionnalité, les intérêts de toutes les parties concernées et de tiers, et notamment les secrets d'affaires, sont pris en compte. Si un défendeur ne se conforme pas à l'injonction de divulguer ou de

---

303 Voir le ch. 5.

304 Proposition de directive sur la responsabilité en matière d'IA (n. 302), exposé des motifs, 3.

305 Proposition de directive sur la responsabilité en matière d'IA (n. 302), exposé des motifs, 7.

306 JO C 404/107, disponible sous https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52020IP0276 (consulté le 28 août 2024).

307 Voir GERHARD WAGNER, Haftung für Künstliche Intelligenz - Eine Gesetzesinitiative des Europäischen Parlaments, ZEuP 2021, 545 ss, 556 ss.

308 Proposition de directive sur la responsabilité en matière d'IA (n. 302), exposé des motifs, 7.

309 Proposition de directive sur la responsabilité en matière d'IA (n. 302), consid. 17.

conserver les éléments de preuve, il y a *présomption de non-respect d'un devoir de vigilance* (article 3[5] de la proposition de directive).

- **Présomption de causalité :** à certaines conditions, une *présomption de causalité* est prévue tant pour les systèmes d'IA à haut risque que pour les autres systèmes d'IA. Cette présomption vise à alléger la charge de la preuve qu'une donnée spécifique entrée par la personne potentiellement responsable a conduit le système d'IA à produire un résultat spécifique à l'origine du dommage en cause. Il y a présomption du lien de causalité entre la faute du défendeur et le résultat produit par le système d'IA ou le fait qu'aucun résultat n'a été produit. La proposition ne renverse pas le fardeau de la preuve, elle crée une présomption réfragable.[310] Elle ne dispense pas non plus de prouver la faute, c'est-à-dire, en général, un manquement à un devoir de vigilance. Il est fait référence à des devoirs de vigilance prévus par le droit de l'UE ou le droit national dont le but direct est d'empêcher le dommage de se produire (art. 4[1][a] de la proposition de directive). Le texte établit donc un lien direct avec le règlement sur l'IA. La présomption de causalité s'applique aux conditions suivantes :

    - Pour les fournisseurs de systèmes d'IA à haut risque, s'il peut être prouvé que les exigences du règlement sur l'IA en matière d'entraînement, de transparence, de contrôle, d'exactitude, de robustesse et de cybersécurité n'ont pas été respectées ou que les mesures correctives nécessaires n'ont pas été prises immédiatement (art. 4[2] de la proposition de directive).

    - Pour les utilisateurs de systèmes d'IA à haut risque, s'il peut être prouvé que les exigences du règlement sur l'IA en ce qui concerne l'utilisation, la surveillance ou la destination des données d'entrée n'ont pas été respectées (art. 4[3] de la proposition de directive).

    - Pour les systèmes d'IA qui ne sont pas à haut risque, la présomption ne s'applique que si la juridiction nationale estime qu'il est excessivement difficile pour le demandeur de prouver le lien de causalité entre le dommage et la faute.

La proposition de directive de la Commission européenne est actuellement (au 31 août 2024) soumise à la commission des affaires juridiques du Parlement européen.[311] Selon les médias, celle-ci a commandé une étude complémentaire sur la nécessité de cette directive supplémentaire – en plus du règlement sur l'IA et de la directive actualisée sur la responsabilité du fait des produits défectueux[312], qui ont toutes deux été adoptées par le Parlement en mars

[310] Proposition de directive sur la responsabilité en matière l'IA (n. 302), exposé des motifs, p. 7 (proportionnalité).

[311] 2022/0303(COD) : Adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle, https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2022%2F0303(COD)&l=fr (consulté le 26 août 2024).

[312] Voir à ce sujet le ch. 6.3.2.

2024.[313] L'avenir de la proposition de directive sur la responsabilité en matière d'IA semble donc incertain.

6.3.1.2 Droit suisse

De manière générale, le droit civil suisse, et notamment le droit de la responsabilité civile, avec ses clauses générales ouvertes, est en mesure de faire face aux évolutions techniques et fournit aux tribunaux des outils pour parvenir à des solutions équitables dans chaque cas d'espèce.[314] C'est pourquoi, par le passé, le Conseil fédéral s'est prononcé à plusieurs reprises contre des modifications législatives dues à des innovations techniques, car celles-ci risquaient d'être dépassées en peu de temps.[315] Une législation trop axée sur les détails techniques risque de créer des lacunes. Néanmoins, du fait de la numérisation, il convient de vérifier régulièrement la compatibilité des bases légales en vigueur avec les dernières évolutions technologiques et, le cas échéant, de combler les lacunes existantes afin de maintenir la sécurité du droit.

La Suisse ne dispose pas de règles de responsabilité civile visant spécifiquement l'IA, à l'instar des règles européennes mentionnées. En vertu de la clause générale de la responsabilité civile définie à l'art. 41, al. 1, CO, celui qui cause, d'une manière illicite, un dommage à autrui, soit intentionnellement, soit par négligence ou imprudence, est tenu de le réparer. La responsabilité présuppose donc le dommage, l'illicéité, le lien de causalité et la faute. En vertu de l'art. 8 CC[316], chaque partie doit, si la loi ne prescrit le contraire, prouver les faits qu'elle allègue pour en déduire son droit. Conformément à l'art. 55, al. 1, CPC[317], elle a donc la charge d'alléguer les faits sur lesquels elle fonde ses prétentions et de produire les preuves qui s'y rapportent. Elle doit soumettre au tribunal une réquisition dans laquelle les preuves sont identifiées ou clairement identifiables (art. 152, al. 1, CPC).[318] Les réquisitions purement exploratoires (« fishing expeditions ») ne sont pas admissibles. La partie adverse est tenue de collaborer et doit notamment produire les titres précisément désignés (art. 160, al. 1, let. b, CPC). Une partie n'a pas le droit de refuser de collaborer pour se soustraire à sa propre responsabilité civile.[319] Si une partie refuse de collaborer sans motif valable, le tribunal en tient compte

---

313 https://www.euractiv.com/section/digital/news/picking-up-the-ai-liability-directive-after-the-tech-policy-spree/ (consulté le 26 août 2024)

314 Voir le Rapport défis de l'intelligence artificielle (n. 1), 36 s.

315 Sur les principes de la politique fédérale en matière de nouvelles technologies, voir le Rapport défis de l'intelligence artificielle (n. 1), 34 ss ; Rapport du Conseil fédéral « Bases juridiques pour la *distributed ledger technology* et la *blockchain* en Suisse » du 14 décembre 2018, 13 ss, disponible sous https://www.newsd.admin.ch/newsd/message/attachments/55152.pdf (consulté le 31 juillet 2024) ; cf. également le rapport du Conseil fédéral « La responsabilité civile des fournisseurs de services Interne du 11 décembre 2015 », 99 ss, disponible sous : www.ofj.admin.ch > Publications & services > Rapports, avis de droit et décisions > Rapports et avis de droit > La responsabilité civile des fournisseurs de services Internet (consulté le 31 juillet 2024).

316 RS **210**.

317 RS **272**.

318 BSK ZPO-GUYAN, art. 152 N 3 s.

319 Message CPC du 28 juin 2006, FF **2006** 6841, 6926.

lors de l'appréciation des preuves (art. 164 CPC). Il n'en résulte cependant pas automatiquement une présomption en défaveur de la partie qui refuse de collaborer.[320] Dans certaines circonstances, le tribunal peut administrer les preuves à titre préventif afin d'évaluer les chances de succès de la procédure (art. 158, al. 1, let. b, CPC).[321]

Un allégement de la charge de la preuve est possible dans certains cas. Selon la jurisprudence et la doctrine, une vraisemblance prépondérante[322] suffit lorsqu' « une preuve stricte n'est pas seulement impossible à apporter dans un cas particulier, mais est exclue ou n'est pas raisonnablement exigible en raison de la nature même de l'affaire et que l'on se trouve ainsi dans un ‹ état de nécessité en matière de preuve › (‹ Beweisnot ›) ».[323] C'est le cas de la preuve du lien de causalité en droit de la responsabilité civile.[324]

L'on peut également noter que ces règles valent pour les procédures soumises à la maxime des débats au sens strict. Dans le cadre de la procédure simplifiée, qui s'applique aux litiges d'une valeur litigieuse qui ne dépasse pas 30 000 fr. (art. 243, al. 1, CPC), le rôle du tribunal est plus actif dans la mesure où son devoir d'interroger les parties est accru (art. 247, al. 1, CPC).

6.3.1.3 Appréciation

Étant donné que les exigences de la proposition de directive sur la responsabilité en matière d'IA se fondent sur les dispositions du règlement sur l'IA, il n'est pas envisageable d'introduire la directive de manière isolée. Toutefois, si le règlement sur l'IA était transposé en droit suisse, les mesures d'accompagnement de la proposition de directive ne seraient pas totalement étrangères à l'ordre juridique suisse. Les instruments de procédure civile en vigueur permettent déjà, dans certains cas, la production d'une documentation identifiable avec précision et détenue par l'autre partie. Les conséquences juridiques d'un refus de collaborer ou de divulguer des éléments de preuve sont également comparables, mais la prise en compte de ce refus dans l'appréciation des preuves va moins loin en droit suisse que la présomption de non-respect d'un devoir de vigilance prévue par la proposition de directive. La présomption de causalité prévue par cette dernière ne peut pas non plus être déduite du droit suisse en vigueur, même si la charge de la preuve est allégée dans ce cas. La reprise de la directive

---

320 ATF **140** III 264, consid. 2.3.

321 BSK ZPO-GUYAN, Art. 158 N 5.

322 « Selon le degré de la preuve de la vraisemblance prépondérante […], une preuve est considérée comme rapportée lorsqu'en examinant objectivement les choses l'exactitude d'une allégation de fait repose sur des motifs si importants que d'autres possibilités imaginables n'entrent raisonnablement pas en ligne de compte » ; ATF **132** III 715, traduit dans JdT 2009 I 183, consid. 3.1. et les réf. citées.

323 ATF **132** III 715, consid. 3.1. et les réf. citées, traduit dans JdT 2009 I 183, consid. 3.1. et les réf. citées.

324 ATF **132** III 715, consid. 3.2. et les réf. citées : « Un fait est déjà rendu vraisemblable si certains éléments parlent en faveur de son existence, même si le tribunal tient encore pour possible qu'il ne se soit pas produit » ; ATF **132** III 715, traduit dans JdT 2009 I 183, consid. 3.1. et les réf. citées.

sur la responsabilité en matière d'IA pourrait donc éventuellement faciliter l'invocation de prétentions civiles et contribuer à une meilleure mise en œuvre dans le secteur privé des dispositions de la convention sur l'IA du Conseil de l'Europe (notamment ses articles 8, 9 et 14 ; cf. ch. 4.3). Il convient toutefois d'attendre les analyses complémentaires et les résultats des discussions au sein de l'UE.

Il est à noter que le champ d'application de ces règles générales de responsabilité civile extracontractuelle serait très limité.[325] En effet, pour diverses utilisations possibles des systèmes d'IA, le droit en vigueur prévoit déjà une responsabilité pour risque qui offre une meilleure protection aux personnes potentiellement lésées, à savoir :

- la responsabilité civile du détenteur de véhicule automobile selon l'art. 58 LCR ;
- la responsabilité civile de l'exploitant d'un aéronef selon l'art. 64 de la loi fédérale du 21 décembre 1948 sur l'aviation (LA)[326], et
- la responsabilité du producteur selon la loi fédérale du 18 juin 1993 sur la responsabilité du fait des produits (LRFP)[327] (voir à ce sujet le ch. 6.3.2).

Dans d'importants domaines d'application pratique de l'IA (voitures autonomes, drones), il existe donc déjà des règles de responsabilité qui garantissent la protection des personnes lésées. La responsabilité civile du détenteur de véhicule – associée à une obligation d'assurance (art. 63 LCR et 70 LA) – permet d'éviter les lacunes en matière de responsabilité.

6.3.2 Droit de la responsabilité du fait des produits

La responsabilité du fait des produits prévoit une responsabilité sans faute des producteurs pour les dommages matériels et corporels causés par des produits défectueux. Elle ne couvre donc que les dommages dus à des produits défectueux et non les dommages causés au produit lui-même.

---

[325] ISABELLE WILDHABER, KI und Haftung : Lösungsansätze für die Schweiz, Jusletter IT du 4 juillet 2024, N 52 s. ; voir aussi : ISABELLE WILDHABER, Eine Einführung in die ausservertragliche Haftung für Künstliche Intelligenz (KI), in : HAVE (éd.), Haftpflichtprozess 2021, Zurich 2021, 1 ss, 45 ; BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie aus schweizerischer Sicht, RSJ 2023, 627 s., 637.

[326] RS **748.0**.

[327] RS **221.112.944**.

6.3.2.1 Directive européenne actualisée sur la responsabilité du fait des produits défectueux

La directive sur la responsabilité du fait des produits défectueux[328], adoptée par le Parlement européen le 12 mars 2024, remplacera l'actuelle directive 85/374/CEE du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux[329] (directive sur la responsabilité du fait des produits). Elle fait suite à la proposition de directive du Parlement européen et du Conseil du 28 septembre 2022 relative à la responsabilité du fait des produits défectueux[330]. La directive actualisée prévoit d'importantes innovations afin de soumettre *les logiciels, y compris les systèmes d'IA*, à la responsabilité sans faute du producteur. Bien qu'elle ait été proposée dans le cadre du train de mesures sur l'IA, elle ne vise pas en premier lieu les applications d'IA, mais plutôt une modernisation générale de la responsabilité du producteur pour les produits défectueux dans le contexte de la numérisation croissante. Elle prend en compte le fait qu'aujourd'hui, les fabricants ne perdent pas nécessairement le contrôle de leurs produits ou toute influence sur ceux-ci dès leur mise sur le marché.[331]

Tant le *cercle des responsables* (fabricants de logiciels, plateformes en ligne et exploitants de systèmes d'IA) que la *notion de faute* (mises à jour défectueuses, cybersécurité insuffisante) sont élargis. La responsabilité du fabricant est également étendue aux dommages résultant de la perte ou de la falsification de données, en plus des dommages matériels et corporels. Comme dans la proposition de directive sur la responsabilité en matière d'IA, la difficulté d'obtention des éléments de preuve est contrecarrée par des *obligations de produire des moyens de preuve* et *des présomptions de causalité*. En outre, le texte recoupe deux autres directives européennes de droit civil en ce qui concerne les obligations de mise à jour :

- *directive (UE) 2019/771*[332] du 20 mai 2019 relative à certains aspects concernant les contrats de vente de biens (*directive sur la vente de biens [UE] 2019/771*) ;

- *directive (UE) 2019/770*[333] du Parlement européen et du Conseil du 20 mai 2019 relative à certains aspects concernant les contrats de fourniture de contenus numériques et de services numériques *(directive [UE] 2019/770 sur le contenu et les services numériques).*

---

[328] Résolution législative du Parlement européen du 12 mars 2024 sur la proposition de directive du Parlement européen et du Conseil sur la responsabilité du fait des produits défectueux, disponible sous : https://www.europarl.europa.eu/doceo/document/TA-9-2024-0132_FR.html (consulté le 26 août 2024).

[329] Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux, JO L 210 du 7 août 1985, 29 à 33, disponible sous https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=celex%3A31985L0374 (consulté le 26 août 2024).

[330] Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, COM/2022/495 final, 28 septembre 2022, disponible sous https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52022PC0495 (consulté le 26 août 2024).

[331] BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie (n. 325), 630.

[332] Directive (UE) 2019/771 du 20 mai 2019 relative à certains aspects concernant les contrats de vente de biens, modifiant le règlement (UE) 2017/2394 et la directive 2009/22/CE et abrogeant la directive 1999/44/CE, JO L 136 du 22 mai 2019, 28 ss.

[333] JO L 136 du 22 mai 2019, 1 ss.

Au 31 août 2024, la directive adoptée par le Parlement le 12 mars 2024 devait encore être confirmée par le Conseil européen et publiée au Journal officiel. Une fois la directive entrée en vigueur, les États membres disposeront de deux ans pour la mettre en œuvre.[334]

6.3.2.2 Droit suisse

La Suisse a introduit la LRFP dans le cadre du train de mesures Swisslex, qui correspond en grande partie à l'ancienne directive sur la responsabilité du fait des produits (directive 85/374/CEE).[335] En vertu de la LRFP, les fabricants peuvent être tenus responsables de leurs produits défectueux. Depuis des années, la question de savoir si les logiciels seuls, c'est-à-dire les logiciels qui ne sont pas liés à un produit physique, peuvent également être considérés comme des produits au sens de la LRFP, est controversée en doctrine.[336] Le Tribunal fédéral ne s'est pas encore prononcé sur cette question. Une grande partie de la doctrine récente est favorable à une interprétation large de l'art. 3, al. 1, LRFP, qui entrerait en contradiction avec le texte, mais souhaite que le législateur clarifie la situation.[337]

6.3.2.1 Appréciation

L'assujettissement des logiciels à la responsabilité du fabricant définie par la LRFP, tel qu'il est prévu par la proposition de directive européenne, est réclamé en Suisse depuis longtemps par la doctrine.[338] La LRFP en vigueur ne contient pas non plus les autres nouveautés de la proposition de directive européenne actualisée, qui tiennent compte de la numérisation croissante des produits.[339]

Dans son rapport « Modernisation du droit de la garantie de la chose vendue » du 16 juin 2023, le Conseil fédéral a constaté, en ce qui concerne la *garantie du vendeur pour les défauts de la chose*, que le droit en vigueur, axé sur un échange singulier entre un bien et une somme d'argent, n'est pas adapté aux produits numériques ou comportant des éléments numériques.[340] Il s'est donc prononcé en faveur de l'introduction d'une obligation de mise à jour pour les produits numériques et les produits comportant des éléments numériques, sur le modèle de la directive sur les contenus et services numériques (UE) 2019/770 et de la directive sur la vente de biens (UE) 2019/771. Lors de la session d'hiver 2023, le Conseil fédéral a été

---

[334] Art. 22 de la directive relative à la responsabilité du fait des produits défectueux, disponible sous https://www.europarl.europa.eu/doceo/document/TA-9-2024-0132_FR.html (consulté le 26 août 2024).

[335] Voir le message sur le programme consécutif au rejet de l'Accord EEE du 24 février 1993, FF **1993** I 757, 833.

[336] Voir BSK CO-FELLMANN, Art. 3 PrHG N 10 et les réf. citées.

[337] Voir FELLMANN, Haftpflichtrecht im Zeichen der Digitalisierung, HAVE 2021, 105 ss ; ISABELLE WILDHABER, Eine Einführung (n. 325), 26 ; BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie aus schweizerischer Sicht (n. 325), 638, et les réf. citées ; pour un assujettissement des développeurs d'IA à la responsabilité du producteur *de lege lata*, voir notamment : ARIANE MORIN, L'opposabilité de la LRFP au fournisseur de l'intelligence artificielle, in : Damiano Canapa/Alexandre Richa (éds.), Aspects juridiques de l'intelligence artificielle, Berne 2024, 117 ss, 120 s.

[338] Voir les références à la nbp 337 ainsi qu'ISABELLE WILDHABER, KI und Haftung (n. 325), n. 26 ss.

[339] Sur le besoin d'adaptation de la LRFP en détail : BERNHARD A. KOCH/PASCAL PICHONNAZ, Der Entwurf einer neuen EU-Produkthaftungsrichtlinie (n. 325), 637 ss.

[340] Rapport du Conseil fédéral « Modernisation du droit de la garantie de la chose vendue » du 16 juin 2023, ch. 4.1 ss, disponible sous https://www.newsd.admin.ch/newsd/message/attachments/79588.pdf (consulté le 31 juillet 2024).

chargé d'élaborer un projet en ce sens par les motions identiques 23.4316 et 23.4345 des Commissions des affaires juridiques des deux Chambres.[341]

Le Conseil fédéral et le Parlement ne se sont pas encore prononcés sur une éventuelle révision de la LRFP.

6.3.3 Protection de la personnalité

Enfin, dans le contexte des applications d'IA, les moyens de la protection générale de la personnalité en droit civil doivent également être abordés. Ceux-ci sont particulièrement pertinents dans le cas du phénomène important des *deepfakes*[342]. Une personne qui subit une atteinte illicite à sa personnalité peut agir en justice pour sa protection contre toute personne qui y participe (art. 28, al. 1, CC). La personne concernée peut notamment requérir le juge de faire cesser une atteinte à la personnalité existante ou d'interdire une atteinte imminente (art. 28*a*, al. 1, ch. 1 et 2, CC). Cette norme est technologiquement neutre.[343] La personnalité (y compris le droit à l'image et à la voix) est un bien juridique protégé de manière absolue. L'art. 28 CC contient une interdiction générale tacite de porter atteinte sans autorisation à la personnalité d'autrui.[344] Les atteintes à la personnalité par le biais d'une application d'IA sont donc aisément couvertes par le droit en vigueur. Il est à noter que les manipulations d'images existaient déjà avant la diffusion des applications d'IA. Ainsi, le Tribunal fédéral a déjà eu à connaître de photomontages et de manipulations d'images et il n'a pas remis en question le fait que des images manipulées peuvent porter atteinte à la personnalité des personnes concernées.[345] Cela a également été confirmé par une récente décision de première instance concernant une vidéo créée par IA : un homme politique a utilisé à des fins électorales une vidéo truquée d'une opposante, créée à l'aide d'une application d'IA. Le tribunal civil de Bâle a considéré qu'il s'agissait d'une atteinte à la personnalité.[346] L'art. 28*a*, al. 1, ch. 1 et 2, CC, permet également d'ordonner la suppression et le blocage de contenus illicites sur Internet. Des actions en dommages-intérêts, en réparation du tort moral ainsi qu'en remise du gain sont également envisageables (art. 28*a*, al. 3, CC).

6.3.4 Conclusion

De manière générale, le droit civil suisse, et notamment le droit de la responsabilité civile, avec ses clauses générales ouvertes, est en mesure de faire face aux développements techniques et fournit aux tribunaux des outils pour parvenir à des solutions équitables dans

[341] 23.4316 Mo. CAJ-E « Modernisation du droit de la garantie » et 23.4345 Mo. CAJ-N « Modernisation du droit de la garantie ».

[342] 23.3563 Mo. Mahaim « Réglementer les "deep fakes" ».

[343] Sur la neutralité technologique de la législation suisse et la prise en compte des *deepfakes* en droit pénal, voir le ch. 6.6.

[344] HEINZ HAUSHEER/REGINA E. AEBI-MÜLLER, Das Personenrecht des Schweizerischen Zivilgesetzbuches, 5e édition, Berne 2020, N 490 ss, 544.

[345] Voir l'arrêt du TF 5A_553/2012 du 14 avril 2014 ; arrêt du TF 5A_376/2013 du 29 octobre 2013.

[346] Medialex newsletter 01/24, disponible sous https://medialex.ch/2024/02/07/newsletter-01-24/ (consulté le 31 juillet 2024).

chaque cas d'espèce. La responsabilité pour risque et les obligations d'assurance existantes dans le domaine de la circulation routière et de l'aviation permettent en outre d'éviter les lacunes en matière de responsabilité dans des domaines essentiels. De même, les règles existantes permettent a priori de faire valoir les droits en justice. L'éventuelle adoption de la directive européenne sur la responsabilité en matière d'IA pourrait toutefois faciliter l'invocation des prétentions de droit civil et contribuer à une meilleure mise en œuvre dans le secteur privé des dispositions de la convention sur l'IA du Conseil de l'Europe (notamment ses art. 8, 9 et 14, voir le ch. 4.3). Introduire la directive seule – sans reprendre en même temps le règlement sur l'IA – n'est toutefois pas envisageable. Il convient également d'attendre les analyses complémentaires et les résultats des discussions au sein de l'UE sur la directive relative à la responsabilité en matière d'IA, dont l'avenir est incertain.

En raison de l'évolution technologique des produits – en partie du fait de l'IA – un besoin général de modernisation se dessine en ce qui concerne la loi sur la responsabilité du fait des produits.

6.4 Droit général des contrats

6.4.1 Loi type de la CNUDCI sur les contrats automatisés

La Commission des Nations Unies pour le droit commercial international (CNUDCI) a adopté en juillet 2024 la loi type sur les contrats automatisés.[347] Celle-ci avait été élaborée auparavant au sein du « Groupe de travail IV : Commerce électronique » de la CNUDCI – avec la participation active de la Suisse.[348] La loi type dispose que la validité des contrats ne peut être niée du seul fait qu'un système automatisé est intervenu dans leur conclusion. D'autres règles concernent le traitement des messages générés par des systèmes automatisés. L'adoption de la loi type par la Suisse, si tant est que le droit en vigueur ne remplisse pas déjà les mêmes exigences, sera examinée en temps voulu.

6.4.2 Droit suisse

6.4.2.1 Imputation de la manifestation de volonté et responsabilité contractuelle

L'engagement contractuel présuppose un consensus réel ou *normatif*, une volonté de conséquence juridique déclarée expressément ou *fondée sur la confiance*. Le critère déterminant est l'adoption par une partie d'un comportement à partir duquel l'autre partie pouvait de bonne foi conclure à l'existence d'une telle volonté.[349] Aujourd'hui déjà, des contrats sont conclus en ligne grâce à l'utilisation de logiciels avec un degré élevé d'automatisation, sans que

[347] Voir le communiqué de presse sous https://unis.unvienna.org/unis/pressrels/2024/unisl362.html (consulté le 26 août 2024).

[348] Tous les documents de travail sont disponibles sous https://uncitral.un.org/fr/working_groups/4/electronic_commerce (consulté le 26 août 2024).

[349] ATF **116** II 695 consid. 2.

cela ne pose de problème pour l'attribution de la manifestation de volonté. La question de savoir si l'autonomie croissante des applications d'IA pourrait entraîner à l'avenir des lacunes dans la réglementation, par exemple si la manifestation de volonté ne peut plus être clairement attribuée à la responsabilité d'une partie contractante, est discutée de manière isolée dans la doctrine.[350]

Lorsqu'une action est intentée sur la base d'une violation du contrat, il existe un allégement pour les personnes lésées dans la mesure où elles n'ont pas à prouver la faute de la personne qui a violé le contrat. Au contraire, la personne qui n'a pas exécuté le contrat ou alors imparfaitement doit prouver qu'aucune faute ne lui est imputable (art. 97, al. 1, CO). Là encore, la doctrine se demande si, à l'avenir, il ne serait pas (trop) facile de se disculper en cas de dysfonctionnement d'une application d'IA en prouvant qu'aucun devoir de diligence n'a été violé. Il est question d'un rapprochement avec la responsabilité pour des auxiliaires visée à l'art. 101, al. 1, CO, qui permettrait d'attribuer les erreurs des applications autonomes à la partie qui les utilise.[351] Il n'y a toutefois pas de raison de penser qu'il existe une lacune dans la législation, car l'utilisation d'une application d'IA peut également constituer une violation du devoir de diligence dans certains cas.[352] Il ne faut pas non plus oublier que les applications d'IA n'ont pas de personnalité juridique propre, contrairement aux personnes auxquelles on confie l'exécution d'une obligation contractuelle. En fin de compte, il s'agit d'outils techniques utilisés par une partie contractante à ses propres risques. Une analogie hâtive avec les règles relatives aux représentants ou aux auxiliaires pourrait créer des lacunes en matière de responsabilité et ne semble pas appropriée en l'état actuel des choses.

6.4.2.2 Smart contracts

Le Conseil fédéral s'est prononcé sur les *smart contracts* dans son rapport du 14 décembre 2018 intitulé « Bases juridiques pour la *distributed ledger technology* et la *blockchain* en Suisse – État des lieux avec un accent sur le secteur financier ».

Un s*mart contract* est un protocole informatique, généralement basé sur un système de blockchain décentralisé, qui permet l'exécution autonome d'un contrat entre deux ou plusieurs parties dont les données ont été préalablement introduites dans le code informatique.[353] D'après l'inventeur de ce concept, la forme la plus simple de *smart contract* est le distributeur automatique qui libère la marchandise dès que le prix a été payé.[354] La doctrine majoritaire admet que malgré son nom, un *smart contract* n'est pas un contrat au sens du

---

[350] MICHAEL MARTIN KIANIČKA, Die Agentenerklärung. Elektronische Willenserklärungen und künstliche Intelligenz als Anwendungsfalls der Rechtsscheinhaftung, Zurich 2012; MALTE GRÜTZMACHER/JÖRN HECKMANN, Autonome Systeme und KI – vom vollautomatisierten zum autonomen Vertragsschluss? Die Grenzen der Willenserklärung, Computer und Recht 2019, 553 ss.

[351] CHRISTAPOR YACOUBIAN, Digitale Systeme als «Erfüllungsgehilfen» - Relevanz der fehlenden Rechtsfähigkeit, PJA 2023, 412 ss ; CHRISTAPOR YACOUBIAN, Der Analogieschluss zu Art. 101 Abs. 1 OR und die Grenzen zulässiger Rechtsfortbildung, PJA 2024, 195 ss ; MELINDA F. LOHMANN/THERESA PRESSLER, Algorithmische Vertragserfüllung (Teil 1) – Eine zukunftsgewandte Betrachtung der Rechtsfiguren der Erfüllungsgehilfin und der Substitutin unter Analyse des Urteils BGer 4A_305/2021 vom 2. November 2021, RSJ 2023, 879 ss, 884 ss.

[352] Voir CORINNE WIDMER-LÜCHINGER, Apps, Algorithmen und Roboter in der Medizin: Haftungsrechtliche Herausforderungen, HAVE 2019, 3 ss.

[353] Définition inspirée de la doctrine relativement uniforme, voir notamment LEE BACON/GEORGE BAZINAS, « Smart Contracts » : The next big Battleground ?, Jusletter IT du 18 mai 2017, 2 ; MARKUS KAULARTZ/JÖRG HECKMANN, Smart contracts - Applications of Blockchain Technology, Computer und Recht 2016, 618 ss, 618 ; STEPHAN D. MEYER/BENEDIKT SCHLUPPI, "Smart Contracts" und deren Einordnung in das schweizerische Vertragsrecht, recht 2017, 204 ss, 207 ; ROLF H. WEBER, Smart Contracts : Vertrags- und verfügungsrechtlicher Regelungsbedarf?, sic! 2018, 291 ss.

[354] NICK SZABO, Formalizing and securing relationships on public networks, First Monday Internet Journal, 1997, 1.

droit des obligations, mais plutôt une « technologie » informatique d'exécution des contrats.[355] Des problèmes peuvent survenir en raison de la rigidité de l'exécution préprogrammée du contrat, qui rend impossible son adaptation au cas par cas, comme le prévoit le droit suisse – par exemple par la *clausula rebus sic stantibus*. En l'état actuel des choses, la doctrine recommande aux parties de prévoir, lors de la conclusion d'un *smart contract*, des mécanismes appropriés pour faire face à d'éventuels changements de situation et pour régler les litiges.[356] L'anonymat des parties, répandu dans le contexte de la *blockchain*, constitue un obstacle majeur à la justiciabilité d'éventuelles prétentions. Ce problème ne peut toutefois pas être résolu par la voie législative. Dans son rapport « Bases juridiques pour la *distributed ledger technology* et la *blockchain* en Suisse », le Conseil fédéral est arrivé à la conclusion qu'il y aura certainement d'autres développements dans le domaine des *smart contracts*, mais qu'ils ne font que commencer et ne requièrent pas impérativement une réglementation.

6.4.3 Appréciation

Les discussions sur l'utilisation des applications d'IA dans les relations contractuelles n'en sont qu'à leurs débuts et aucun problème n'est apparu dans la pratique jusqu'à présent. Les développements doivent être suivis de près, mais une intervention du législateur ne semble pas appropriée aujourd'hui.

6.5 Droit du travail

6.5.1 Introduction

Ce chapitre approfondit des questions relatives au droit du travail. En effet, l'utilisation de l'IA pour gérer les relations de travail s'est fortement développée et elle est courante aujourd'hui dans de nombreuses entreprises. Des questions juridiques nouvelles et des questions spécifiques se posent en outre dans ce domaine, au vu en particulier des nombreuses règles de protection des travailleurs qui le caractérisent. Enfin, ces développements ont déjà eu des conséquences au niveau juridique, car des règles spéciales ou des initiatives pour légiférer ont vu le jour, au niveau de l'UE ou au niveau suisse.

---

[355] ANDREAS FURRER, Die Einbettung von Smart Contracts in das schweizerische Privatrecht, Revue de l'avocat 2018, 103 ss, 109 ; GABRIEL OLIVIER BENJAMIN JACCARD, Smart contracts and the role of Law, Jusletter du 23 novembre 2017, n° 8-9 ; STEPHAN D. MEYER/BENEDIKT SCHLUPPI, Smart Contracts (n. 353), 208 ; ROLF H. WEBER, Leistungsstörungen und Rechtsdurchsetzung bei Smart Contracts : Eine Auslegeordnung möglicher Problemstellungen, Jusletter du 4 décembre 2017, ch. 2. Sur les controverses : HANS RUDOLF TRÜEB, Smart Contracts, in : Pascal Grolimund et al. (éds.), Festschrift für Anton K. Schnyder, Zurich 2018, 723 ss, 725.

[356] Voir ROLF H. WEBER, Leistungsstörungen (n. 355), N 33 ss.

6.5.2 Au niveau européen

6.5.2.1 Directive visant à améliorer les conditions de travail des travailleurs des plateformes

Il n'y a à la connaissance de l'OFJ pas de projet législatif au niveau de l'UE qui traite de manière générale de l'utilisation de l'IA dans les relations de travail. Par contre, l'UE est en passe d'adopter une directive visant à améliorer les conditions de travail des travailleurs des plateformes, qui contient également des règles concernant l'utilisation d'algorithmes dans ce cadre.[357] Les négociateurs du Parlement et du Conseil sont parvenus à un accord politique sur le texte le 8 février 2024.[358] Le Conseil de l'UE a confirmé cet accord le 11 mars 2024[359] et le Parlement européen en date du 24 avril 2024[360]. Il ne manque plus que l'adoption formelle par le Conseil des Ministres, et la signature des colégislateurs pour que le texte soit définitivement adopté.[361]

La directive vise à améliorer les conditions de travail et la protection des données à caractère personnel des personnes exécutant un travail via une plateforme, en promouvant notamment la transparence, l'équité, le contrôle humain et la responsabilité dans la gestion algorithmique du travail de plateforme (art. 1, par. 1 let. b). Elle contient aussi des règles visant à améliorer la protection des personnes physiques lors du traitement de leurs données personnelles en prévoyant des mesures en matière de gestion algorithmique (art. 1, par. 2). La directive s'applique au travail de plateforme effectué dans l'UE, indépendamment du siège de la plateforme et du droit applicable (art. 1, par. 3). Des plateformes basées en Suisse pourraient donc devoir appliquer ce texte si des prestataires accomplissent leur travail pour elles dans l'UE. Le travail pour une plateforme digitale, selon la définition de la directive, couvre des prestations offertes par des canaux digitaux comme un site Internet ou une application mobile (art. 2, par. 1, point 1, let. a). La directive ne concerne donc pas les entreprises qui organisent leur travail de manière traditionnelle. L'utilisation d'instruments de monitorage ou de prise de décision automatisée fait également partie des éléments constitutifs cumulatifs de la définition

---

[357] Proposition de Directive du Parlement européen et du Conseil relative à l'amélioration des conditions de travail dans le cadre du travail via une plateforme, 9 décembre 2021, COM(2021) 762 final, disponible sous https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A52021PC0762 (consulté le 28 août 2024).

[358] Cf. https://www.europarl.europa.eu/news/en/press-room/20240205IPR17417/provisional-deal-on-first-eu-wide-rules-for-platform-workers (consulté le 28 août 2024) et https://data.consilium.europa.eu/doc/document/ST-7212-2024-ADD-1/fr/pdf. (consulté le 28 août 2024).

[359] Cf. https://www.consilium.europa.eu/fr/press/press-releases/2024/03/11/platform-workers-council-confirms-agreement-on-new-rules-to-improve-their-working-conditions/ (consulté le 28 août 2024).

[360] Cf. https://www.europarl.europa.eu/news/fr/press-room/20240419IPR20584/le-parlement-adopte-la-directive-sur-le-travail-des-plateformes (consulté le 28 août 2024).

[361] Le Parlement a adopté, en date du 4 juillet 2024, un rectificatif (version française du 8 juillet 2024 disponible sous https://www.europarl.europa.eu/doceo/document/TA-9-2024-0330-FNL-COR01_FR.pdf) qui apporte des modifications rédactionnelles ou formelles au texte adopté le 24 avril 2024. Les différences avec le texte adopté le 24 avril sont indiquées en note en bas de page.

de plateforme digitale (art. 2, par. 1, point 1, let. d). Les instruments de soutien à la décision sont inclus (art. 2, par. 1, point 9[362]).

Le chapitre III (art. 7 ss) traite de la gestion algorithmique et prévoit notamment les droits et obligations suivants :

- L'art. 7, par. 1, interdit la collecte ou le traitement de données liées à la vie privée du travailleur par des systèmes de monitorage ou de décision automatisée, comme les conversations privées, la prédiction de la manière dont le travailleur va exercer ses droits fondamentaux ou les données sur des caractéristiques personnelles sensibles (par ex., origine raciale ou ethnique, état de santé), ou le traitement de données biométriques ou liées à l'état émotionnel ou psychologique.

- L'art. 8 établit que le traitement de données personnelles par un système de monitorage ou de prise de décision automatisé est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques au sens de l'art. 35, par. 1, RGPD et implique l'obligation d'effectuer une analyse d'impact en matière de protection des données. L'analyse d'impact est remise aux représentants des travailleurs (art. 8, par. 2).

- L'art. 9 prévoit une obligation d'informer sur les systèmes de monitorage ou de décision automatisés, même lorsque les décisions n'ont pas une incidence notable sur les personnes exécutant un travail via une plateforme.

- L'art. 10 prévoit une obligation de surveillance et d'évaluation des systèmes automatisés. Certaines décisions, comme la suspension ou la terminaison de la relation contractuelle, doivent être prises par un être humain (art. 10, par. 5).

- L'art. 11 prévoit une obligation d'expliquer une décision prise par un système automatisé et le droit de faire réviser et le cas échéant de rectifier la décision (art. 11, par. 2 et 3).

- L'art. 12 prévoit des obligations en matière d'évaluation et de prévention des risques pour la santé et la sécurité.

- L'art. 13 prévoit une obligation d'informer et de consulter les représentants des travailleurs sur des décisions portant sur l'introduction ou la modification substantielle de systèmes automatisés. La représentation du personnel est par ailleurs destinataire de l'obligation d'informer prévue à l'art. 9, est impliquée dans la surveillance et l'évaluation des systèmes d'IA (art. 10) et peut faire réviser des décisions prises par de tels systèmes (art. 11).

[362] Art. 2, par. 1, let. j selon le rectificatif.

- L'art. 21 prévoit enfin une obligation de produire tout moyen de preuve relevant et les art. 22 et 23 requièrent une protection contre des représailles ou le licenciement, suite à l'invocation des droits résultant de la directive.

- Les règles de la directive sont impératives, c'est-à-dire que des règles moins favorables ne sont pas admises, que ce soit dans les législations nationales des États membres ou dans des conventions collectives (art. 26, par. 2). Des règles plus favorables sont par contre possibles. La directive s'écarte volontairement du RGPD sur certains points. En particulier, elle exclut le consentement comme fait justificatif d'un traitement de données (voir consid. 40).

6.5.2.2 Règlement de l'UE sur l'intelligence artificielle et droit du travail

Le règlement sur l'IA (cf. ch. 5), malgré son caractère général, contient des règles qui concernent spécifiquement le monde du travail. Il procède en effet à une classification d'applications destinées au monde du travail suivant les degrés de risques définis.

Des systèmes qui infèrent les émotions sont ainsi classés dans la catégorie des applications interdites s'ils sont utilisés sur la place de travail, sauf si l'application est utilisée à des fins médicales ou de sécurité (art. 5, par. 1, point f). Une série d'applications sont ensuite classées comme étant à haut risque (art. 6, par. 2 et Annexe III, point 4) : les systèmes servant au recrutement ou à la sélection de personnes physiques, notamment pour le placement ciblé d'offres d'emploi, pour l'analyse et le filtrage de dossiers de candidature et pour l'évaluation des candidats ; les systèmes destinés à prendre des décisions influençant les conditions de travail, la promotion ou la fin des rapports de travail ou destinés à attribuer des tâches sur la base de comportements individuels, de traits de la personnalité ou de caractéristiques personnelles, et les systèmes de contrôle et d'évaluation. Les obligations attachées à la classification dans la catégorie à haut risque, de même que les exceptions à cette classification, ont été exposées dans la partie consacrée au règlement sur l'IA (cf. ch. 5.2.7).

La question se pose de savoir si l'employeur peut être destinataire d'obligations définies dans le règlement sur l'IA, notamment celles prévues au chapitre 3 du règlement concernant la catégorie à haut risque. L'employeur sera en principe une personne qui déploie le système d'IA (déployeur) au sens de l'art. 3, point 4 du règlement sur l'IA. L'art. 26, par. 7, du règlement prévoit ainsi notamment que les employeurs informent les représentants des travailleurs et les travailleurs concernés qu'ils seront soumis à l'utilisation d'un système d'IA à haut risque (voir pour d'autres obligations à la charge des déployeurs le ch. 5.2.7.3.3). Cependant, sur la base de l'art. 25, par. 1, du règlement sur l'IA, l'employeur pourra également être considéré comme un fournisseur, car toute personne qui fait des modifications substantielles à un système à haut risque ou qui en modifie l'utilisation prévue sera considérée comme un fournisseur avec les obligations qui s'y rapportent.

L'interaction entre ces règles générales et les règles prévues en droit du travail au niveau de l'UE a été discutée.[363] Il est tout d'abord clair que le règlement a un caractère horizontal et donc subsidiaire, et que toutes les obligations découlant du droit du travail restent applicables. De plus, il se pourra que, d'une part, une action au moment de la conception du système intégrant l'IA soit nécessaire pour le rendre compatible avec les règles existantes de droit du travail. Toutefois, celles-ci incombent à l'employeur qui n'est pas nécessairement le fournisseur et concepteur du système. D'autre part, l'employeur qui décide d'introduire un système intégrant l'IA devra faire en sorte de respecter ses obligations découlant du droit du travail, ce qui implique qu'il dispose des informations et des moyens pour en adapter le fonctionnement si nécessaire.

6.5.3 Situation et discussions en droit suisse du travail

6.5.3.1 Motion 23.4492 Gysi « Intelligence artificielle. Renforcer les droits de participation des travailleurs »

La motion 23.4492 déposée le 22 décembre 2023 demande « de renforcer au niveau de la loi les droits de participation des travailleurs, lorsque [des] systèmes [algorithmiques] sont utilisés pour des recommandations, des prévisions, des décisions, etc. qui concernent les travailleurs ou qui utilisent des données sur ces derniers ». Pour ce faire « il convient d'élargir le droit d'être consulté, de renforcer les droits à l'information, de créer des droits de recours collectifs et d'examiner les possibilités de sanctions ». Cette motion se fonde sur une prise de position du syndicat syndicom et de l'ONG Algorithmwatch, elle-même basée sur une étude mandatée par ces organisations.[364]

6.5.3.2 Utilisation de l'intelligence artificielle dans le monde du travail en Suisse

Une enquête auprès des grandes entreprises suisses, doublée de 5 études de cas, menée en 2018 et renouvelée en 2020, donne une idée de l'extension de l'utilisation de l'IA dans les entreprises en Suisse.[365] Deux tiers environ de celles-ci indiquent utiliser de tels instruments et l'on constate une augmentation entre 2018 et 2020.[366]

S'agissant des types d'utilisation, les plus cités sont le recrutement, la fidélisation et la transition, la performance, l'aménagement du poste de travail, la compliance, ainsi que les décisions de sélection ou de licenciement. L'étude montre également les finalités d'utilisation les

---

[363] Cf. notamment AUDE CEFALIELLO/MIRIAM KULLMANN, Offering false security: How the draft artificial intelligence act undermines fundamental workers rights, European Labour Law Journal 2022, 542 ss.

[364] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden beim Einsatz von ADM-Systemen am Arbeitsplatz, novembre 2023, disponible sous https://syndicom.ch/fileadmin/user_upload/Web/Website/Dossiers/KI/2023_Rechtsgutachten_final.pdf (consulté le 28 août 2024).

[365] Cette enquête a été menée dans le cadre du Projet national de recherche « Big Data » (PNR 75), cf. https://www.nfp75.ch/fr (consulté le 28 août 2024).

[366] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 5.

plus fréquentes. Ce sont la fidélisation des collaborateurs et la transition entre anciens et nouveaux (63 %), la gestion de la performance (47 %) et le recrutement (39 %). Ainsi, un premier tri des candidatures pour le recrutement ou des outils de surveillance des frappes à l'ordinateur ou de l'utilisation d'Internet sont répandus. Il faut se demander dans ces cas quel est le rôle de l'IA voire même si ces outils entrent dans la définition de l'IA.[367] L'étude en question se focalise sur des systèmes de soutien à la décision (automatisation partielle) ou sur des systèmes totalement automatisés.

6.5.3.3 Défis posés par l'intelligence artificielle en droit du travail en Suisse

Les défis suivants sont identifiés en lien avec l'IA au travail, dans l'étude mentionnée ci-dessus ou dans d'autres publications :

- Protection des données[368] : Ces systèmes peuvent potentiellement récolter de manière massive et continue des données personnelles des travailleurs. La récolte peut de plus être indifférenciée et les finalités multiples ou mal définies.

- Protection de la santé : Les outils de mesure de performance ou de surveillance peuvent générer une atteinte à la santé psychique et au bien-être des travailleurs liée au fait de se sentir constamment évalué.[369] D'autre part, des situations de surmenage peuvent survenir si un outil intégrant l'IA surveille l'exécution des tâches sur la base de paramètres d'exécution qui ne tiennent pas compte de la réalité.

- Biais dans l'évaluation, rigidité de la prise de décision : Les paramètres mesurés ou saisis dans le système ne tiennent pas nécessairement compte de toutes les particularités d'une situation (obstacles sur le trajet usuel qui empêchent de transporter des charges dans le temps programmé, embouteillage qui ne permet pas d'effectuer un trajet selon une durée standard). L'utilisation de l'IA instaure également une logique de performance fondée sur des données chiffrées qui ne tiennent pas compte d'éléments importants, mais non chiffrables comme les qualités personnelles et les compétences sociales, ou alors des difficultés non mesurables.[370] Un logiciel qui saisit les frappes sur le clavier pourrait amener à des

---

[367] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 10.

[368] ISABELLE WILDHABER, Die Roboter kommen – Konsequenzen für Arbeit und Arbeitsrecht, RDS 2016, 315 ss, 346 s. ; ISABELLE WILDHABER, Répercussions de la robotique et de l'intelligence artificielle sur le lieu de travail, in : Jean-Philippe Dunand/Pascal Mahon/Aurélien Witzig (éds.), La révolution 4.0 au travail – Une approche multidisciplinaire, Genève/Zurich/Bâle 2019, 201 ss, 228 s. ; WOLFGANG DÄUBLER, Digitalisierung und Arbeitsrecht, Frankfurt am Main 2022, §9, N 5 ss.

[369] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 18 : 22 % des participants à l'enquête indiquent que le comportement des travailleurs et la collaboration entre eux sont surveillés. Egalement AUDE CEFALIELLO/MIRIAM KULLMANN, Offering false security (n. 363), 559.

[370] Sur cet aspect, voir JEAN-CHRISTOPHE SCHWAAB, Les nouvelles tendances en matière d'évaluation du personnel et le droit du travail, Revue de droit du travail et d'assurance-chômage, 2019, 103 ss, 114 s.

évaluations négatives infondées si la personne travaille sur un sujet difficile et s'arrête souvent pour réfléchir. Les discussions entre collègues, leur qualité et leur apport positif sont également difficiles à saisir au moyen d'instruments intégrant l'IA. Ces limites de la prise de décision automatisée dans le contexte du travail ont ainsi amené à réfléchir à diverses mesures pour y remédier, comme la nécessité d'une prise de décision ou d'un contrôle effectués par un humain, l'autonomie décisionnelle de l'humain par rapport à l'output de la machine ou encore le droit à l'explication de la décision émanant d'un système d'IA ou fondée sur ses appréciations, ou sa révision par un humain.[371] Comme exposé au ch. 6.5.2 ces remèdes ont été intégrés dans les règles adoptées dans l'UE. C'est également en partie le cas en Suisse avec la nouvelle LPD (voir ci-dessous ch. 6.5.3.4).

- Biais aboutissant à des discriminations : la question du potentiel de discrimination a été très discutée en lien avec l'intégration de l'IA dans les processus décisionnels au travail, notamment le recrutement.[372] Ce thème est traité de manière particulière dans la présente analyse (cf. ch. 4.3.2.5), mais est mentionné tout de même dans ce contexte. Les biais résultent notamment du fait que les masses de données et de décisions existantes, utilisées par les instruments intégrant l'IA pour « se former », peuvent contenir les biais discriminatoires existants dans la société.

- Transparence et motivation des décisions : la doctrine relève que des instructions données par des systèmes automatisés sont un phénomène qui est devenu une réalité du monde du travail.[373] Or, ces instructions doivent respecter le cadre légal du droit du travail, ce qui implique de savoir sur quelle base la décision est prise et donc de connaître la manière dont les systèmes d'IA sont programmés. Mais le travailleur qui souhaite contester une instruction ne le sait pas. Cela vaut pour toute autre décision prise par un système d'IA ou fondée sur un tel système. Ce thème a aussi été abordé de manière générale plus haut (cf. ch. 4.3.2.3).

6.5.3.4 Règles applicables en droit suisse

Les règles qui entrent en ligne de compte pour répondre à ces problèmes sont de divers ordres :

- Les règles sur la protection des données sont tout d'abord à mentionner. La LPD s'applique aux relations de travail dans le secteur privé. Aux règles générales s'ajoute l'art. 328*b* CO, qui règle spécialement la question pour le contrat de travail. L'art. 328*b*, 1re

---

[371] AIDA PONCE DEL CASTILLO, Le travail à l'ère de l'IA : pourquoi la réglementation est nécessaire pour protéger les travailleurs, etui. 2020, 13 ; JEREMIAS ADAMS-PRASSL/ HALEFOM ABRAHA/AISLINN KELLY-LYTH et al., Regulating algorithmic management : A blueprint, European Labour Law Journal 2023, 124 ss, 139-140, 143.

[372] Entre autres ISABELLE WILDHABER, Répercussions de la robotique (n. 368), 213 ss. Voir aussi https://www.rts.ch/info/sciences-tech/13464935-la-place-de-lintelligence-artificielle-dans-le-recrutement.html, https://www.humanrights.ch/fr/nouvelles/discrimination-algorithmique-protection , https://algorithmwatch.ch/de/findhr/ (consultés le 28 août 2024).

[373] ISABELLE WILDHABER, Répercussions de la robotique (n. 368), 210 s.

phrase, CO définit deux finalités possibles pour la récolte des données par l'employeur : déterminer l'aptitude du travailleur à exécuter la prestation de travail et servir l'exécution du contrat de travail. Cette règle est de droit relativement impératif. La deuxième phrase rappelle l'application de la LPD.

La question de savoir si l'employeur peut récolter des données pour d'autres finalités que celles prévues à l'art. 328*b* CO a été longtemps controversée, particulièrement en lien avec le consentement du travailleur, qui est un motif justificatif selon la LPD (art. 31, al. 1). Si une bonne partie de la doctrine s'est accordée à dire que le consentement ne peut justifier un traitement de données à d'autres fins[374], le Tribunal fédéral a récemment admis cette possibilité[375] : les deux hypothèses énoncées à l'art. 328*b* CO posent une présomption de licéité du traitement, les traitements pour d'autres finalités devant reposer sur les faits justificatifs généraux prévus par la LPD. Malgré cette ouverture du Tribunal fédéral, beaucoup de réserves sont exprimées par la doctrine sur la possibilité de donner un consentement libre au sens de l'art. 6, al. 6, LPD, en droit du travail, ou dans toute situation impliquant un déséquilibre structurel entre les parties.[376]

Les problèmes posés par l'IA peuvent être appréhendés par le biais des règles générales (art. 328*b* CO et la LPD). Tout d'abord, les règles relatives à la finalité du traitement posent certaines limites. L'employeur qui utilise un système d'IA sera ainsi tenu par les finalités posées à l'art. 328*b* CO. Il devra justifier la collecte de données pour d'autres finalités. La finalité doit de plus, selon l'art. 6, al. 3, LPD, être déterminée et reconnaissable pour la personne concernée. Les autres principes de la LPD doivent aussi être respectés, notamment la proportionnalité (art. 6, al. 2, LPD). Le respect de ces principes constitue un point de tension avec les systèmes d'IA notamment en raison de la collecte souvent massive de données inhérente au développement des systèmes, en raison des finalités diverses qui sont poursuivies ou de l'utilisation à d'autres fins que celle prévue initialement. L'enquête susmentionnée montre ainsi que l'introduction de systèmes d'IA par certains employeurs a été abandonnée, car ces systèmes ne respectaient pas les principes ci-dessus.[377] Diverses dispositions de la LPD qui concernent particulièrement le monde du travail peuvent être mentionnées au surplus. Tout d'abord, les données collectées pourront être des données sensibles au sens de l'art. 5,

---

[374] Voir notamment Rapport du Conseil fédéral du 16 novembre 2016 en réponse au postulat 12.3166 Meier-Schatz, Conséquences juridiques du télétravail, ch. 7.7.1 et les réf. citées, disponible sous https://www.newsd.admin.ch/newsd/message/attachments/80239.pdf (consulté le 28 août 2024).

[375] Arrêt du TF, 4A_518/2020, consid. 4.2.4.

[376] CÉLIAN HIRSCH, Droit du travail et intelligence artificielle : défis des décisions automatisées pour les employeurs, in : Valérie Défago/Jean-Philippe Dunand/Pascal Mahon/Samantha Posse/David Raedler (éds.), La protection des données dans les relatons de travail à la lumière de la nouvelle loi fédérale sur la protection des données, Genève/Zurich 2024, 95 ss, 112 et les réf. citées ; ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 16.

[377] ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 8.

let. c, LPD, à l'exemple des opinions syndicales ou des données sur la santé, et/ou servir de base à du profilage à risque élevé. La collecte de telles données ainsi que les profilages à risque élevé requièrent le respect de conditions spécifiques, notamment, le cas échéant, le consentement exprès de la personne concernée (art. 6, al. 7, let. a, LPD) et l'obligation d'effectuer une analyse d'impact si le traitement de données sensibles est effectué à grande échelle (art. 22, al. 1 et 2, let. a, LPD).

Il convient encore de mentionner l'art. 21, al. 1 et 2, LPD, qui se rapporte spécifiquement aux systèmes d'IA et prévoit des droits et obligations d'information spécifiques en cas de décision individuelle automatisée (cf. ch. 4.3.2.3). On rappellera qu'en principe l'employeur sera libéré de ses obligations s'il obtient le consentement exprès du travailleur (art. 21, al. 3, let. b, LPD), sous réserve du caractère problématique du consentement en droit du travail, évoqué ci-dessus. La notion de décision individuelle automatisée est relativement étroite, car elle n'inclut pas les systèmes livrant une évaluation servant de base à une décision qui sera prise par un humain.[378]

- Les règles sur la protection de la santé peuvent intervenir à plus d'un titre. La protection de la santé est réglée principalement à l'art. 328 CO et à l'art. 6 LTr. L'obligation de protéger la santé est concrétisée dans les ordonnances relatives à la LTr.

  Une règle particulièrement pertinente dans ce contexte est l'art. 26 OLT 3, dont l'al. 1 interdit « des systèmes de surveillance ou de contrôle destinés à surveiller le comportement des travailleurs à leur poste de travail ». Il s'agit de tout système technique permettant d'enregistrer des données sur le comportement des travailleurs, comme des caméras, des dispositifs d'écoute des conversations téléphoniques ou de surveillance de l'activité à l'ordinateur, des systèmes de localisation ou des outils informatiques utilisant l'IA.[379] Ces systèmes de surveillance sont autorisés s'ils poursuivent d'autres buts légitimes, comme la sécurité des lieux ou le contrôle de qualité ou du rendement, pour peu qu'ils respectent le principe de proportionnalité et que les travailleurs concernés en aient été informés.[380]

  D'autres obligations sont pertinentes dans ce contexte : l'art. 6, al. 2, LTr précise que l'employeur doit protéger les travailleurs contre le surmenage et l'art. 2, al. 1, OLT 3 précise que la protection porte sur la santé physique et psychique, et qu'elle comporte notamment le fait d'éviter des efforts excessifs. Une obligation d'information et d'instruction incombe à l'employeur (art. 5, al. 1, OLT 3), de même que l'obligation d'informer et

[378] CÉLIAN HIRSCH, Droit du travail et intelligence artificielle (n. 376), 104 s., malgré le fait que la jurisprudence dans l'UE semble aller vers une interprétation large ; ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 15.

[379] SECO, commentaire OLT 3 (n. 126), art. 26, 326-2.

[380] ATF **130** II 425, consid. 4.4 ; SECO, commentaire OLT 3 (n. 126), art. 26, 326-1.

de consulter les travailleurs ou leurs représentants sur les questions relatives à la protection de la santé (art. 48, al. 1, let. a, LTr).

- La loi sur la participation[381] règle l'information et la consultation des travailleurs dans l'entreprise. Il s'agit d'une loi-cadre qui est concrétisée par une série de règles spéciales, dont certaines sont énumérées à l'art. 10. La loi instaure un droit d'élire une représentation du personnel dans les entreprises employant plus de 50 travailleurs (art. 3). Elle établit un droit, pour la représentation des travailleurs, d'être informée sur toutes les affaires dont la connaissance lui est nécessaire pour s'acquitter convenablement de ses tâches (art. 9, al. 1). Au-delà de ce droit général à l'information, les droits de participation dépendent des règles spéciales. Un droit général d'être consulté n'est en particulier pas prévu. L'art. 10 énumère ainsi quatre domaines où des droits de participation sont prévus : sécurité et santé au travail ; transfert de l'entreprise ; licenciements collectifs ; affiliation à une institution de la prévoyance professionnelle et résiliation de l'affiliation.

- Les limites posées au droit de donner des instructions seront aussi pertinentes : le droit de donner des instructions prévu à l'art. 321*d* CO doit en effet être conforme au droit impératif, notamment l'art. 328 CO (protection de la personnalité), au contrat et au principe de la bonne foi.[382] Toute forme d'instruction générée par une IA doit donc se situer dans ce cadre et le travailleur peut refuser de l'exécuter si ce n'est pas le cas. L'employeur ne peut en particulier pas invoquer le caractère automatisé de la décision pour se justifier ou reporter sa responsabilité sur le concepteur du système.[383] Ces obligations impliqueront donc pour l'employeur de pouvoir établir la licéité de la décision prise, ce qui requiert de connaître la manière dont le système d'IA fonctionne et est programmé et de pouvoir le communiquer au travailleur.

- Enfin, la protection de la personnalité de manière générale sera pertinente (art. 328 CO) : toute décision liée aux tâches à accomplir fondée sur une évaluation générée par l'IA pourra être attentatoire à la personnalité du travailleur si elle ne repose pas sur la réalité objective du travail accompli. Cela pourra aussi être le cas d'une affectation qui péjore la situation du travailleur ou d'une sanction quelconque.

6.5.3.5 Appréciation

L'adoption imminente de la directive sur les travailleurs de plateforme impliquera une divergence législative entre le droit du travail de l'UE et le droit suisse. Ce dernier ne prévoit en ef-

---

381 Loi fédérale sur l'information et la consultation des travailleurs dans les entreprises (RS **822.14**).

382 ATF **132** III 115, consid. 5.2.

383 ISABELLE WILDHABER, Die Roboter kommen (n. 368), 330 s.

fet pas de règles spéciales en matière de gestion algorithmique du travail, que ce soit spécifiquement pour les plateformes digitales qui offrent des prestations de travail ou de manière générale pour toute relation de travail.

Des lacunes sont identifiées dans la prise de position du syndicat syndicom et de Algorithmwatch[384], qui se fonde sur l'étude WILDHABER/EBERT citée plus haut. Ces positions sont reprises dans la motion 23.4492 Gysi. Ces demandes se focalisent sur la participation des travailleurs dans l'entreprise, qui doit être renforcée selon elles.

Certaines demandes visent un renforcement du droit de la participation de manière générale, sans que cela ne concerne spécifiquement l'avènement de l'IA. Les autres demandes se rapportent spécifiquement à l'utilisation de l'IA. C'est le cas d'éventuelles règles sur l'information et la consultation des travailleurs en cas de mise en place et d'utilisation de systèmes d'IA par l'employeur. Actuellement, au-delà du droit général à l'information prévu à l'art. 9, al. 1 de la loi sur la participation, de tels droits dépendent de l'existence de règles spéciales. Or, les seules qui sont pertinentes en cas de mise en place d'un système d'IA sont celles relatives à la protection de la santé (art. 48, al. 1, let. a, LTr ; cf. ch. 6.5.3.4). Des droits de participation ne pourront donc découler du droit en vigueur qu'indirectement, en cas d'atteinte potentielle à la santé. Par ailleurs, même si l'art. 9, al. 1 de la loi sur la participation prévoit un droit à l'information, sa formulation est très générale. Pour ce qui est des règles sur la protection des données, elles permettent de fonder un droit du travailleur à être informé, mais ce droit est de nature individuelle et n'existe pas à un niveau collectif, alors que les systèmes d'IA posent des questions qui concernent l'ensemble des travailleurs.[385] Cette dimension collective n'est peut-être pas spécifique à la relation de travail, mais le droit du travail contient des règles sur la représentation collective des intérêts des travailleurs.

Enfin l'analyse montre que les règles générales du droit du travail encadrent à divers niveaux la mise en place et de l'utilisation de systèmes d'IA par l'employeur. Ce dernier devra ainsi successivement se demander si son système est interdit sur la base de l'art. 26, OLT 3, s'il répond aux exigences de la protection des données, notamment en matière de finalité ou de proportionnalité, ou en lien avec des obligations plus spécifiques rattachées au traitement de données sensibles ou au profilage à risque élevé. Il devra ensuite respecter les obligations prévues à l'art. 21 LPD concernant les décisions individuelles automatisées. Il devra enfin faire en sorte de pouvoir motiver les décisions prises par ou à l'aide de systèmes d'IA, par exemple pour établir la licéité d'instructions données au travailleur ou encore pour motiver un licenciement si un travailleur le demande (art. 335, al. 2, CO). La situation légale actuelle a le mérite de poser des limites assez nombreuses, mais a pour inconvénient que chaque norme pertinente ne couvre qu'un aspect du problème et repose sur une notion différente, comme « système de surveillance ou de contrôle destiné à surveiller le comportement », « décision individuelle automatisée » ou « profilage à risque élevé ». Cette diversité rend la situation juridique compliquée et aura pour conséquence que divers régimes juridiques s'appliqueront aux systèmes d'IA selon qu'ils relèvent d'une, de plusieurs ou d'aucune catégorie définie par ces

[384] À télécharger sous https://syndicom.ch/fr/themes/dossier/intelligenceartificielleia/algorithmes-au-travail (consulté le 28 août 2024).

[385] Dans ce sens ISABELLE WILDHABER/ISABEL EBERT, Beteiligung der Arbeitnehmenden (n. 364), 16 s.

diverses règles. Il existe par ailleurs une incertitude sur la portée du consentement selon la LPD, dans les relations fondées sur un contrat de travail.

Droit de l'UE : il existe un écart au vu de la nouvelle directive sur la protection des travailleurs de plateforme. De plus, le droit suisse ne connaît pas d'équivalent aux règles spécifiques aux relations de travail contenues dans le règlement sur l'IA.

Les règles générales du droit du travail et de la protection des données permettent toutefois de protéger le travailleur. Malgré cela, le système normatif mis en place dans l'UE, qui accompagne les systèmes d'IA depuis leur conception, jusqu'à leur mise en place et tout au long de leur utilisation, n'a pas d'équivalent dans le droit en vigueur. Les règles spécifiques relatives à la participation des travailleurs n'ont également pas non plus d'équivalent en droit suisse, même si les règles actuelles sur la protection de la santé et la sécurité au travail couvrent une partie des situations.

Interventions parlementaires : Des lacunes sont également mises en avant par la motion 23.4492 Gysi et les documents qui lui ont servi de base. Une partie d'entre elles n'est pas spécifique à l'IA. Par contre, un droit à l'information et à la consultation spécifique à la mise en place et à l'utilisation de systèmes d'IA pourrait faire sens au vu du caractère très général de la loi sur la participation.

L'analyse au niveau des besoins de régulation en droit du travail doit se faire en coordination avec l'analyse générale sur les besoins de régulation en matière d'IA. Le comblement des besoins de légiférer transversaux identifiés ci-dessus, en particulier en lien avec une éventuelle ratification de la convention sur l'IA (cf. ch. 4.6), permettrait déjà d'appréhender un certain nombre de défis en lien avec le monde du travail. On songe notamment aux principes de transparence et contrôle (art. 8), d'égalité et non-discrimination (art. 10), et de protection des données (art. 11). La portée de la convention sur les relations de droit privé constitue toutefois une limite. En effet, la convention se limite aux situations où les droits fondamentaux, la démocratie ou l'État de droit sont touchés (cf. ch. 4.2.3.1). Une seconde limite a trait à la participation des travailleurs et aux règles qui s'y rapportent, qui sont spécifiques au droit du travail.

En outre, dans l'hypothèse où la Suisse devait choisir de se rapprocher du règlement sur l'IA en prévoyant par exemple que certaines applications problématiques soient interdites dans le secteur privé, cela aurait aussi des incidences en droit du travail. Si la Suisse ne choisit pas cette voie, l'on pourra se demander si une reprise des règles spécifiques au droit du travail ferait sens.

En l'état, un besoin de légiférer spécifique en matière de droit du travail ne semble donc donné que sur certains aspects ponctuels. Il conviendra cependant d'observer l'évolution du cadre légal général afin d'intervenir si besoin avec des mesures particulières.

6.6 Droit pénal

6.6.1 Applicabilité de principe

Le code pénal suisse (CP) est rédigé en suivant le principe de neutralité technologique, de sorte qu'il reste applicable quel que soit la technologie ou l'instrument utilisé par l'auteur.[386] La portée matérielle de la protection est donc en principe à même et conçue pour couvrir l'utilisation de différentes technologies par un auteur ou un groupe d'auteurs. Si l'auteur se sert de systèmes d'IA pour commettre un délit contre l'honneur ou la sphère privée, par exemple au moyen de *deepfakes*, les infractions correspondantes (art. 173 ss CP, et en particulier l'usurpation d'identité au sens de l'art. 179^{decies} CP) sont applicables. Si l'auteur suit de plus vastes desseins par l'utilisation de systèmes d'IA, par exemple l'obtention d'un avantage patrimonial par la tromperie, les infractions contre le patrimoine sont également applicables lors de l'utilisation d'une nouvelle technologie en cours de développement (p. ex. l'escroquerie au sens de l'art. 146 CP ou la falsification de marchandises au sens de l'art. 155 CP). Tant que l'auteur utilise des systèmes d'IA et remplit les conditions de l'infraction, la responsabilité reste en principe celle de l'individu ou du cercle d'auteurs qui utilise cette technologie. Si les technologies sont particulièrement sophistiquées dans certains cas ou si la subtilité de la tromperie augmente, comme dans l'exemple des *deepfakes*, il est tout à fait possible, lors de la fixation de la peine, de reprocher à l'auteur un procédé particulièrement astucieux ou une intention de tromper et d'envisager une augmentation de la peine.

Les dispositions qui sanctionnent les infractions contre l'intégrité sexuelle sont également conçues en principe de manière technologiquement neutre. Pour les *deepfakes*, l'art. 197 CP (pornographie) est notamment pertinent. En principe, la production, la possession et la diffusion de *deepfakes* pornographiques, considérés comme de la « pornographie dure », sont punissables en application de l'art. 197, al. 4 et 5, CP.[387] Indépendamment de la technique utilisée, tous les actes de production de pornographie constituent l'infraction, c'est-à-dire non seulement les nouvelles fabrications, mais aussi les copies, reproductions ou adaptations de représentations pornographiques créées d'une autre manière.[388] Le CP accorde une importance particulière à l'intégrité sexuelle des enfants et des adolescents dans ce contexte et sanctionne également les objets et représentations ayant comme contenu des actes d'ordre sexuel *non effectifs* avec des mineurs (p. ex. des images ou représentations générées par des moyens informatiques ; art. 197, al. 4 et 5, CP). Cela signifie que même les représentations pornographiques avec des mineurs inventées de toutes pièces au moyen d'un système d'IA constituent généralement l'infraction. En outre, les articles 187 et 197 CP offrent un point

---

[386] Voir à ce sujet l'avis du Conseil fédéral du 16 août 2023 sur la Mo. 23.3563, Réglementer les « deep fakes », disponible sous https://www.parlament.ch/fr/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20233563 (consulté le 29 août 2024).

[387] Pour plus de détails, voir BRIGITTE TAG/MARTIN WYSS, Die strafrechtliche Einordnung von pornografischen Deepfakes, Jusletter du 29 avril 2024, 1 ss.

[388] ULRICH WEDER, in : Andreas Donatsch (éd.), StGB/JStG, Mit weitere Erlassen und Kommentar zu den Strafbestimmungen des SVG, BetmG, AIG und OBG, Orell Füssli Kommentar, Zurich 2022, Art. 197 N 18, et les réf. citées ; PK StGB-TRECHSEL/BERTOSSA, Art. 197 N 15 ; BRIGITTE TAG/MARTIN WYSS, Die strafrechtliche Einordnung (n. 387), 11.

d'ancrage pour la punissabilité des infractions commises à l'étranger sur des mineurs (art. 5 CP).

Les dispositions qui sanctionnent la provocation publique au crime ou à la violence (art. 259 CP) ou la discrimination et l'incitation à la haine (art. 261bis CP) sont également applicables quelle que soit la technologie utilisée pour rabaisser ou discriminer une personne ou un groupe de personnes et quel que soit le procédé utilisé pour inciter autrui à commettre un acte de violence.

6.6.2 Responsabilité pénale

6.6.2.1 Remarques générales

Le droit pénal suisse, et donc la responsabilité pénale, repose sur le principe de la culpabilité.[389] Pierre angulaire du droit pénal, la culpabilité est la condition préalable à toute sanction (art. 19, al. 1, CP) et la base de la fixation de la peine (art. 47, al. 1, CP). Par conséquent, il n'y a en principe pas de peine sans culpabilité (*nulla poena sine culpa*). La responsabilité pénale est fondamentalement liée à la capacité de l'auteur à commettre une faute, ce qui signifie qu'une technologie, une machine ou un système (comme un robot) ne sont pas en soi des sujets de droit capables de porter une responsabilité pénale. La doctrine s'est certes déjà demandé si une technologie autonome peut ou doit être tenue pour coupable et donc pénalement responsable, et si les principes fondamentaux du droit pénal doivent être modifiés dans ce sens, mais la majorité des auteurs y sont défavorables.[390] Il est tout sauf certain qu'une telle approche réglementaire du droit pénal apporterait une valeur ajoutée. En effet, les systèmes d'IA ne sont en principe pas en mesure d'agir de manière coupable. Alors qu'au moment de l'infraction, il est possible pour un humain de décider et de concrétiser sa volonté de commettre un acte répréhensible ou de l'accepter, un système d'IA « décide » seulement à partir d'algorithmes et de jeux de données.[391]

Étant donné qu'un système d'IA ne peut pas être lui-même poursuivi pénalement, la question de l'imputation et du transfert de la responsabilité en cas de dommages se pose plutôt entre l'utilisateur et le développeur (et éventuellement le détenteur) du système d'IA automatisé (voire autonome). L'entreprise qui a développé le système pourrait être tenue pour pénalement responsable si le dommage a été causé dans l'exercice d'activités commerciales conformes aux buts de l'entreprise et ne peut être imputé à aucun collaborateur déterminé (art. 102 CP).

---

[389] ATF **123** IV 1, consid. 2.

[390] Sur la possibilité d'une responsabilité pénale primaire de la technologie, voir avec les réf. citées MONIKA SIMMLER/NORA MARKWALDER, Roboter in der Verantwortung ? - Zur Neuauflage der Debatte um den funktionalen Schuldbegriff, Zeitschrift für die gesamte Strafrechtswissenschaft 01/2017, 20 ss, 22 ; NORA MARKWALDER/MONIKA SIMMLER, Roboterstrafrecht, Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz, PJA 02/2017, 171 ss, 172.

[391] ANNA LOHMANN, Strafrecht im Zeitalter von Künstlicher Intelligenz, Der Einfluss von autonomen Systemen und KI auf die tradierten strafrechtlichen Verantwortungsstrukturen, Baden-Baden 2021.

En principe, les fabricants doivent respecter les mêmes devoirs de vigilance à l'égard des systèmes d'IA que pour les autres produits techniques. Toutefois, ces obligations de vigilance peuvent varier en fonction du type d'IA, de son domaine d'application et du risque qui en résulte pour les biens juridiques protégés par le droit pénal.

L'un des défis liés à l'utilisation de l'IA est l'attribution de la responsabilité pénale lorsqu'une machine prend le contrôle d'une situation de manière largement automatisée (voire autonome) et prend seule, à partir d'algorithmes, des « décisions » ou effectue des actions qui étaient auparavant exclusivement réservées à des êtres humains (par exemple dans le domaine médical, dans la circulation routière ou même dans le contexte d'un usage moderne des armes).

Plus un système basé sur l'IA est complexe, plus il sera difficile pour le juge de déterminer, en cas de dommage, où se situe le défaut et qui en est responsable (problème de la boîte noire). Contrairement aux produits traditionnels, les produits d'IA présentent différents niveaux de développement et d'automatisation, qui leur permettent parfois d'être largement « autonomes ». Dans une certaine mesure, c'est justement l'intérêt de ces produits. Certaines technologies fondent leur apprentissage automatique et leur traitement des données sur des algorithmes qui peuvent évoluer et s'adapter *de manière autonome*.[392] Par conséquent, les processus d'un produit basé sur l'IA (ou « intelligent ») ne sont pas toujours prévisibles *ex ante* et ne sont parfois pas même transparents ou compréhensibles *ex post*.[393] L'attribution concrète des risques et des causes prévisibles, et dès lors de la responsabilité pénale, est souvent très complexe et dépend d'une multitude de facteurs spécifiques à chaque cas, comme le comportement concret de l'utilisateur et la prévisibilité et le caractère évitable des défauts du système. Des questions épineuses subsistent sur ces points, auxquelles il n'est pas possible de répondre une fois pour toutes. Compte tenu des développements significatifs en matière d'IA, les questions de délimitation trouveront probablement une réponse plus précise en fonction des évolutions technologiques concrètes, de leur utilisation et de la nature du danger.

6.6.2.2 Exemple : utilisation de systèmes d'intelligence artificielle dans la conduite automatisée

Dans le cas de la conduite automatisée, le véhicule prend en grande partie le contrôle, ce qui peut entraîner des accidents en cas d'erreur de détection d'un objet et d'évitement de la collision par la machine. De tels accidents, qui impliquent non seulement le conducteur, mais aussi des systèmes d'IA, représentent un défi pour l'attribution de la responsabilité pénale. En principe, le conducteur doit maîtriser son véhicule à tout moment, conformément à l'art. 31

---

[392] MARCO SCHREYER/ANITA GIERBL/T. FLEMMING RUUD/DAMIAN BORTH, Stichprobenauswahl durch Anwendung von Künstlicher Intelligenz, Expert Focus 2/22, 10 ss, 13 ss ; OMLOR SEBASTIAN, Methodik 4.0 für ein KI-Deliktsrecht, in : Zeitschrift zum Innovations- und Technikrecht 04/2020, 221 ss, 221.

[393] LEA BACHMANN, Prozedurale Entlastung von Herstellern « smarter » Produkte im Strafrecht ?, Revue suisse de droit pénal 140/2022, 77 ss, 78.

LCR. Ce principe peut toutefois s'avérer beaucoup plus complexe à appliquer dans le contexte des systèmes d'IA et il est au moins partiellement relativisé dans le cas des véhicules automatisés.[394] L'imputabilité d'un dommage dépend toujours de la capacité du conducteur à influencer le véhicule. Il y a donc faute lorsque cette personne, en dépit de ses obligations, de ses possibilités et de ce qui est raisonnablement exigible d'elle, n'interrompt pas ou n'empêche pas le comportement erroné d'un système basé sur l'IA, ou omet d'atténuer ou d'annuler le dommage. Dans le contexte de la conduite automatisée, la responsabilité pénale tend à se déplacer vers le fabricant au fur et à mesure que l'automatisation des véhicules progresse (*accountability shift*).[395] En d'autres termes, la responsabilité pénale peut, selon les cas, incomber aussi bien à l'utilisateur qu'au fabricant (et, conformément à la LCR, également au détenteur).[396]

L'utilisateur d'un véhicule automatisé peut être tenu pénalement responsable s'il ne respecte pas son devoir de vigilance, c'est-à-dire s'il utilise un véhicule automatisé en toute connaissance de cause ou par négligence, ou s'il ne le surveille pas de manière adéquate et qu'il était dans l'obligation de détecter et d'éviter une erreur.[397]

En revanche, le fabricant de véhicules automatisés porte la responsabilité du fait des produits (en principe quel que soit le degré d'automatisation)[398] et doit respecter certaines obligations de diligence en ce qui concerne le risque admissible, qui est parfois défini par des règlements techniques et des homologations. On part généralement du principe que le fabricant peut largement maîtriser le risque par une programmation et une instruction minutieuses.[399] Par conséquent, il peut être tenu pour responsable des dommages causés par des erreurs ou des défauts du produit s'il est prouvé qu'il est coupable du dysfonctionnement. Les dommages qui entrent en ligne de compte ne sont pas seulement ceux causés aux objets, mais aussi aux personnes, comme des lésions corporelles ou la mort. Si un dysfonctionnement était connu et n'a pas été corrigé, ou si un fabricant savait par exemple qu'un véhicule contrôlé par un système d'IA avait des difficultés à détecter des objets immobiles, qu'il n'a pas donné d'instructions détaillées au client potentiel et qu'un accident est survenu, le fabricant peut en principe être tenu pénalement responsable d'une violation de son devoir de diligence.[400] Dans le cas des systèmes d'IA qui seront souvent entraînés à partir des données d'utilisation générées après leur mise sur le marché et où les utilisateurs influent sur l'IA en choisissant la méthode

---

394 Dans le cas des véhicules automatisés de niveau 3 et 4, les conducteurs sont partiellement déchargés de leurs obligations de vigilance et de maîtrise du véhicule. Le conducteur d'un véhicule de niveau 3 peut se consacrer à d'autres choses lorsque le système est activé, mais reste obligé de reprendre le contrôle du véhicule dès que le système l'y invite. Voir FF **2021** 3026, p. 12.

395 NADINE ZURKINDEN, Vertrauen in Fahrzeugautomatisierung als strafmindernder Umstand? Anmerkungen zur Urteilsbegründung des Regionalgerichts Emmental-Oberaargau vom 30. Mai 2018, PEN 17 16 DIP, Jusletter du 3 décembre 2018, 1 ss ; CHRISTOF RIEDO/STEFAN MAEDER, Die Benutzung automatisierter Motorfahrzeuge aus strafrechtlicher Sicht, in : Thomas Probst/Franz Werro (éds.), Strassenverkehrsrechts-Tagung des 21 et 22 juin 2016, 85 ss, 94.

396 Le détenteur a également une responsabilité primaire de détenteur pour les véhicules automatisés, c'est-à-dire une responsabilité pénale pour la surveillance du véhicule et des causes de danger (art. 93, al. 2, let. b, 95, al. 1, let. e, et 96, al. 3, LCR). Le détenteur est donc en principe (co)responsable, entre autres, de la sécurité du véhicule.

397 Pour plus de détails sur ce point et sur la question du degré d'automatisation à partir duquel un utilisateur peut encore être considéré comme un conducteur au sens de la LCR, voir CHRISTOF RIEDO/STEFAN MAEDER, Die Benutzung automatisierter Motorfahrzeuge (n. 395), 91 s.

398 La loi fédérale sur la sécurité des produits (LSPro, RS 930.11) et la loi fédérale sur la responsabilité du fait des produits (LRFP, RS 221.112.944) sont pertinentes dans ce contexte.

399 Voir l'aperçu proposé par MELINDA F. LOHMANN, Ein zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz, Warum die Schweiz ihr Produkthaftungsrecht aktualisieren muss, Haftung und Versicherung 04/2021, 120.

400 NADINE ZURKINDEN, Vertrauen in Fahrzeugautomatisierung (n. 395), 14 s.

d'apprentissage, les données d'entraînement et la durée du processus d'apprentissage,[401] la possibilité de contrôle et l'influence des fabricants peut être plus restreinte.[402] En ce qui concerne les véhicules automatisés, il convient toutefois de noter que le logiciel du système d'IA fait partie intégrante de la réception par type et que celui-ci ne peut pas évaluer lui-même les connaissances acquises par l'entraînement. Au contraire, le constructeur doit faire évoluer le logiciel sur la base des nouvelles connaissances, par exemple par des mises à jour, et obtenir au besoin une nouvelle réception par type. Toutefois, si l'influence de l'utilisateur ou son interaction avec le système d'IA est à l'origine d'un dommage, elle peut néanmoins avoir une incidence sur la responsabilité pénale de l'utilisateur.[403]

6.6.3 Difficultés d'application du droit

Les explications ci-dessus concernant les dispositions de droit pénal montrent que le droit matériel en vigueur, grâce à sa neutralité technologique, est fondamentalement à même de sanctionner les infractions liées à l'utilisation de systèmes d'IA. Le droit pénal reste applicable et opérationnel indépendamment de la technologie ou des instruments spécifiques employés par l'auteur. En outre, le droit pénal actuel prévoit des devoirs de diligence et des responsabilités (pour les fabricants, les détenteurs et les utilisateurs). Toutefois, comme dans d'autres domaines du droit, il est difficile d'identifier les auteurs des infractions commises à l'aide de systèmes d'IA ou de dommages pénalement répréhensibles causés par l'IA, tout comme il est complexe de démontrer que les défauts du système étaient prévisibles et évitables lorsque des devoirs de diligence existent. Identifier les auteurs, obtenir des éléments de preuves, déceler et démontrer les sources d'erreurs et les violations du devoir de diligence constituera sans aucun doute un défi particulier, y compris dans le contexte pénal. Dans la pratique, les systèmes d'IA posent des problèmes considérables en termes de transparence et de traçabilité de leurs étapes de décision et d'apprentissage automatisées. Les difficultés ne sont donc pas nécessairement liées au droit matériel lui-même, mais souvent à l'application du droit lorsque l'auteur, les technologies et les preuves qu'elles recèlent ou les fabricants se trouvent à l'étranger, ou que la cause exacte du dommage ne peut pas être déterminée précisément après coup (problème de la boîte noire). L'application du droit suisse peut également atteindre ses limites lorsque les auteurs ou les développeurs à l'étranger sont soumis à une législation moins stricte ou que les modes de fonctionnement complexes et les liens de causalité des systèmes d'IA sont difficiles à élucider, à comprendre et à justifier par le ministère public ou le juge. Cela signifie que le problème de la boîte noire des systèmes automatisés, d'une part, et l'anonymat des auteurs d'infractions à l'étranger, d'autre part (par ex. en cas de

---

401 HERBERT ZECH, Entscheidungen digitaler autonomer Systeme : Empfehlen sich Regelungen zu Verantwortung und Haftung ? Gutachten für den 73. Deutschen Juristentag, Munich 2020, A 35 ss ; MELINDA F. LOHMANN, Roboter als Wundertüten - eine zivilrechtliche Haftungsanalyse, PJA 2017, Sonderheft Roboterrecht, 152 ss, 158.

402 HERBERT ZECH, Entscheidungen (n. 401), A 89 ; MELINDA F. LOHMANN, Roboter (n. 401), 158.

403 NADJA BRAUN BINDER/THOMAS BURRI/MELINDA FLORINA LOHMANN/MONIKA SIMMLER/FLORENT THOUVENIN/KERSTIN NOËLLE VOKINGER, Künstliche Intelligenz : Handlungsbedarf im Schweizer Recht, Jusletter du 28 juin 2021, 21.

deepfakes), entraînent des difficultés dans la pratique qui ne peuvent pas nécessairement être surmontées par des normes matérielles plus poussées.

Si l'auteur ou le fabricant se trouve à l'étranger, le principal problème pour l'obtention d'éléments de preuves est généralement l'absence de coopération internationale fonctionnelle et rapide, plutôt qu'une lacune du droit national lui-même. La Suisse a donc participé à l'élaboration d'une convention des Nations Unies sur la cybercriminalité et elle est partie à la convention correspondante du Conseil de l'Europe du 23 novembre 2001[404]. Ces accords portent essentiellement sur la sécurité et la coopération internationales dans le cadre de la cybercriminalité transfrontalière et visent à améliorer l'application du droit. Bien que la convention sur la cybercriminalité du Conseil de l'Europe, qui est entrée en vigueur pour la Suisse le 1er janvier 2012, prévoie déjà quelques instruments dans ce domaine, d'autres développements sont en cours. Ainsi, les règles de l'UE en matière d'e-evidence représentent une étape importante pour l'accès transfrontalier aux données qui constituent des éléments de preuves dans le cadre de procédures pénales en Europe.[405] Le système du US CLOUD Act suit une approche similaire.[406] La Suisse se penche sur les développements au sein de l'UE et aux États-Unis et décidera en temps utile de la manière dont elle souhaite les aborder, en tenant compte des possibilités politiques et juridiques.

6.6.4 Résumé et perspectives

Les considérations qui précèdent montrent que le droit pénal matériel offre en principe un cadre juridique adéquat et tout à fait applicable pour délimiter les responsabilités pénales et les devoirs de diligence lors de l'utilisation de systèmes d'IA. Le droit en vigueur se fonde sur le principe de culpabilité, qui reste pertinent dans le contexte de ces systèmes, tant pour les infractions intentionnelles que pour celles commises par négligence.[407] Cela signifie que le facteur déterminant dans le cas d'espèce est la culpabilité du ou des responsables potentiels. Dans le cas d'infractions intentionnelles qui visent à causer des dommages ou à obtenir un avantage patrimonial à l'aide de technologies d'IA, comme les *deepfakes*, c'est en général moins la situation juridique matérielle que l'application du droit qui constitue un défi, car les

---

404 RS **0.311.43**.

405 Voir le rapport sur le projet e-Evidence de l'UE, avis de l'Office fédéral de la justice du 24 octobre 2023 (disponible sous : www.ofj.admin.ch > Publications & services > Rapports, avis de droit et décisions > Rapports et avis de droit > Rapport sur le projet e-evidence de l'UE, consulté le 29 août 2024).

406 Voir le rapport sur le US CLOUD Act, avis de l'Office fédéral de la justice du 17 septembre 2021 (disponible sous : www.ofj.admin.ch > Publications & services > Rapports, avis de droit et décisions > Rapports et avis de droit > Rapport sur le US CLOUD Act, consulté le 29 août 2024) ainsi que le « Data Access Agreement » entre les Etats-Unis et le Royaume-Uni qui en découle et qui est entré en vigueur en octobre 2022.

407 MONIKA SIMMLER, Strafrechtliche Verantwortung im Zeitalter autonomer Technik: Vom Individual- zum Unternehmensstrafrecht?, in : Daniel Fink et al. (éds.), Strafjustiz zwischen künstlicher Intelligenz und prädiktiven Algorithmen, Bâle 2021.

auteurs peuvent souvent dissimuler leurs traces numériques et se cacher à l'étranger ou agir sous couvert d'anonymat.

En ce qui concerne la responsabilité pénale par négligence en rapport avec un système d'IA, il importe en premier lieu de savoir si son utilisation, son développement ou sa mise en service constitue en soi une violation d'un devoir de diligence ou le dépassement d'un risque admissible, mais aussi de déterminer quels devoirs de diligence et sources d'erreurs étaient évitables *ex ante*.[408] Dès lors, une responsabilité objective aggravée sans faute en cas de dommage est en principe exclue en droit pénal.[409] Par conséquent, une personne ne peut être tenue pénalement responsable d'une violation de son devoir de diligence que si cette violation était identifiable et évitable.[410] Il est cependant impossible de délimiter clairement et concrètement les responsabilités (en particulier entre celles l'utilisateur, du fabricant et du détenteur), et il sera difficile, sinon inutile, de graver ces limites dans le marbre compte tenu de l'autonomie croissante des technologies et des processus. Par conséquent, l'interprétation et l'application concrètes de ces obligations de diligence resteront un défi. La jurisprudence jouera un rôle déterminant à cet égard.

Les récents développements dans le domaine de la conduite automatisée plaident en faveur d'un déplacement de la responsabilité pénale vers le fabricant ou les collaborateurs de l'entreprise du fabricant.[411] Il est évident que des questions pratiques de délimitation subsisteront concernant les sphères de risque et d'influence des constructeurs et des utilisateurs[412] (et, selon le domaine, des détenteurs) ainsi que les devoirs de diligence concrets dans le con-

[408] BRAUN BINDER et al., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (n. 403). Sur le risque admissible lors de l'utilisation de systèmes basés sur l'IA dans les véhicules autonomes, voir NADINE ZURKINDEN, Strafrecht und selbstfahrende Autos – ein Beitrag zum erlaubten Risiko, recht 2016, p. 144 ss.

[409] En revanche, dans certains domaines, comme la responsabilité civile liée aux véhicules automatisés, l'utilisation de l'IA peut donner lieu à certains cas de responsabilité sans faute de l'utilisateur ou du détenteur (voir l'art. 58 LCR). Ces responsabilités doivent toutefois être distinguées des responsabilités pénales. Sur ce point, voir MELINDA F. LOHMANN, Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts, Baden-Baden 2016, 211 ss.

[410] BRAUN BINDER et al., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (n. 403). Sur les questions similaires concernant les robots sociaux (autonomes), voir MONIKA SIMMLER/OLIVIA ZINGG, Rechtliche Aspekte sozialer Roboter, Gutachten im Auftrag der TA-SWISS, 2011, 19 ss et 48 ss.

[411] Voir à ce sujet NADINE ZURKINDEN, Vertrauen in Fahrzeugautomatisierung (n. 395), 1 ss, avec une référence aux considérants de la décision du Tribunal régional d'Emmental-Haute-Argovie du 30 mai 2018, PEN 17 16 DIP. Sur cette tendance au déplacement de la responsabilité, voir également MONIKA SIMMLER, Strafrechtliche Verantwortung (n. 407).

[412] BRAUN BINDER et al., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht (n. 403), 21.

texte de la responsabilité pénale des entreprises ou de la complicité par négligence, notamment dans le cas des systèmes largement automatisés, voire autonomes.[413] À cet égard, une clarification des devoirs de diligence en matière de sécurité, d'autorisation et d'utilisation de l'IA, par exemple par le biais d'exigences concrètes d'autorisation ou de sécurité, peut améliorer sa prévisibilité et donc contribuer à la sécurité du droit. On pensera en particulier aux devoirs de diligence, d'une part pour l'utilisation et la mise en circulation de systèmes d'IA par des particuliers ou des entreprises, et d'autre part en relation avec la sécurité des produits.

Dans le cadre de l'utilisation de systèmes d'IA dans la circulation routière, le Conseil fédéral a ouvert, lors de sa séance du 18 octobre 2023, la consultation sur deux nouvelles ordonnances par lesquelles il entend réglementer la conduite automatisée. La consultation a duré jusqu'au 2 février 2024. Il s'agit d'une part de l'ordonnance sur la conduite automatisée (OCA), pertinente dans le présent contexte, et d'autre part de l'ordonnance concernant les aides financières destinées à promouvoir des solutions innovantes pour la circulation sur la voie publique (OAFC).[414] Ces textes visent à établir différents devoirs de diligence pour le fabricant, le distributeur, le détenteur et l'utilisateur du véhicule et à punir d'une amende le non-respect d'obligations particulièrement importantes. Compte tenu de la rapidité des évolutions dans le domaine de la conduite automatisée, il convient d'attendre de pouvoir observer les effets de ces réglementations et d'évaluer leur valeur ajoutée et leurs interactions avec les développements internationaux dans ce domaine.

En résumé, le droit pénal suisse, grâce à sa neutralité technologique, est fondamentalement à même de sanctionner les infractions liées à l'utilisation de systèmes d'IA, en particulier dans le cas d'infractions intentionnelles.

Les plus grands défis sont donc en premier lieu la délimitation pratique des responsabilités, l'application de la loi et, de manière générale, la justiciabilité. La concrétisation des devoirs de diligence dans l'utilisation des systèmes d'IA, telle qu'elle est prévue par l'ordonnance sur la conduite automatisée (OCA), permettra de gagner en clarté, du moins dans ce domaine. Une intervention plus poussée du législateur semble prématurée à l'heure actuelle. Le cas échéant, il faudrait qu'elle vise un domaine d'application précis. En outre, les devoirs de diligence raisonnable prévus par l'OCA contribueraient déjà, dans leur forme actuelle, à la mise en œuvre des principes de transparence et de responsabilité découlant de la convention du Conseil de l'Europe sur l'IA (art. 8 et 9, voir les ch. 4.3.2.3 et 4.3.2.4), si la Suisse ratifie cette convention. Une transparence accrue pourrait en outre – dans la mesure du possible – résoudre du moins en partie la problématique de la boîte noire.

[413] Sur les défis correspondants, voir MONIKA SIMMLER/OLIVIA ZINGG, Rechtliche Aspekte sozialer Roboter (n.410), 55 ss ; MONIKA SIMMLER, Strafrechtliche Verantwortung (n. 407) ; BRAUN BINDER et al., Künstliche Intelligenz : Handlungsbedarf im Schweizer Recht (n. 403), 23.

[414] Communiqué de presse du 18 octobre 2023, Le Conseil fédéral souhaite autoriser la conduite automatisée, disponible sous www.admin.ch > Documentation > Communiqués > Le Conseil fédéral souhaite autoriser la conduite automatisée (consulté le 29 août 2024).

Compte tenu des évolutions en cours au niveau suisse, européen et international, il conviendra de poursuivre leur analyse et d'en tirer des conclusions sur la nécessité de légiférer en Suisse. Par exemple, le Conseil de l'Europe, à l'issue de la 78e session plénière du Comité européen pour les problèmes criminels (CDPC) en novembre 2020, a également envisagé de réglementer les systèmes d'IA dans le domaine du droit pénal.[415] Il n'est donc pas exclu qu'en plus de la Convention sur l'IA, des discussions et des développements sur des points spécifiques puissent suivre.

[415] L'intelligence artificielle et le droit pénal – Comité européen pour les problèmes criminels (https://www.coe.int/fr/web/cdpc/artificial-intelligence-and-criminal-law, consulté le 29 août 2024).

7 Conclusions

L'analyse a permis de confirmer que le droit suisse contient déjà des dispositions qui s'appliquent également en matière de systèmes d'IA. L'IA ne se développe donc pas dans un vide juridique.

Il convient ici de mentionner que l'OFJ n'a pas analysé de manière exhaustive l'entier du droit fédéral existant. L'analyse présente donc le plus souvent des conclusions intermédiaires, qu'il conviendra encore d'approfondir.

Comme indiqué dans la méthodologie, et afin d'éviter finalement de dupliquer l'analyse du Rapport du groupe de travail interdépartemental « Intelligence artificielle » de 2019[416], le besoin de légiférer a été évalué principalement à l'aune de l'hypothèse d'une ratification par la Suisse de la convention sur l'IA du Conseil de l'Europe. L'analyse fait en outre état de considérations juridiques sur le règlement sur l'IA en cas de rapprochement de la législation suisse avec le droit de l'UE. Ces deux hypothèses dépendent toutefois de décisions politiques. Par ailleurs, l'analyse a également porté sur la situation dans d'autres domaines du droit spécifiques, où des développements existent et où des défis se posent en matière d'IA.

Pour ce qui est de la convention sur l'IA, il apparaît que le droit suisse existant permettrait de mettre en œuvre en partie cette dernière si la Suisse devait la ratifier. Des interventions législatives semblent toutefois devoir être envisagées, notamment s'agissant du renforcement de la transparence, du cadre de gestion des risques et des impacts ou encore des mécanismes de contrôle (cf. ch. 4.6). S'agissant du secteur privé, l'analyse a démontré que la portée de la convention sur l'IA en droit privé se limite aux cas où un effet horizontal direct ou indirect des droits fondamentaux entre privés est reconnu ou devait l'être à l'avenir (cf. ch. 4.2.3.1).

Un rapprochement du droit suisse avec le règlement sur l'IA nécessiterait une intervention plus large du législateur. En effet ce texte contient des prescriptions très spécifiques sur les différents systèmes d'IA en fonction des risques qu'ils impliquent, avec de nombreuses obligations à charge des différents opérateurs. À ce stade de nombreuses questions restent ouvertes, notamment politiques, telle celle de savoir si l'adoption d'une législation équivalente permettrait d'accéder plus efficacement au marché de l'UE, en obtenant par exemple une reconnaissance mutuelle de la législation en matière d'IA dans le cadre d'une extension de l'ARM au domaine de l'IA (cf. ch. 5.4). Des analyses plus approfondies, permettant par exemple de préciser les contours d'une éventuelle législation suisse qui se rapprocherait du règlement sur l'IA, seraient en outre nécessaires en cas de volonté politique d'aller dans cette direction. Elles porteraient également sur les directives européennes en droit privé (cf. ch. 6.3 ss).

S'agissant des domaines du droit spécifiques qui ont été examinés (cf. ch. 6), l'analyse a démontré que certaines questions se posent, mais que, en principe, les règles en vigueur apportent des réponses. Une amélioration de la protection pourrait aussi passer par l'adoption

[416] Rapport défis de l'intelligence artificielle (n. 1).

de normes générales pour se conformer à la convention sur l’IA, par exemple renforçant la transparence des systèmes d’IA.

Pour aller plus loin dans l’analyse juridique et cibler les approfondissements à effectuer, il convient d’apporter des réponses politiques aux deux hypothèses de travail ci-dessus, à savoir celle de la ratification de la convention sur l’IA par la Suisse, et celle d’un éventuel rapprochement avec le règlement européen sur l’IA.

Table des abréviations

al.	alinéa
ARM	Accord entre la Confédération suisse et la Communauté européenne relatif à la reconnaissance mutuelle en matière d'évaluation de la conformité, RS **0.946.526.81**
art.	article
ATF	Recueil officiel des arrêts du Tribunal fédéral
BGÖ	Bundesgesetz über das Öffentlichkeitsprinzip der Verwaltung, **RS 152.3**
BSK [Loi] – AUTEUR	Commentaire bâlois
BV	Bundesverfassung, RS **101**
c.	contre
CAI	Commitee on artificial intelligence / Comité sur l'intelligence artificielle
CC	Code civil suisse, RS **210**
CEDH	Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales, RS **0.101**
CEN	Comité européen de normalisation
CENELEC	Comité européen de normalisation électrotechnique
cf.	*confer*
ch.	chapitre
CJUE	Cour de justice de l'Union européenne
CNAI	Réseau de compétences en intelligence artificielle
CNUDCI	Commission des Nations Unies pour le droit commercial international

CO	Loi fédérale complétant le Code civil suisse (Livre cinquième : Code des obligations), RS **220**
CourEDH	Cour européenne des droits de l'homme
consid.	considérant
convention 108+	Convention modernisée du Conseil de l'Europe pour la protection des personnes à l'égard du traitement des données à caractère personnel
convention sur l'IA	Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit
CP	Code pénal suisse, RS **311.0**
CPC	Code de procédure civile, RS **272**
CPS-N	Commission de la politique de sécurité du Conseil national
CR [Loi] – AUTEUR	Commentaire romand
Cst.	Constitution fédérale de la Confédération suisse, RS **101**
DDIP	Direction du droit international public
DETEC	Département fédéral de l'environnement, des transports, de l'énergie et de la communication
DFAE	Département fédéral des affaires étrangères
DFI	Département fédéral de l'intérieur
DFJP	Département fédéral de justice et police
DSG	Bundesgesetz über den Datenschutz, RS **235.1**
éd(s).	éditeur(s)
et al.	*et alii*
etc.	*et caetera*

ex.	exemple
ETSI	Institut européen des normes de télécommunication
FF	Feuille fédérale
modèle GPAI	general-purpose AI model / modèle d’intelligence artificielle à usage général
IA	intelligence artificielle
IPI	Institut Fédéral de la Propriété Intellectuelle
JO	Journal officiel de l’Union européenne
LA	Loi fédérale sur l’aviation, RS **748.0**
LAVS	Loi fédérale sur l’assurance-vieillesse et survivants, RS **831.10**
LBI	Loi fédérale sur les brevets d’invention, RS **232.14**
let.	lettre
LCo	Loi fédérale sur la procédure de consultation, RS **172.061**
LCR	Loi fédérale sur la circulation routière, RS **741.01**
LD	Loi fédérale sur les douanes, RS **631.0**
LDA	Loi fédérale sur le droit d’auteur et les droits voisins, RS **231.1**
LDP	Loi fédérale sur les droits politiques, RS **161.1**
LEg	Loi fédérale sur l’égalité entre femmes et hommes, RS **151.1**
LHand	Loi fédérale sur l’élimination des inégalités frappant les personnes handicapées, RS **151.3**
LMETA	Loi fédérale sur l’utilisation des moyens électroniques pour l’exécution des tâches des autorités, RS **172.019**

LPD	Loi fédérale sur la protection des données, RS **235.1**
LRFP	Loi fédérale sur la responsabilité du fait des produits, RS **221.112.944**
LRTV	Loi fédérale sur la radio et la télévision, RS **784.40**
LTr	Loi fédérale sur le travail dans l'industrie, l'artisanat et le commerce, RS **822.11**
LTrans	Loi fédérale sur le principe de la transparence dans l'administration, RS **152.3**
Mo.	motion
N	numéro(s) de paragraphe (de la source citée)
n.	note en bas de page (du présent document)
nbp	note en bas de page (de la source citée)
OAFC	Ordonnance concernant les aides financières destinées à promouvoir des solutions innovantes pour la circulation sur la voie publique
OCA	Ordonnance sur la conduite automatisée
OCDE	Organisation de coopération et de développement économiques
OCPD	Ordonnance sur les certifications en matière de protection des données, RS **235.13**
OFCOM	Office fédéral de la communication
OFJ	Office fédéral de la justice
OFROU	Office fédéral des routes
OFS	Office fédéral de la statistique
OLT 3	Ordonnance 3 relative à la loi sur le travail, RS **822.113**

OPDo	Ordonnance sur la protection des données, RS **235.11**
OR	Bundesgesetz betreffend die Ergänzung des Schweizerischen Zivilgesetzbuches (Fünfter Teil: Obligationenrecht), RS **220**
ORTV	Ordonnance sur la radio et la télévision, RS **784.401**
p. ex.	par exemple
PA	Loi fédérale sur la procédure administrative, RS **172.021**
Pacte I	Pacte international relatif aux droits économiques, sociaux et culturels, RS **0.103.1**
Pacte II	Pacte international relatif aux droits civils et politiques, RS **0.103.2**
par.	paragraphe
PFPDT	Préposé fédéral à la protection des données et à la transparence
réf.	référence(s)
règlement sur l'IA	Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles concernant l'intelligence artificielle et modifiant les règlements (CE) no 300/2008, (UE) no 167/2013, (UE) no 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle)
RGPD	Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données)
RS	Recueil systématique
s.	et suivant(e)(s)

SECO	Secrétariat d’État à l’économie
SGK [Loi]– AUTEUR	Commentaire st-gallois
SRC	Service de renseignement de la Confédération
ss	suivant(e)s
StGB	Schweizerisches Strafgesetzbuch, RS **311.0**
TA-SWISS	Fondation pour l’évaluation des choix technologiques
TF	Tribunal fédéral
TUE	Traité sur l’Union européenne
UE	Union européenne
ZGB	Schweizerisches Zivilgesetzbuch, RS **210**

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de justice et police DFJP

Office fédéral de la justice OFJ
Domaine de direction Droit public
Unité Projets législatifs I

Annexe 1

Le tableau ci-dessous a pour objectif de proposer une comparaison entre la convention sur l'IA et le règlement sur l'IA, lorsque cela est possible. La présentation n'est pas exhaustive. Les articles de la convention sur l'IA sont présentés dans la colonne de gauche et les dispositions du règlement sur l'IA qui peuvent être mises en relation avec les articles de la convention sur l'IA sur la droite. Seules les dispositions les plus pertinentes du règlement sur l'IA sont indiquées.

Convention sur l'IA du Conseil de l'Europe	**Règlement sur l'IA de l'Union européenne**
Art. 1, par. 1 Objet et but	Art. 1, par. 1 Objet
Art. 1, par. 2 Mesures graduées et différenciées	Approche fondée sur les risques (art. 5, art. 6 ss, art. 50, art. 51 ss)
Art. 2 Systèmes d'intelligence artificielle	Art. 3, point 1 La définition est sensiblement la même.
Art. 3 Champ d'application	Art. 2 Application dans le secteur public et privé. Comme la convention sur l'IA, le règlement sur l'IA ne s'applique pas aux systèmes d'IA utilisés exclusivement à des fins militaires, de défense ou de sécurité nationale. Exclusion de la recherche et du développement. Le règlement sur l'IA prévoit d'autres exceptions.
Art. 4 Protection des droits de l'homme	Art. 1, par. 1 Le règlement sur l'IA a pour but notamment d'améliorer le fonctionnement du marché intérieur de l'UE en lien avec les produits d'IA, tout en garantissant la protection des droits fondamentaux. Très peu de droits individuels, mais but de protéger les droits fondamentaux en établissant des règles sur la sécurité des produits. Art. 77 Pouvoirs des autorités de protection des droits fondamentaux

Art. 5 Intégrité des processus démocratiques et respect de l'État de droit	Art. 1, par. 1 Le règlement sur l'IA a pour but notamment d'améliorer le fonctionnement du marché intérieur de l'UE en lien avec les produits d'IA, tout en garantissant la protection des droits fondamentaux. Très peu de droits individuels, mais but de protéger les droits fondamentaux en établissant des règles sur la sécurité des produits.
Art. 6 Approche générale	Pas pertinent
Art. 7 Dignité humaine et autonomie individuelle	Art. 5 Pratiques interdites en matière d'IA
Art. 8 Transparence et contrôle	Art. 8 ss Exigences applicables aux systèmes d'IA à haut risque, en particulier cf. art. 11 Documentation technique, art. 12 Enregistrement, art. 13 Transparence et fourniture d'informations aux déployeurs, art. 14 Contrôle humain Art. 26 Obligations incombant aux déployeurs (ex. devoir d'informer) Art. 50 Exigences de transparence Art. 49 et 71 Enregistrement et base de données Art. 86 Droit à l'explication des décisions individuelles
Art. 9 Obligation de rendre des comptes et responsabilité	Catalogue d'obligations pour les différents acteurs dans la chaîne d'approvisionnement des systèmes d'IA (fournisseur, déployeur, etc.) Art. 17, par. 1, point m Cadre de responsabilisation interne Art. 99 ss Sanctions
Art. 10 Egalité et non-discrimination	Art. 10 Données et gouvernance des données, en particulier Art. 10 par. 2, point f Art. 15 Exactitude, robustesse et cybersécurité
Art. 11 Respect de la vie privée et protection des données à caractère personnel	Art. 10 Données et gouvernance des données Art. 15 Exactitude, robustesse et cybersécurité
Art. 12 Fiabilité	Art. 14 Contrôle humain Art. 15 Exactitude, robustesse et cybersécurité

	Art. 40 ss Normes harmonisées Art. 56 ss Codes de bonne pratique Art. 95 Codes de conduite
Art. 13 Innovation sûre	Art. 57 ss Mesures de soutien à l'innovation
Art. 14 Recours	Art. 85 Réclamation Art. 86 Droit à l'explication des décisions individuelles
Art. 15 Garanties procédurales	Art. 26 Obligations incombant aux déployeurs (ex. devoir d'informer) Art. 50 Exigences de transparence Art. 86 Droit à l'explication des décisions individuelles
Art. 16 Cadre de gestion des risques et des impacts	Art. 9 Système de gestion des risques Art. 11 Documentation technique Art. 12 Enregistrement Art. 16 Système de gestion de la qualité Art. 27 Analyse d'impact des systèmes d'IA à haut risque sur les droits fondamentaux Art. 5 Pratiques interdites
Art. 17 Non-discrimination	Pas pertinent, concerne la mise en œuvre de la convention sur l'IA
Art. 18 Droits des personnes handicapées et des enfants	Pas pertinent, concerne la mise en œuvre de la convention sur l'IA
Art. 19 Consultation publique	Art. 67 Forum consultatif Inclusion des parties prenantes dans l'élaboration de normes harmonisées (art. 40 ss), codes de bonne pratique (art. 56), codes de conduite (art. 95), etc.
Art. 20 Maîtrise du numérique et compétences numériques	Art. 4 Maîtrise du numérique Art. 13 Fourniture d'informations aux déployeurs

Art. 21 Sauvegarde des droits de l’homme reconnus	Le règlement sur l’IA précise à plusieurs reprises qu’il s’applique en sus du cadre légal existant et qu’il ne vise en aucun cas à le restreindre. Cela concerne en particulier la protection des données (cf. consid. 10 du règlement).
Art. 22 Protection plus étendue	Pas pertinent
Art. 23 Conférence des Parties	Pas pertinent
Art. 24 Obligation de rapport	Pas pertinent
Art. 25 Coopération internationale	Art. 57, par. 13 Coopération internationale bacs à sable
Art. 26 Mécanismes de contrôle effectifs	Art. 28 ss Autorités de notification et organismes modifiés Art. 40 ss Normes et évaluation de la conformité Art. 64 ss Gouvernance Art. 74 ss Surveillance du marché Art. 99 ss Sanctions
Art. 27 ss Clauses finales	Pas pertinents