Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Justiz 15. Oktober 2025

Office fédéral de la justice 15 octobre 2025

Partizipative Arbeitstagung zur Umsetzung der KI-Konvention des Europarats

Journée de travail participative sur la mise en œuvre de la Convention sur l'intelligence artificielle du Conseil de l'Europe

Material zur Vorbereitung

Matériel de préparation

Liebe Teilnehmende der Arbeitstagung zur Umsetzung der KI-Konvention,

wir freuen uns sehr, am 27. Oktober auf Ihren Beitrag zählen zu können!

Zur Vorbereitung auf Ihre Mitarbeit in den thematischen Breakout-Sessions bitten wir Sie, die Factsheets zu den beiden Schwerpunkten, die Sie bearbeiten werden, zu lesen. Ihre Gruppenzuordnung finden Sie auf der Website für Teilnehmende. Dort stehen Ihnen auch weiterführende Ressourcen zur Verfügung, falls Sie sich vertiefter mit dem Thema der Tagung und der Konvention befassen möchten.

Vielen Dank und bis bald!

Chers participantes et participants à la journée de travail sur la mise en œuvre de la convention sur l'IA,

Nous sommes ravis de pouvoir compter sur votre contribution le 27 octobre !

Afin de vous préparer à participer aux sessions thématiques en petits groupes, nous vous prions de bien vouloir lire les fiches d'information sur les deux thèmes-clés que vous traiterez. Vous trouverez la répartition des groupes sur le site web destiné aux participants. Vous y trouverez également des ressources supplémentaires si vous souhaitez approfondir le thème de la journée de travail et la convention.

Merci beaucoup et à bientôt !

Inhaltsverzeichnis

Table des matières

1. Folgenabschätzung

Um die Grundrechte zu schützen, muss die Schweiz einen Risikomanagement-Rahmen schaffen, der den gesamten Lebenszyklus von KI-Systemen abdeckt. Wie kann eine Folgenabschätzung gestaltet werden, die Risiken frühzeitig erkennt und verhindert?

Worum geht es?

Der breite Einsatz von KI-Systemen birgt sowohl für die Verwaltung als auch für Unternehmen Risiken, die **ganzheitlich** erfasst werden müssen – nicht nur punktuell durch Massnahmen zur Transparenz oder zum Datenschutz. Ein wirksamer **Rahmen für das Risikomanagement** ermöglicht es, solche Risiken frühzeitig zu erkennen, im besten Fall ihr Eintreten zu verhindern und im ungünstigsten Fall ihre Auswirkungen zu mindern. Ein zentrales Instrument dafür ist die **Grundrechte-Folgenabschätzung**.

Was sieht die KI-Konvention des Europarats vor?

Artikel 16 der KI-Konvention verpflichtet die Vertragsstaaten zur Schaffung eines **Rahmens für das Risiko- und Folgenmanagement** für den gesamten Lebenszyklus von KI-Systemen.

- Der Rahmen muss die **Ermittlung, Bewertung, Vermeidung und Minderung** der von KI-Systemen ausgehenden Risiken ermöglichen, unter Berücksichtigung der tatsächlichen und potenziellen Auswirkungen auf die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit.
- Er muss den **ganzen Lebenszyklus des KI-Systems** abdecken und **iterativ** sein, d. h. regelmässig neu beurteilt werden, weil sich bestimmte Risiken erst nach einem breiten Einsatz oder nach einer technischen Weiterentwicklung zeigen;
- Er umfasst **abgestufte und differenzierte Massnahmen**, die dem Kontext, dem vorgesehenen Verwendungszweck sowie der Schwere oder Wahrscheinlichkeit der Auswirkungen angemessen sind
- Die Massnahmen berücksichtigen die Sichtweise der **relevanten beteiligten Parteien** (betroffene Personen, externe Fachleute, Zivilgesellschaft);
- Zu den Massnahmen gehört auch die **Dokumentation** der Risiken, der realen und potenziellen Auswirkungen, und den Ansatz für das Risikomanagement;
- Die Massnahmen erfordern möglicherweise **eine Vorabprüfung** der KI-Systeme vor ihrem erstmaligen Einsatz sowie nach wesentlichen Änderungen.

Eine der zentralen vorgesehenen Verpflichtungen ist die Durchführung einer spezifischen Folgenabschätzung für KI-Systeme. Darüber hinaus hat der Ausschuss für künstliche Intelligenz (CAI) mit **HUDERIA**[1] einen unverbindlichen, aber relativ detaillierten methodischen Leitfaden entwickelt, der ein praktisches Vorgehen zur Umsetzung von Artikel 16 beschreibt und von den einzelnen Staaten frei an ihre Gegebenheiten angepasst werden kann.

Wie stellt sich die aktuelle Situation in der Schweiz dar?

Das schweizerische Recht kennt bereits die Pflicht zur Durchführung einer **Datenschutz-Folgenabschätzung (DSFA)** in bestimmten Situationen. Gemäss Artikel 22 des **Datenschutzgesetzes (DSG)** muss ein privater Verantwortlicher oder ein Bundesorgan eine DSFA durchführen, wenn eine Datenbearbeitung ein hohes **Risiko** für die Persönlichkeit oder die Grundrechte mit sich bringt. Die wichtigsten Merkmale sind:

[1] Risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law.

- Die Risikobewertung obliegt dem Verantwortlichen; das Gesetz nennt lediglich **Beispiele** für Situationen mit hohem Risiko;
- Die DSFA muss die geplante Bearbeitung beschreiben, die Risiken (physische, materielle oder immaterielle) bewerten und **Schutzmassnahmen** vorsehen;
- Es ist keine einheitliche Methode vorgeschrieben, es existieren jedoch **praktische Leitfäden**;
- Das DSG schützt die **Persönlichkeit** und **bestimmte Grundrechte, die eng mit der Bearbeitung von Personendaten verbunden sind** (persönliche Freiheit, informationelle Selbstbestimmung, Privatsphäre), aber andere Rechte, wie etwa der Schutz vor Diskriminierung, sind insbesondere im privaten Bereich nicht direkt erfasst.

Der Verantwortliche konsultiert in der Regel den **Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten (EDÖB),** wenn trotz der vorgesehenen Massnahmen ein hohes Risiko bestehen bleibt (Art. 23 DSG). Der EDÖB teilt dem Verantwortlichen daraufhin seine allfälligen Einwände mit und schlägt geeignete Massnahmen vor. Private Verantwortliche können jedoch **auf die Konsultation** des EDÖB **verzichten**, wenn sie ihre Datenschutzberaterin oder ihren Datenschutzberater beigezogen haben. In der Praxis kommt es daher im Privatsektor nur **sehr selten** zu Konsultationen beim EDÖB.

Selbst wenn eine DSFA nach Artikel 22 DSG erforderlich ist, wird das Unterlassen einer solchen **nicht unmittelbar sanktioniert**. Der EDÖB kann jedoch jederzeit die Durchführung einer DSFA verlangen, sobald er von einem Fall Kenntnis erhält, und seine Anordnung mit der Androhung einer strafrechtlichen Sanktion verbinden, mit einer Busse von bis zu 250 000 Franken. Zudem kann er die Aussetzung oder Einstellung der Bearbeitung anordnen.

Eine DSFA reicht allein nicht aus, um die Anforderungen der Konvention zu erfüllen. Sie deckt nicht alle Situationen ab, in denen durch ein KI-System Risiken entstehen können, etwa wenn das Training oder die Nutzung des Systems **keine Bearbeitung von Personendaten erfordert**. Zudem umfasst sie, wie oben skizziert, nicht unbedingt **den gesamten Katalog** an Grundrechten, der gemäss KI-Konvention relevant ist.

Welche Herausforderungen stellen sich?

Die Pflicht, vor dem Einsatz eines KI-Systems sowie gegebenenfalls bei wesentlichen Änderungen eine Folgenabschätzung durchzuführen, bildet das **Kernstück des Risikomanagements** im Sinne von Artikel 16 der KI-Konvention. Dabei ergeben sich jedoch mehrere Herausforderungen:

- Gestaltung einer Methodik, die präzise genug ist, um ihrem präventiven Zweck gerecht zu werden
- Vermeidung eines zu schwerfälligen Rahmens, der die praktische Anwendung behindert
- Koordinierung dieses neuen Instrumentes mit der bestehende DSFA zur Vermeidung von Doppelspurigkeit.

Weitere Informationen: **Rechtliche Basisanalyse des BJ** (Deutsch / Französisch): Kapitel 4.3.4 **Erläuternder Bericht zur KI-Konvention** (Englisch / Französisch): N 105 – 112	 **HUDERIA:** Allgemeine Informationen HUDERIA-Methode (Englisch / Französisch) **Merkblatt zur DSFA des EDÖB** (Deutsch / Französisch)

1. Analyse d'impact

Pour protéger les droits fondamentaux, la Suisse doit se doter d'un cadre de gestion des risques couvrant tout le cycle de vie des systèmes d'IA. Quelle forme donner à une analyse d'impact capable d'identifier et prévenir ces risques ?

De quoi s'agit-il ?

L'utilisation de systèmes d'intelligence artificielle (IA) à grande échelle – dans l'administration comme dans les entreprises – comporte des risques qui doivent être appréhendés dans leur **globalité**, pas seulement par des mesures ponctuelles de transparence ou de protection des données. Un **cadre de gestion des risques** efficace permet de les anticiper et, au mieux, prévenir leur survenance, au pire, en réduire leurs impacts, notamment par le biais d'une **analyse d'impact sur les droits fondamentaux (AIDF)**.

Que prévoit la convention du Conseil de l'Europe ?

L'article 16 de la Convention sur l'intelligence artificielle impose aux États parties d'instaurer un **cadre de gestion des risques et des impacts** tout au long du cycle de vie des systèmes d'IA.

- Proposer une méthodologie suffisamment précise pour atteindre son objectif de prévention ;
- Le cadre doit permettre **d'identifier, évaluer, prévenir et atténuer** les 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 ;
- Il doit couvrir **l'entier du cycle de vie** du système d'IA et être **itératif**, c'est-à-dire réévalué de façon régulière, car certains risques n'apparaissent qu'après un déploiement à large échelle ou une évolution technologique ;
- Il comprend des mesures **graduées et différenciées**, proportionnées au contexte, à l'utilisation prévue et à la gravité ou probabilité des impacts ;
- Les mesures prennent en compte le point de vue des **parties prenantes** pertinentes (personnes concernées, experts externes, société civile) ;
- Les mesures comprennent la **documentation** des risques, des impacts réels et potentiels, et de l'approche de la gestion des risques ;
- Le cas échéant, les mesures exigent **l'essai préalable** des systèmes d'IA avant leur première utilisation et lorsqu'ils subissent des modifications significatives.

L'obligation centrale envisagée est la mise en place d'une **analyse d'impact dédiée aux systèmes d'IA**. Le Comité sur l'intelligence artificielle (CAI) a par ailleurs développé **HUDERIA**[2], un guide méthodologique non contraignant et relativement détaillé qui propose une démarche pratique pour appliquer l'article 16 et peut être adapté librement par chaque pays.

Quelle est la situation actuelle en Suisse ?

Le droit suisse connaît déjà une obligation d'effectuer une **analyse d'impact relative à la protection des données personnelles (AIPD)** dans certaines situations. Selon l'article 22 de la **loi sur la protection des données (LPD)**, un responsable de traitement privé ou un organe fédéral doit effectuer une AIPD lorsqu'un traitement de données présente un **risque**

[2] Risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law.

élevé pour la personnalité ou les droits fondamentaux. Ses principales caractéristiques sont les suivantes :

- L'évaluation du risque relève du responsable de traitement, la loi donnant seulement des **exemples** de situations à risque élevé ;
- L'AIPD doit décrire le traitement envisagé, évaluer les risques (physiques, matériels ou immatériels) et prévoir les **mesures de protection** ;
- Aucune méthode unique n'est imposée, mais des **guides pratiques** existent ;
- La LPD vise à protéger la **personnalité** et **certains droits fondamentaux en lien étroit avec le traitement de données personnelles** (liberté personnelle, autodétermination informationnelle, sphère privée), mais d'autres droits, comme la protection contre les discriminations, ne sont pas directement couverts, en particulier dans le secteur privé.

Le responsable consulte en principe le **préposé fédéral à la protection des données (PFPDT)** si un risque élevé subsiste malgré les mesures prévues (art. 23 LPD). Le PFPDT doit ensuite communiquer ses éventuelles objections au responsable de traitement, en proposant des mesures appropriées. Cependant, le responsable du traitement privé peut **renoncer à consulter** le PFPDT s'il a consulté sa conseillère ou son conseiller à la protection des données. Les consultations du PFPDT sont donc **très rares** en pratique pour le secteur privé.

Même lorsqu'elle elle est requise par l'article 22 LPD, l'absence de réalisation d'une AIPD ne fait pas directement l'objet **d'une sanction**. En revanche, le PFPDT peut toutefois en tout temps exiger la réalisation d'une telle AIPD s'il a connaissance du cas et peut assortir sa décision de la menace d'une sanction pénale, avec une amende pouvant aller jusqu'à 250 000 francs. Il peut également exiger la suspension ou cessation du traitement.

L'AIPD ne suffit pas en tant que telle à remplir les exigences de la convention. En effet, elle ne couvre pas toutes les situations présentant des risques en lien avec un système d'IA, par exemple lorsque l'entraînement ou l'usage du système n'implique **pas de traitement de données personnelles**. En outre, comme esquissé ci-dessus, elle ne couvre pas nécessairement **tout le catalogue** des droits fondamentaux pertinents aux yeux de la convention.

Quels sont les enjeux ?

L'obligation d'établir une analyse d'impact en amont du déploiement d'un système d'IA, et éventuellement lors de changements majeurs, constitue **le cœur du cadre de gestion des risques** au sens de l'article 16 de la convention. Cependant, elle soulève plusieurs enjeux :

- Proposer une méthodologie suffisamment précise pour atteindre son objectif de prévention ;
- Éviter un cadre trop lourd qui découragerait son application ;
- Coordonner ce nouvel instrument avec l'AIPD existante pour prévenir les doublons.

Informations supplémentaires : **Analyse jurisdique de base de l'OFJ** (Allemand / Français) : chapitre 4.3.4 **Rapport explicatif de la convention** (Anglais / Français) : paragraphes 105-112	**HUDERIA:** Allgemeine Informationen Méthodologie HUDERIA (Anglais / Français) **Aide-mémoire sur l'AIPD du PFPDT (Allemand / Français)**

2. Rote Linien

Bestimmte Anwendungen von KI können mit Grundrechten unvereinbar sein. Sollten Verbote oder Moratorien vorgesehen werden, um die Gesellschaft zu schützen?

Worum geht es?

Der Einsatz bestimmter Systeme künstlicher Intelligenz kann völlig unvereinbar mit den Grundrechten, dem Funktionieren der Demokratie und dem Rechtsstaat sein. In solchen Fällen kann der Gesetzgeber **auf gesetzlicher Ebene rote Linien ziehen**, indem er bestimmte Systeme verbietet oder Strafnormen einführt, die bestimmte Verwendungen unter Strafe stellen. Solche roten Linien können auch durch die **Rechtsprechung** anhand von Einzelfällen entwickelt werden.

Es stellt sich die Frage, ob der Schweizer Gesetzgeber in der künftigen Regulierung zur KI rote Linien vorsehen soll und gegebenenfalls in welcher Form.

Was sieht die KI-Konvention des Europarats vor?

Die Konvention enthält keine unmittelbar anwendbaren Bestimmungen, die den Einsatz bestimmter KI-Systeme verbieten oder unter Strafe stellen. Allerdings verweisen mindestens zwei Bestimmungen auf die Thematik einer mit Grundrechten, Demokratie und Rechtsstaat unvereinbaren Nutzung:

- Gemäss Artikel 7 «Menschenwürde und individuelle Autonomie» trifft jeder Staat Massnahmen zur **Wahrung der Menschenwürde und der individuellen Autonomie** in Bezug auf Tätigkeiten innerhalb des gesamten Lebenszyklus von KI-Systemen. Diese Bestimmung unterstreicht die Bedeutung der Menschenwürde und der individuellen Autonomie im Rahmen einer menschenzentrierten Regulierung. So dürfen Tätigkeiten im Lebenszyklus von KI-Systemen nicht zur **Entmenschlichung** des Einzelnen beitragen.
- Artikel 16 Absatz 4 («Rahmen für das Risiko- und Folgemanagement») verpflichtet die Staaten, die Notwendigkeit eines **Moratoriums**, eines **Verbots** oder **anderer geeigneter Massnahmen** zu prüfen, wenn sie bestimmte Anwendungen von KI-Systemen als **unvereinbar** mit den Menschenrechten, der Funktionsweise der Demokratie oder dem Rechtsstaat ansehen. Laut dem erläuternden Bericht zur Konvention liegt es somit in der Verantwortung der Staaten, festzulegen, was als «unvereinbar» gilt, ebenso wie die Beurteilung der Notwendigkeit eines Moratoriums, eines Verbots oder anderer Massnahmen.

Wie stellt sich die aktuelle Situation in der Schweiz dar?

- **Verbote:**
  - **Im öffentlichen Sektor:** Artikel 36 der Bundesverfassung enthält die Voraussetzungen für die Einschränkung von Grundrechten. Ein Eingriff in Grundrechte durch ein KI-System, das auf keiner gesetzlichen Grundlage beruht oder den Kerngehalt des betroffenen Rechts verletzt, ist unzulässig und somit bereits *per se* **durch die Verfassung verboten**. Daher ergäbe im öffentlichen Sektor ein Verbot oder Moratorium nur dann Sinn, wenn eine **politische Entscheidung** getroffen würde, **bestimmte Praktiken ausdrücklich zu untersagen**. Zu beachten ist, dass der Gesetzgeber ein solches Verbot jederzeit wieder aufheben oder ändern kann, sei es generell oder gezielt. Die Tragweite eines solchen Verbots ist also im Wesentlichen symbolisch – ausser es wird in die Verfassung aufgenommen. Denkbar wäre hingegen, dass das Gesetz bestimmte Verwendungen von KI davon abhängig macht, dass eine vom

Parlament beschlossene gesetzliche Grundlage besteht (vgl. z. B. Art. 34 des Bundesgesetzes über den Datenschutz).
    - **Im privaten Sektor**: Hier gilt Artikel 36 der Bundesverfassung nicht. Der Gesetzgeber müsste daher **explizite Normen** erlassen, um bestimmte Nutzungen zu verbieten.
- **Strafrecht**: Das Strafgesetzbuch ist grundsätzlich **technologieneutral** und gilt unabhängig davon, welches Instrument der Täter verwendet – also auch für KI. Artikel 179^{decies} (*«Identitätsmissbrauch»*), der die unbefugte Nutzung der Identität einer anderen Person zum Zweck der Schädigung oder zur Erlangung eines unrechtmässigen Vorteils unter Strafe stellt, ist ein Beispiel für ein Delikt, das typischerweise auch durch den Einsatz von KI (z. B. Deepfakes) begangen werden kann.

Welche Herausforderungen gibt es?

Die Konvention verbietet den Einsatz bestimmter KI-Systeme nicht direkt. Sie verpflichtet die Staaten jedoch, die Notwendigkeit von Verboten, Moratorien oder anderen geeigneten Massnahmen zu prüfen, wenn bestimmte Nutzungen mit den Grundrechten, der Demokratie und die Rechtsstaatlichkeit unvereinbar sind. Die Schweiz muss daher erwägen, ob in der künftigen KI-Regulierung **rote Linien eingeführt werden sollen**. Diese Frage ist primär politischer Natur und wirft verschiedene Herausforderungen auf, insbesondere:

- Auswahl an Praktiken, die gegebenenfalls verboten werden sollen (z. B. Emotionserkennung, soziale Bewertung)
- gegebenenfalls Art und Weise der Definition solcher Verbote, wobei rechtliche Klarheit gewährleistet sein muss
- gegebenenfalls mögliche Formulierung von Ausnahmen
- Anwendungsbereich im öffentlichen und privaten Sektor

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.3.2.2 et 4.3.4.	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 53 – 55 und 111 –112

2. Lignes rouges

Certaines applications de l'IA peuvent être incompatibles avec les droits fondamentaux. Faut-il prévoir des interdictions ou des moratoires pour protéger la société ?

De quoi s'agit-il ?

Certains systèmes d'intelligence artificielle peuvent être totalement incompatibles avec les droits fondamentaux, le fonctionnement de la démocratie et l'État de droit. Dans ce cas, le législateur peut introduire des **lignes rouges au niveau législatif**, en prévoyant notamment l'interdiction de certains systèmes ou l'introduction de dispositions pénales criminalisant certaines utilisations. De telles lignes rouges peuvent aussi être développées par la **jurisprudence** lorsque des cas individuels sont jugés par les tribunaux.

La question se pose de savoir si le législateur suisse doit prévoir des lignes rouges dans la future réglementation sur l'IA, et, le cas échéant, sous quelle forme.

Que prévoit la convention du Conseil de l'Europe ?

La convention ne prévoit pas de dispositions directement applicables interdisant ou criminalisant le recours à tel ou tel système d'IA. Cela étant, au moins deux dispositions renvoient à la thématique des usages incompatibles avec les droits fondamentaux, la démocratie et l'État de droit :

- Selon l'art. 7 « Dignité humaine et autonomie personnelle », chaque État adopte ou maintient des mesures en faveur du 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.
- L'art. 16, par. 4 (« Cadre de gestion des risques et des impacts »), prévoit que chaque État évalue la nécessité d'un **moratoire**, d'une **interdiction** ou d'**autres mesures appropriées** concernant certaines utilisations de systèmes d'IA lorsqu'il considère que ces utilisations sont **incompatibles** avec le respect des droits de l'homme, le fonctionnement de la démocratie ou l'État de droit. Selon le rapport explicatif de la convention, la détermination de ce qui est « incompatible » relève donc des États, de même que l'évaluation de la nécessité d'un moratoire, d'une interdiction, ou d'autres mesures appropriées.

Quelle est la situation actuelle en Suisse ?

- La sélection des pratiques à interdire (p. ex. reconnaissance des émotions, notation sociale), le cas échéant ;
- La manière de définir ces interdictions, le cas échéant, en garantissant la clarté juridique ;
- **Interdictions** :
    - **Dans le secteur public** : En droit suisse, l'art. 36 de la Constitution fédérale pose des conditions à la restriction des droits fondamentaux. Une atteinte aux droits fondamentaux provenant d'un système d'IA qui ne repose pas sur une base légale, ou qui touche l'essence même du droit en cause, est injustifiée et donc *per se* **déjà interdite par la Constitution**. Dès lors, dans le secteur public, une interdiction ou un moratoire ne ferait du sens qu'en cas de **décision politique d'intervenir activement** pour interdire explicitement certaines pratiques. À noter que si le législateur

suisse interdit une certaine pratique, rien ne l'empêche ultérieurement de revenir sur ce choix, soit de manière générale, soit de manière ciblée. La portée de l'interdiction est donc essentiellement symbolique, sauf si elle figure dans la Constitution. En revanche, on pourrait imaginer que la loi conditionne certains usages de l'IA à l'adoption d'une base légale adoptée par le Parlement (cf. p. ex. l'art. 34 de la loi fédérale sur la protection des données).
    - **Dans le secteur privé** : La situation est différente dans le **secteur privé**, où l'art. 36 de la Constitution ne s'applique pas. Le législateur devrait donc édicter des **normes explicites** pour interdire certaines utilisations.
- **Droit pénal** : le Code pénal est en principe **technologiquement neutre** et s'applique quel que soit l'instrument utilisé par l'auteur, y compris l'IA. L'article 179^{decies} (« Usurpation d'identité »), qui punit quiconque utilise l'identité d'une autre personne sans son consentement dans le dessin de lui nuire ou de se procurer ou de procurer à un tiers un avantage illicite, constitue un exemple d'infraction pouvant typiquement être réalisée au moyen d'un système d'IA (p. ex. au moyen de *deepfakes*).

Quels sont les enjeux ?

La convention n'interdit pas directement le recours à certains systèmes d'IA. Elle demande toutefois aux États d'évaluer la nécessité de prévoir des interdictions, moratoires ou autres mesures appropriées en cas d'incompatibilité de certaines utilisations avec les droits fondamentaux, la démocratie et l'État de droit. La Suisse doit donc réfléchir à **l'introduction d'éventuelles lignes rouges** dans la réglementation en matière d'IA. Cette question, principalement d'ordre politique, soulève plusieurs enjeux, à savoir notamment :

- La sélection des pratiques à interdire (p. ex. reconnaissance des émotions, notation sociale), le cas échéant ;
- La manière de définir ces interdictions, le cas échéant, en garantissant la clarté juridique ;
- Comment formuler d'éventuelles exceptions, le cas échéant ;
- Le champ d'application dans le secteur public et privé ;

Informations supplémentaires :	
Analyse jurisdique de base de l'OFJ (Allemand / Français) : chapitre 4.3.2.2 et 4.3.4.	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 53 - 55 et 111-112

3. Demokratie

Die Schweiz verfügt über wirksame Instrumente zum Schutz der Demokratie im KI-Zeitalter. Ist eine Neubewertung angesichts neuer KI-Anwendungen angezeigt?

Worum geht es?

Künstliche Intelligenz kann die Beteiligung der Bürgerinnen und Bürger und das Funktionieren der Demokratie unterstützen, sie kann die Institutionen und den Rechtsstaat aber auch schwächen. KI-Systeme können insbesondere:

- **Desinformation** verstärken und die Grenze zwischen wahr und falsch verwischen (Deepfakes, Chatbots, automatisiert agierende Bots);
- **Politisches Targeting** erleichtern, indem sie Personendaten zur Beeinflussung von Meinungen und Wahlverhalten nutzen;
- **Intrusive Überwachung** ermöglichen (Gesichtserkennung, Massenanalysen), welche sich auf Meinungs- und Versammlungsfreiheit auswirkt.

Diese Risiken bedrohen die freie Meinungsbildung, den gleichberechtigten Zugang zur öffentlichen Debatte und die Unabhängigkeit der Institutionen. Sie können sowohl bei **Abstimmungen** als auch **Wahlen** auftreten (z. B. Verbreitung falscher Informationen, gezieltes Profiling) als auch ausserhalb (z. B. Überwachung). Zum Schutz der Demokratie angesichts der rasanten Entwicklungen im KI-Bereich ist ein Rahmen für Prävention und Reaktion unerlässlich.

Was sieht die KI-Konvention des Europarats vor?

Artikel 5 der KI-Konvention verpflichtet die Staaten, Massnahmen zu ergreifen oder beizubehalten, die sicherstellen, dass KI-Systeme nicht dazu eingesetzt werden, die Integrität, **Unabhängigkeit** und **Wirksamkeit** der **demokratischen Institutionen und Prozesse** zu beeinträchtigen – einschliesslich des Prinzips der Gewaltenteilung, der Unabhängigkeit der Justiz sowie des Zugangs zur Justiz. Dieser Rahmen:

- deckt den gesamten **Lebenszyklus** von KI-Systemen ab, einschliesslich des gleichberechtigten Zugangs zur öffentlichen Debatte und die Möglichkeit der freien Meinungsbildung
- lässt den Staaten einen **breiten Ermessensspielraum** bei der Wahl der konkreten Massnahmen, um sich an aktuelle und künftige Risiken anpassen zu können.

Die Bestimmung ist bewusst offen formuliert, sodass auch noch unbekannte Bedrohungen berücksichtigt werden können. Sie ergänzt weitere Artikel der Konvention, insbesondere das **Transparenzprinzip** (Art. 8) und den **Rahmen für das Risikomanagement** (Art. 16).

Wie stellt sich die aktuelle Situation in der Schweiz dar?

Das schweizerische Recht enthält bereits eine Reihe von Garantien, die Demokratie und Rechtsstaat vor Risiken durch KI-Systeme schützen:

- Artikel 34 der Bundesverfassung garantiert die freie Willensbildung und die unverfälschte Stimmabgabe. Gestützt auf das **Gesetz über die politischen Rechte** sowie die Rechtsprechung des Bundesgerichts kann das Ergebnis einer Abstimmung aufgehoben werden, wenn die freie Willensbildung nicht zuverlässig und die Stimmabgabe nicht unverfälscht gewährleistet war. Eine amtliche Richtigstellung kann angeordnet werden, wenn offensichtlich falsche Informationen das Ergebnis erheblich beeinflussen könnten.

- Das **Datenschutzgesetz (DSG)** regelt die Bearbeitung besonders schützenswerter Daten, darunter auch politische Meinungen. Automatisiertes Profiling oder Profiling mit hohem Risiko (Art. 5 DSG) verlangt erhöhte Transparenz und kann eine **Datenschutz-Folgenabschätzung** erforderlich machen (Art. 22 DSG). Diese Regeln betreffen insbesondere gezielte politische Werbung und den Einsatz von Algorithmen zur Beeinflussung von Wählergruppen. Zudem ist für die Bearbeitung besonders schützenswerter Daten durch Bundesorgane eine gesetzliche Grundlage in Form eines Bundesgesetzes erforderlich (Art. 34 Abs. 2 DSG).
- Der Schutz der Demokratie und des Rechtsstaats geht auch mit dem Schutz der Grundrechte einher. Die Bundesverfassung (Art. 16, 17, 22) und die Europäische Menschenrechtskonvention (EMRK) schützen **die Meinungs- und Informationsfreiheit, die Medienfreiheit sowie die Versammlungsfreiheit.** Die Rechtsprechung des Europäischen Gerichtshofs für Menschenrechte betont die Grenzen von Massenüberwachung, insbesondere durch Gesichtserkennung.[3]
- Das **Strafgesetzbuch** ahndet Diskriminierung und Aufruf zu Hass (Art. 261bis), Identitätsmissbrauch (Art. 179^{decies}) und Vergehen gegen den Volkswillen (Art. 279 ff.). Im **Zivilrecht** erstreckt sich der Persönlichkeitsschutz (Art. 28 ff. ZGB) auch auf Deepfakes oder die Verbreitung manipulierten Bildmaterials.
- Bis heute ist **politische Werbung** in Radio und Fernsehen verboten, für Online-Werbung hingegen gibt es keine Regulierung.

Der Bundesrat hat zudem das UVEK (BAKOM) beauftragt, einen Vorentwurf für ein Gesetz zur Regulierung **grosser Online-Plattformen** auszuarbeiten. Vorgesehen sind verschiedene Pflichten wie erhöhte Transparenz, Meldeverfahren für Hassrede, eine Anlaufstelle in der Schweiz und Rechtsmittel für Nutzer und Nutzerinnen. Dieses Gesetzgebungsprojekt gehört jedoch **nicht** zu den Arbeiten des BJ im Bereich KI.

Insgesamt deckt der hier dargestellte Rechtsrahmen die Anforderungen der KI-Konvention weitgehend ab. Die hohe Geschwindigkeit der technologischen Entwicklungen erfordert jedoch ständige Wachsamkeit und Beobachtung der Rechtsprechung.

Welche Herausforderungen stellen sich?

Gemäss der rechtlichen Basisanalyse des BJ wird die Integrität demokratischer Prozesse und die Einhaltung der Rechtsstaatlichkeit im schweizerischen Recht durch eine breite Palette von Normen sichergestellt. Entsprechend kommt die Analyse zum Schluss, dass keine unmittelbaren Gesetzesänderungen erforderlich sind, jedoch mehrere Herausforderungen bestehen bleiben:

- die schnelle Entwicklung von Techniken der Desinformation, des politischen Profilings und der algorithmischen Manipulation
- die Transparenz bei KI-generierten Inhalten (Kennzeichnung, Identifizierung), zur Gewährleistung der freien Meinungsbildung
- die Koordination mit anderen Bestimmungen der Konvention (Transparenz, Risikomanagement) zur Ergänzung des bestehenden Schutzes.

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.3.1.2	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 42 – 48

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

3. Démocratie

La Suisse dispose déjà d'instruments solides pour préserver la démocratie à l'ère de l'IA. Une réévaluation s'impose-t-elle face à de nouvelles formes d'utilisation de l'IA ?

De quoi s'agit-il ?

L'intelligence artificielle peut soutenir la participation citoyenne et le fonctionnement de la démocratie, mais elle peut aussi fragiliser les institutions et l'État de droit. Les systèmes d'IA peuvent en particulier :

- Amplifier la **désinformation** et brouiller la frontière entre vrai et faux (deepfakes, chatbots, bots automatisés) ;
- Faciliter le **ciblage politique** en utilisant des données personnelles pour influencer l'opinion ou le comportement électoral ;
- Permettre une **surveillance intrusive** (reconnaissance faciale, analyses de masse) qui affecte la liberté d'expression et de réunion.

Ces risques menacent la libre formation de l'opinion, l'égalité d'accès au débat public et l'indépendance des institutions. Ils peuvent apparaître tant dans le contexte de **votations ou élections** (p. ex. diffusion de fausses informations, profilage ciblé), qu'en dehors (p. ex. surveillance). Un cadre de prévention et de réaction s'avère donc essentiel pour protéger la démocratie face aux évolutions rapides de l'IA.

Que prévoit la convention du Conseil de l'Europe ?

L'article 5 de la convention impose aux États d'adopter ou maintenir des 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**, y compris au principe de la séparation des pouvoirs, au respect de l'indépendance de la justice et à l'accès à la justice. Ce cadre :

- Couvre l'ensemble du **cycle de vie** des systèmes d'IA, en incluant l'accès équitable au débat public et la capacité des personnes à se forger librement une opinion ;
- Laisse aux États une **large marge d'appréciation** quant aux mesures concrètes, afin de s'adapter aux risques actuels et futurs.

La disposition est volontairement ouverte, ce qui permet d'intégrer des menaces encore inconnues. Elle complète d'autres articles de la convention, notamment le principe de **transparence** (art. 8) et le **cadre de gestion des risques** (art. 16).

Quelle est la situation actuelle en Suisse ?

Le droit suisse comporte déjà un ensemble de garanties protégeant la démocratie et l'État de droit face aux risques posés par les systèmes d'IA :

- L'article 34 de la Constitution fédérale garantit la libre formation de la volonté et l'expression non faussée du vote. La **loi sur les droits politiques** et la jurisprudence du Tribunal fédéral permettent d'annuler un scrutin si la volonté libre n'a pas été exprimée de manière fiable et non faussée. Une rectification officielle peut être imposée lorsque des informations manifestement fausses risquent d'influencer gravement le résultat.
- La **loi sur la protection des données (LPD)** encadre le traitement de données sensibles, dont les opinions politiques. Le profilage automatisé ou le profilage à risque élevé (art. 5 LPD) exige une transparence accrue et peut nécessiter une **analyse d'impact**

(art. 22 LPD). Ces règles couvrent notamment la publicité politique ciblée et l'utilisation d'algorithmes pour influencer des groupes d'électeurs. En outre, les organes fédéraux ne peuvent traiter des données sensibles qu'en présence d'une base légale dans une loi fédérale (art. 34, al. 2 LPD).

- La protection de la démocratie et de l'État de droit va aussi de pair avec la protection des droits fondamentaux. La Constitution fédérale (art. 16, 17, 22) et la Convention européenne des droits de l'homme (CEDH) protègent la **liberté d'opinion et d'information**, la **liberté des médias** et la **liberté de réunion**. La jurisprudence de la Cour européenne des droits de l'homme souligne notamment les limites de la surveillance de masse, en particulier par reconnaissance faciale[4].
- Le code **pénal** sanctionne la discrimination et l'incitation à la haine (art. 261bis), l'usurpation d'identité (art. 179^{decies}) et les infractions contre la volonté populaire (art. 279 ss). En droit **civil**, la protection de la personnalité (art. 28 ss du code civil) peut s'appliquer aux deepfakes ou à la diffusion d'images manipulées.
- À ce jour, la **publicité politique** est interdite à la radio et à la télévision, mais aucune réglementation n'encadre la publicité politique en ligne.

Le Conseil fédéral a par ailleurs chargé le DETEC (OFCOM) de rédiger un avant-projet législatif visant à réguler les **grandes plateformes en ligne** en leur imposant différentes obligations, par exemple : transparence accrue, procédures de signalement des discours haineux, point de contact en Suisse et voies de recours pour les utilisateurs. Ce projet législatif ne fait **pas** partie des travaux de l'OFJ relatifs à l'IA.

Dans l'ensemble, le dispositif juridique esquissé ci-avant couvre largement les exigences de la convention. Toutefois, la rapidité des évolutions technologiques nécessite une vigilance permanente et un suivi de la jurisprudence.

Quels sont les enjeux ?

Selon l'analyse juridique de base de l'OFJ, l'intégrité des processus démocratiques et le respect de l'État de droit, sont appréhendés en droit suisse par un large éventail de normes. Dès lors, l'analyse conclut qu'aucune modification législative immédiate n'est nécessaire, mais que plusieurs défis demeurent :

- L'évolution rapide des techniques de désinformation, de profilage politique et de manipulation algorithmique ;
- La transparence sur les contenus générés par l'IA (marquage, identification) afin de préserver la libre formation de l'opinion ;
- Une coordination avec d'autres dispositions de la convention (transparence, gestion des risques) pour compléter la protection existante.

Informations supplémentaires : **Analyse jurisdique de base de l'OFJ** (Allemand / Français) : chapitre 4.3.1.2	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 42-48

[4] CourEDH, *Glukhin c. Russie*, 11519/20 (4 juillet 2023)

4. Gleichstellung und Nichtdiskriminierung

KI-Systeme können diskriminierende Entscheidungen treffen und bestehende Ungleichheiten in unserer Gesellschaft in grossem Umfang reproduzieren. Welche Ergänzungen sind im Schweizer Recht erforderlich, um algorithmische Diskriminierungen zu bekämpfen?

Worum geht es?

Der Einsatz künstlicher Intelligenz stellt eine **grosse Herausforderung** im Hinblick auf Gleichstellung und das Diskriminierungsverbot dar. **KI ist nämlich nicht neutral**: Sie **reproduziert** die in unserer Gesellschaft bereits bestehenden **Ungleichheiten** und kann diese sogar verstärken. Hinzu kommt, dass KI-Systeme zunehmend von Behörden und privaten Akteuren eingesetzt werden – etwa im Rahmen von Rekrutierungsverfahren oder bei der Gewährung von Leistungen. Dadurch besteht die Gefahr, dass Diskriminierungen nicht nur verfestigt werden, sondern eine **systemische Dimension** annehmen. **Diskriminierende Verzerrungen** können insbesondere aus den verwendeten Daten resultieren (z. B. mangelnde Repräsentativität, veraltete Daten), aus dem Algorithmus selbst (z. B. Wahl der im Modell berücksichtigten oder nicht berücksichtigten Variablen) oder seines Einsatzmodus. Im Falle eines diskriminierenden Ergebnisses erschwert der sogenannte «Black-Box-Effekt» die Nachvollziehbarkeit der zugrunde liegenden Kriterien erheblich – und damit auch den **Nachweis** einer Diskriminierung.

Was sieht die KI-Konvention des Europarats vor?

Artikel 10 Absatz 1 der Konvention verpflichtet jeden Staat, Massnahmen zu ergreifen oder aufrechtzuerhalten, welche die Achtung der Gleichheit, einschliesslich der Gleichstellung der Geschlechter, sowie das Diskriminierungsverbot gemäss dem anwendbaren Völker- und Landesrecht gewährleisten. Absatz 2 besagt, dass jeder Staat sich verpflichtet, in Übereinstimmung mit seinen nationalen und internationalen Verpflichtungen im Bereich der Menschenrechte bei Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen Massnahmen zu ergreifen oder aufrechtzuerhalten, die darauf abzielen, Ungleichheiten zu beseitigen, um unparteiische, gerechte und faire Ergebnisse zu erzielen.

Es wird dabei darauf verwiesen, dass die Konvention **keine neuen Grundrechte schafft** und somit die Tragweite des Rechts auf Gleichbehandlung und Nichtdiskriminierung, wie sie im geltenden schweizerischen Recht besteht, nicht verändert. Die Konvention verlangt jedoch von den Staaten, im spezifischen Kontext von KI-Systemen **einen ebenso wirksamen und gültigen Diskriminierungsschutz** zu gewährleisten wie ausserhalb des Bereichs von KI.

Wie stellt sich die aktuelle Situation in der Schweiz dar?

- Auf Ebene der **Bundesverfassung (BV)** ist die Rechtsgleichheit in Artikel 8 Absatz 1 BV und das Diskriminierungsverbot in Artikel 8 Absatz 2 BV verankert. Artikel 8 Absatz 3 BV garantiert die rechtliche und tatsächliche Gleichstellung von Mann und Frau. Artikel 8 Absatz 4 BV sieht Massnahmen zur Beseitigung von Benachteiligungen von Behinderten vor. Mit Ausnahme der Lohngleichheit zwischen Mann und Frau (Art. 8 Abs. 3 BV) gelten das Gebot der Rechtsgleichheit und das Diskriminierungsverbot gegenüber dem Staat und binden Private nicht unmittelbar.
- Gemäss Artikel 35 BV ist der Staat verpflichtet, die Grundrechte in der gesamten Rechtsordnung, auch im Verhältnis zwischen Privaten, zur Geltung zu bringen, soweit sich diese dafür eignen. Der schweizerische Gesetzgeber hat das Gleichbehandlungsgebot

und das Diskriminierungsverbot im Verhältnis unter Privaten **nach Diskriminierungsarten oder in spezifischen Bereichen** konkretisiert. **Ein allgemeines Diskriminierungsverbot unter Privaten besteht jedoch nicht**.

- Im **Privatrecht** bestehen gewisse Normen, die indirekt Schutz vor Diskriminierung bieten. So können Betroffene je nach Fall gestützt auf **Artikel 28 ZGB und Artikel 328 OR** die Unterlassung einer Diskriminierung oder eine Entschädigung verlangen. Das **Bundesgesetz über die Gleichstellung von Frau und Mann (GIG)** konkretisiert den Schutz vor Diskriminierungen im Arbeitsverhältnis, auch im privaten Sektor. Zudem gilt das **Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen (BehiG)** für algorithmische Diskriminierungen, die insbesondere in Arbeitsverhältnissen nach dem Bundespersonalgesetz, in der Aus- und Weiterbildung sowie bei bestimmten öffentlich zugänglichen Leistungen auftreten können.[5]
- Schliesslich enthält auch das **Bundesgesetz über den Datenschutz**, obwohl es keinen direkten Diskriminierungsschutz bietet, relevante Schutzmechanismen: verstärkter Schutz besonders schützenswerter Daten, Erfordernis einer gesetzlichen Grundlage für Bundesorgane, Grundsätze der Verhältnismässigkeit, Richtigkeit, Zweckbindung und von Treu und Glauben, Datenschutz durch Technik und datenschutzfreundliche Voreinstellungen, Durchführung von Datenschutz-Folgenabschätzungen bei hohem Risiko, Informationspflicht und Auskunftsrecht – insbesondere bei ausschliesslich automatisierten Entscheidungen – sowie Pflicht zur Dokumentation und Protokollierung.

Trotz dieses Rahmens wird in der Öffentlichkeit teilweise die Ansicht vertreten, dass das schweizerische Recht im Bereich Rechtsgleichheit und Diskriminierungsverbot **Lücken** aufweist – unabhängig davon, ob KI zum Einsatz kommt oder nicht. Es gibt insbesondere Stimmen, die sich für die Schaffung eines allgemeinen Gesetzes zur Gleichbehandlung und Nichtdiskriminierung aussprechen.

Welche Herausforderungen stellen sich?

Der Bundesrat hat anerkannt, dass beim Einsatz von KI-Systemen **erhöhte Diskriminierungsrisiken** bestehen, und zwar sowohl im öffentlichen wie auch im privaten Sektor. Er hat angekündigt, diesem Thema im Rahmen seiner gesetzgeberischen Arbeiten **besondere Aufmerksamkeit** zu widmen. Falls politischer Handlungswille besteht, ist offensichtlich, dass eine generelle Stärkung der Antidiskriminierungsgesetzgebung auch im Kontext von KI von Vorteil wäre.

Abgesehen von dieser Frage besteht die zentrale Herausforderung weiterhin darin, mögliche **gezielte Interventionen** des Gesetzgebers zur Verbesserung des bestehenden rechtlichen Rahmens zu bestimmen. Zu den möglichen Massnahmen zählen insbesondere eine stärkere **Transparenzpflicht** für KI-Systeme sowie die Verpflichtung, **Risikoanalysen** im Hinblick auf Diskriminierungen durchzuführen.

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.3.2.5	**Erläuternder Bericht zur Konvention** (English / Französisch): N 71–78

[5] Der Bundesrat hat Ende 2024 die Botschaft zur Teilrevision des BehiG verabschiedet, die insbesondere eine Ausweitung des Geltungsbereichs des Gesetzes auf alle Arbeitsverhältnisse, unabhängig davon, ob sie dem öffentlichen oder dem privaten Recht unterliegen, sowie eine Verbesserung der Zugänglichkeit von Dienstleistungen, insbesondere von Dienstleistungen, die von Privatpersonen für die Öffentlichkeit erbracht werden, vorsieht.

4. Égalité et non-discrimination

Les systèmes d'IA peuvent prendre des décisions discriminatoires et reproduire à large échelle des inégalités existantes dans notre société. Quels compléments sont nécessaires en droit suisse afin de lutter contre les discriminations algorithmiques ?

De quoi s'agit-il ?

L'utilisation de l'intelligence artificielle représente un **défi majeur** en matière d'égalité de traitement et d'interdiction des discriminations. En effet, **l'IA n'est pas neutre** : elle **reproduit les inégalités** déjà présentes au sein de notre société. En outre, il existe un risque que des systèmes d'IA soient utilisés par un grand nombre d'autorités et entreprises, p. ex. lors de processus de recrutement ou de l'octroi de prestations, avec pour effet d'exacerber les discriminations existantes et de leur donner une **dimension systémique**. Les **biais discriminatoires** peuvent notamment découler des données utilisées (p. ex. manque de représentativité, données obsolètes), de l'algorithme lui-même (p. ex. choix des variables prises en compte ou non dans le modèle), ou de la manière dont il est utilisé. En cas de résultat discriminatoire, l'effet boîte noire rend très difficile de déterminer les critères exacts sur lesquels repose ce résultat et ainsi la **preuve** de la discrimination.

Que prévoit la convention du Conseil de l'Europe ?

L'art. 10, par. 1, de la convention oblige chaque État à adopter ou maintenir des mesures visant à garantir le respect de l'égalité, y compris l'égalité de genre, et l'interdiction de la discrimination conformément au droit international et interne applicable. Le paragraphe 2 prévoit que chaque État 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.

Il est ici rappelé que la convention **ne créé pas de nouveaux droits fondamentaux** et ne change ainsi pas la portée du droit à l'égalité de traitement et à la non-discrimination telle qu'on la connaît actuellement en droit suisse. Cela étant, la convention demande aux États de fournir une **protection tout aussi valable et efficace** contre les discriminations dans le contexte spécifique des systèmes d'IA qu'en dehors.

Quelle est la situation actuelle en Suisse ?

- Au niveau de la **Constitution fédérale**, le principe d'égalité est ancré à l'art. 8, al. 1, et l'interdiction de la discrimination à l'art. 8, al. 2. L'art. 8, al. 3, garantit l'égalité de droit et de fait entre hommes et femmes. L'art. 8, al. 4, impose une obligation d'éliminer les inégalités qui frappent les personnes handicapées. Hormis l'égalité salariale entre femmes et hommes (art. 8, al. 3), le principe d'égalité et l'interdiction de discrimination s'appliquent à l'État et ne lient pas directement les privés.
- En vertu de l'art. 35 Cst., l'État est tenu de réaliser les droits fondamentaux dans l'ensemble de l'ordre juridique, y compris dans les relations entre privés, lorsque cela s'y prête. Le législateur suisse a réalisé le principe d'égalité de traitement et de non-discrimination entre particuliers **par type de discrimination ou dans des domaines spécifiques**. Il n'y a **pas d'interdiction générale de la discrimination entre privés**.
- En **droit privé**, certaines normes protègent indirectement contre la discrimination. Par exemple, on peut selon les cas se fonder sur les **art. 28 du code civil et 328 du code des obligations** pour requérir la cessation d'une discrimination ou une indemnité. La **loi**

fédérale sur l'égalité entre femmes et hommes (LEg) concrétise la protection contre les discriminations dans les relations de travail, y compris dans le secteur privé. En outre, **la loi fédérale sur l'élimination des inégalités frappant les personnes handicapées (LHand)** s'applique aux discriminations algorithmiques qui interviennent notamment dans les rapports de travail régis par la loi fédérale sur le personnel de la Confédération, la formation et la formation continue, et certaines prestations accessibles au public[6].

- Enfin, la **loi fédérale sur la protection des données**, bien qu'elle ne protège pas directement contre les discriminations, contient des garde-fous pertinents, notamment : protection accrue des données sensibles, obligation de base légale pour les organes fédéraux, principes de proportionnalité, d'exactitude, de finalité et de bonne foi, protection des données dès la conception et par défaut, analyses d'impact en cas de risque élevé, devoir d'information et droit d'accès, en particulier en cas de décisions exclusivement automatisées, obligations de documentation et de journalisation.

Malgré ce cadre, une partie de l'opinion estime que le droit suisse en matière d'égalité et non-discrimination reste **lacunaire**, que l'on recoure ou non à l'IA. En particulier, certaines voix se positionnent en faveur de l'élaboration d'une loi générale sur l'égalité de traitement et la non-discrimination.

Quels sont les enjeux ?

Le Conseil fédéral a reconnu qu'il existe des **risques accrus de discrimination** lorsqu'un système d'IA est utilisé, dans le secteur public et dans le secteur privé. Il a indiqué qu'il portera une **attention particulière** à cette thématique dans le cadre de ses travaux législatifs. En cas de volonté politique d'agir, il va de soi que tout renforcement de la législation anti-discrimination de manière générale serait bénéfique dans le contexte de l'IA.

Indépendamment de cette thématique, un des enjeux majeurs demeure celui d'identifier les possibles **interventions ciblées** du législateur, qui pourraient améliorer le cadre juridique existant. Parmi les différentes mesures, on peut citer notamment le renforcement de la **transparence** des systèmes d'IA et l'obligation de procéder à une **analyse des risques** sous l'angle des discriminations.

Informations supplémentaires : **Analyse jurisdique de base de l'OFJ** (Allemand / Français) : chapitre 4.3.2.5	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 71-78

[6] Le Conseil fédéral a adopté le message relatif à la révision partielle de la LHand fin 2024, prévoyant notamment une extension du champ d'application de la loi à tous les rapports de travail, qu'ils soient régis par le droit public ou privé, et une amélioration de l'accessibilité aux prestations de service, en particulier des prestations fournies au public par des particuliers.

5. Automatisierte Einzelentscheidungen

KI kann dazu dienen, Entscheidungen zu treffen, die die Rechtslage einer Person verändern oder sie in gleicher Weise beeinträchtigen. Welche Schutzmassnahmen gibt es, um den Einsatz von KI in Entscheidungsprozessen im öffentlichen und privaten Sektor zu ermöglichen und zu regeln?

Worum geht es?

Die Nutzung von künstlicher Intelligenz bei der Fällung von **Verwaltungs- oder privaten Entscheiden** kann die Wirksamkeit der Rechte der betroffenen Personen beeinträchtigen, insbesondere wenn die Entscheide ganz oder weitgehend von einem automatisierten System getroffen werden. In solchen Fällen ist es möglich, dass die betroffene Person **weder ihren Standpunkt einbringen noch Zugang zu einer menschlichen Aufsicht bekommen kann.** Zudem besteht das Risiko, dass ein Entscheid, der von einem auf maschinellem Lernen basierenden System getroffen wird, nicht nachvollziehbar begründet werden kann (sogenannter «Black-Box»-Effekt). Hinzu kommt, dass das System **diskriminierende Verzerrungen** aufweisen kann, sei es aufgrund mangelhafter Qualität der Trainingsdaten (unausgewogen, stereotypisch oder mit Unterrepräsentation bestimmter Bevölkerungsgruppen), sei es aufgrund der vom System verwendeten Entscheidungslogik (unangemessene Datengewichtung, fragwürdige Korrelationen oder strukturelle Verkürzungen).

Was sieht die KI-Konvention des Europarats vor?

Die KI-Konvention sieht verschiedene Garantien für die Entscheidungsprozesse vor, insbesondere:

- Gemäss Artikel 14 muss jeder Staat gewährleisten, dass zugängliche und wirksame **Rechtsmittel** gegen Menschenrechtsverletzungen bei der Nutzung von KI-Systemen verfügbar sind. Dazu gehört auch die Pflicht, relevante Informationen über die Funktionsweise dieser Systeme zu **dokumentieren und weiterzugeben**, damit die betroffenen Personen Entscheidungen, die von diesen Systemen getroffen wurden oder weitgehend auf ihnen beruhen, vor den zuständigen Behörden anfechten können.
- Gemäss Artikel 15 muss jeder Staat sicherstellen, dass wirksame **Schutzgarantien und Verfahrensrechte** zur Verfügung stehen, wenn ein KI-System erhebliche Auswirkungen auf die Menschenrechte hat. Abhängig vom Kontext müssen Personen, die mit KI-Systemen interagieren, zudem darüber informiert werden, dass sie mit einem solchen System und nicht mit einem Menschen interagieren.

Wie stellt sich die aktuelle Situation in der Schweiz dar?

In der Schweiz gibt es bereits mehrere Mechanismen zur Regelung von Entscheidungsprozessen, die durch ein KI-System ausgeführt oder von einem solchen unterstützt werden:

- **Informationspflicht, Recht auf Stellungnahme und Recht auf Erklärbarkeit:** Das Datenschutzgesetz (DSG) sieht eine Informationspflicht sowie ein Recht auf Darlegung des eigenen Standpunktes bei automatisierten Einzelentscheidungen sowohl im öffentlichen als auch im privaten Bereich vor (Art. 21 DSG). Es gewährt auch das Recht auf die Erteilung der Information zur Logik, auf der die Entscheidung beruht (Art. 25 Abs. 2 lit. f DSG). Diese Rechte stehen jedoch nur natürlichen Personen zu (Art. 1 DSG) und gelten nicht für die Fälle **teilautomatisierter** Entscheidungen (z. B. ein einfacher Score, der anschliessend in einen Entscheid einfliesst)
- Weiter sind im öffentlichen Recht relevant:

- **Recht der betroffenen Person auf rechtliches Gehör:** Das Gesetz garantiert der betroffenen Person das Recht, am Verfahren mitzuwirken, Tatsachen vorzubringen, die Akten einzusehen, Beweismittel einzureichen und die erhobenen Beweise anzufechten (Art. 29 Abs. 2 der Bundesverfassung [BV] und Art. 30 des Bundesgesetzes über das Verwaltungsverfahren [VwVG]). Ausnahmen bestehen insbesondere dann, wenn der Entscheid mit Einsprache angefochten werden kann (Art. 30 Abs. 2 lit. b VwVG) oder wenn er den Begehren der Parteien voll entspricht (Art. 30 Abs. 2 lit. c VwVG).
- **Begründungspflicht**: Die den Entscheid treffende Behörde muss diesen begründen. Auch wenn sich dies je nach Fall unterschiedlich gestalten kann, muss die Begründung den festgestellten Sachverhalt und die zugrunde liegende rechtliche Argumentation umfassen (Art. 29 Abs. 2 BV und 35 VwVG). Auf eine Begründung kann verzichtet werden, wenn der Entscheid den Begehren der Parteien voll entspricht (Art. 35 Abs. 3 VwVG).
- **Untersuchungsmaxime**: Im Verwaltungsverfahren ist die Behörde verpflichtet, den Sachverhalt von Amtes wegen, d. h. selbst zu ermitteln (Art. 12 VwVG). Diese Pflicht wird allerdings durch die Mitwirkungspflicht der Parteien teilweise relativiert (Art. 13 VwVG). Bei automatisierten Verfahren ist sicherzustellen, dass alle relevanten Tatsachen berücksichtigt werden.

Sowohl im öffentlich-rechtlichen als auch im privatrechtlichen Bereich stellt sich die Frage nach einer allfälligen **Ausweitung** der Informationspflicht – resp. der im Rahmen der Ausübung des Auskunftsrechts gemäss DSG zu liefernden Informationen – auch auf teilautomatisierte Entscheidungen.

Welche Herausforderungen stellen sich?

Das schweizerische Recht enthält somit bereits verschiedene einschlägige Bestimmungen im Zusammenhang mit automatisierten Entscheidungen, namentlich im Rahmen des **DSG** und des **Verwaltungsverfahrensrechts**. Diese decken jedoch nicht sämtliche Konstellationen ab, insbesondere nicht **teilautomatisierte Entscheidungen**, z. B. solche, die *mithilfe* von KI-Systemen getroffen werden. Ein Eingreifen des Gesetzgebers erscheint erforderlich, um auch diese Situationen zu erfassen und die diesbezüglichen Anforderungen festzulegen. Hinsichtlich erstinstanzlicher Verwaltungsentscheide versteht es sich von selbst, dass die Anforderungen des rechtlichen Gehörs zu wahren sind. Gleichwohl ist es angezeigt, die Wirksamkeit des bestehenden Rechtsrahmens zu überprüfen und zu analysieren, ob und unter welchen Bedingungen dieser **ergänzt oder neu überdacht werden** muss.

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.3.3.1 bis 4.3.3.3	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 95 – 104

5. Décisions individuelles automatisées

L'IA peut servir à rendre des décisions qui modifient la situation juridique d'une personne ou l'affecte de manière équivalente. Quels garde-fous pour permettre et encadrer l'usage de l'IA dans les processus décisionnels dans le secteur public et dans le secteur privé ?

De quoi s'agit-il ?

L'intégration de l'intelligence artificielle dans les **décisions administratives** ou **privées** peut mettre en danger l'effectivité des droits des personnes concernées lorsque ces décisions sont prises par un système automatisé ou largement fondées sur celui-ci. Il peut en résulter l'impossibilité pour la personne de **faire valoir son point de vue** et d'accéder à un décideur humain ou l'impossibilité d'expliquer les motifs de la décision si celle-ci est prise grâce à des techniques de *machine learning* (effet « boîte noire »). De plus, le système peut être entaché de **biais porteurs de discriminations**, soit en raison de la mauvaise qualité des données d'entraînement (déséquilibrées, stéréotypées ou sous-représentatives de certaines catégories de la population), soit en raison des logiques décisionnelles mises en œuvre par le système (pondérations inadéquates des données, corrélations douteuses, raccourcis structurels).

Que prévoit la convention du Conseil de l'Europe ?

La convention consacre plusieurs garanties se rapportant aux processus décisionnels, notamment :

- Selon l'article 14, chaque État doit garantir l'existence de **voies de recours** accessibles et effectives contre les violations des droits de l'homme liées à l'utilisation de systèmes d'IA. Cela inclut l'obligation **de documenter et de communiquer** des informations pertinentes sur le fonctionnement de ces systèmes, afin de permettre aux personnes concernées de contester auprès des autorités compétentes les décisions prises par ces derniers ou fondées en grande partie sur ceux-ci.
- Selon l'article 15, chaque État doit assurer que lorsqu'un système d'IA a un impact significatif sur les droits de l'homme, les personnes concernées bénéficient de **garanties, de protection et de droits procéduraux** effectifs. En fonction du contexte, les personnes qui interagissent avec des systèmes d'IA doivent en outre être informées du fait qu'elles interagissent avec de tels systèmes et non avec un humain.

Quelle est la situation actuelle en Suisse ?

En droit suisse, plusieurs mécanismes encadrent déjà les processus décisionnels réalisés ou assistés par un système d'IA :

- **Devoir d'informer, droit de faire valoir son point de vue et droit à l'explicabilité** : la loi fédérale sur la protection des données (LPD) impose une obligation d'information et un droit de faire valoir son point de vue en cas de décision individuelle automatisée, tant dans le secteur public que privé (art. 21 LPD). Elle permet aussi d'obtenir des informations sur la logique qui sous-tend une telle décision (art. 25 al. 2 let. f LPD). L'application de ces droits est toutefois limitée aux personnes physiques (art. 1 LPD) et ne couvre pas les cas de décisions **partiellement** automatisées (p. ex. un simple score ensuite utilisé dans le cadre d'une décision).
- En droit public, sont également pertinents :

- **Droit d'être entendu de la personne** : la loi reconnaît à l'administré le droit de participer à la procédure, d'alléguer des faits, de consulter le dossier, de produire des preuves et de contester celles administrées (art. 29 al. 2 de la Constitution fédérale [Cst.] et art. 30 de la loi fédérale sur la procédure administrative [PA]). Des exceptions sont prévues, notamment lorsque la décision est susceptible d'être frappée d'opposition (art. 30 al. 2, let. b PA) ou qu'elle fait entièrement droit aux conclusions des parties (art. 30 al. 2, let. c PA).
- **Devoir de motivation** : l'autorité qui rend la décision doit en exposer les motifs. Même si elle peut varier en fonction des cas, la motivation porte sur l'état de fait retenu et le raisonnement juridique suivi (art. 29 al. 2 Cst. et 35 PA). Il est possible de renoncer à la motivation lorsque la décision fait entièrement droit aux conclusions des parties (art. 35 al. 3 PA).
- **Maxime inquisitoire** : en procédure administrative, l'autorité est tenue d'établir les faits d'office, c'est-à-dire par elle-même (art. 12 PA). La règle est toutefois partiellement relativisée par l'obligation des parties de collaborer (art. 13 PA). En cas de procédure automatisée, il convient de garantir la prise en compte de tous les faits pertinents.

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.

Quels sont les enjeux ?

Le droit suisse contient ainsi déjà diverses dispositions pertinentes dans le contexte des décisions automatisées, notamment dans le cadre de la **LPD** et de la **procédure administrative**. Cependant, celles-ci ne couvrent pas toutes les constellations, à savoir en particulier les décisions **partiellement automatisées**, p. ex, celles prises *à l'aide* des systèmes d'IA. Une intervention du législateur paraît nécessaire afin de couvrir également ces situations et de définir les exigences 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. Néanmoins, il est opportun d'examiner l'efficacité du cadre légal existant et analyser si et à quelles conditions celui-ci doit être **complété ou repensé**.

Informations supplémentaires:	
Analyse jurisdique de base de l'OFJ (Allemand / Français) : chapitres 4.3.3.1 à 4.3.3.3	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 95 à 104

6. Technische Normen und Verhaltenskodizes

Im Bereich der KI gibt es zahlreiche technische Normen und Verhaltenskodizes, weitere befinden sich in Ausarbeitung. Welche Rolle sollen diese Instrumente in der Schweizer KI-Regulierung spielen?

Worum geht es?

Bestimmte Aspekte von künstlicher Intelligenz können mittels sogenannt «weichem Recht» (*soft law*) geregelt werden. Dafür kommen verschiedene Instrumente in Betracht. Zu unterscheiden ist dabei zwischen **Selbstregulierung**, bei der private Akteure ihre eigenen Regulierer sind, und **Koregulierung**, bei welcher der Staat aktiv in diese Selbstregulierung eingreift. Ohne Anspruch auf Vollständigkeit zeigt die nachstehende Tabelle einige Optionen auf:

Selbstregulierung	Koregulierung
• Spontane Verabschiedung branchenspezifischer Verhaltenskodizes ohne staatliches Eingreifen (z. B. Verhaltenskodizes von Berufsverbänden). • Ausarbeitung technischer Normen durch private Organisationen (ISO, IETF, W3C usw.) ohne Mandat oder Genehmigung durch den Gesetzgeber. • Schaffung privater Labels oder Zertifizierungen zur Stärkung des Marktvertrauens, ohne zwingende gesetzliche Anerkennung.	• Der Gesetzgeber legt allgemeine Grundsätze fest und überträgt die Konkretisierung in technische Normen an Normierungsorganisationen (z. B. technische Normen im Bereich Produktsicherheit). • Der Gesetzgeber verweist ausdrücklich auf technische Normen im Gesetz und verleiht ihnen damit rechtlich verbindlichen Charakter. • Der Gesetzgeber verweist indirekt und dynamisch auf den «Stand der Technik» und verpflichtet die Akteure, die Weiterentwicklung privater Normen zu berücksichtigen, um rechtskonform zu bleiben. • Zertifizierungen oder Verhaltenskodizes, die von privaten Akteuren ausgearbeitet, aber von einer öffentlichen Behörde validiert/genehmigt werden (z. B. Zertifizierungen im Bereich Datenschutz).

In diesem Zusammenhang stellen die **«technischen Normen»** nicht zwingende **Regeln, Leitlinien oder Spezifikationen** dar, die von Expertinnen und Experten innerhalb von Normierungsorganisationen ausgearbeitet werden. Technische Normen gelten für Verschiedenes, so für Produkte, Verfahren, Messmethoden, Prozesse und Dienstleistungen, und werden in nahezu allen Branchen und Fachgebieten angewandt. Sie können **rechtliche verbindliche Bestimmungen ergänzen und/oder deren Umsetzung unterstützen.** Technische Normen spielen eine **wichtige Rolle** im Bereich der Digitalisierung. Im Bereich der KI ist die internationale und europäische Normierungslandschaft bereits sehr weit entwickelt (vgl. z. B. Normen von ISO, IEC, ITU, CEN, CENELEC, ETSI).

Ein **«Verhaltenskodex»** hingegen ist ein Referenzrahmen, der von Akteuren eines bestimmten Sektors und nicht von einer offiziellen Organisation aufgestellt wird (vgl. z. B. Verhaltenskodizes nach Artikel 11 des Bundesgesetzes über den Datenschutz).

Was sieht die KI-Konvention des Europarats vor?

Artikel 12 der Konvention («Zuverlässigkeit») sieht vor, dass jeder Staat gegebenenfalls Massnahmen ergreift, um die Zuverlässigkeit von KI-Systemen und das Vertrauen in deren Ergebnisse zu fördern. Dies kann die Festlegung angemessener Anforderungen an Qualität und Sicherheit über den gesamten Lebenszyklus der KI-Systeme beinhalten.

Gemäss dem erläuternden Bericht unterstreicht diese Bestimmung die Bedeutung technischer Normen für die Bewertung und Überprüfung der **Zuverlässigkeit** von KI-Systemen so-

wie für die transparenten Dokumentation und Kommunikation der dabei erbrachten Nachweise. So könnten technische Normen den betroffenen Akteuren eine konkrete Orientierung bieten, wie eine **Folgenabschätzung** von KI-Systemen durchzuführen ist.

Der erläuternde Bericht zur Konvention weist zudem darauf hin, dass technische Normen im Rahmen eines **transparenten und inklusiven Prozesses** ausgearbeitet werden sollten, der die Kohärenz mit internationalen und nationalen Menschenrechtsinstrumenten fördert.

Die Konvention lässt den Vertragsstaaten grossen Spielraum bei der Wahl der geeigneten Massnahmen.

Wie stellt sich die aktuelle Situation in der Schweiz dar?

Die Schweiz wendet bereits gewisse Regeln an, die zwar nicht spezifisch auf KI zugeschnitten sind, in diesem Bereich aber dennoch relevant sind. Dazu gehören gesetzliche Normen im Bereich des Schutzes von Personendaten[7] und der Informationssicherheit[8], die im Zusammenhang mit KI-Systemen deren Zuverlässigkeit sowie deren funktionale Voraussetzungen wie Qualität, Integrität, Datensicherheit und Cybersicherheit verbessern können. Das Bundesgesetz über den Datenschutz sieht die Möglichkeit vor, Verhaltenskodizes (Art. 11) sowie Zertifizierungsmechanismen (Art. 13) im Zusammenhang mit der Bearbeitung personenbezogener Daten auszuarbeiten.

Derzeit sieht das Schweizer Recht keinen **rechtlichen Mechanismus** vor, der die Interaktion zwischen *soft law* und gesetzlichen Normen im Bereich der KI regelt. Solche Mechanismen bestehen hingegen im Datenschutzrecht.

Welches sind die Herausforderungen?

In der Schweiz haben sich gewisse Sektoren bereits Verhaltenskodizes im Bereich der KI gegeben (vgl. z. B. die Leitlinien des Schweizer Presserats zur KI). Zudem werden technische Normen im Bereich KI in der Praxis **bereits angewendet**: Sie stellen für die betroffenen Akteure ein wichtiges Mittel dar, um in ihrem Tätigkeitsbereich günstige Rahmenbedingungen zu schaffen. Auch international entwickelte Normen zur KI können Auswirkungen auf die Schweizer Akteure haben.

Es gilt daher zu **klären**, ***wie* die zukünftige Schweizer Gesetzgebung im Bereich KI mit *soft law* interagieren kann**. Diese Problematik wirft mehrere Fragen auf, insbesondere:

- Welche normative Rolle dem *soft law* im Rahmen der künftigen Schweizer KI-Regulierung zukommen soll;
- die mögliche Integration technischer Normen in die künftige Gesetzgebung oder deren Bezugnahme darauf;
- die Themenbereiche, bei denen es sinnvoll wäre, auf technische Normen zu verweisen (z. B. Anforderungen an Folgenabschätzungen), und gegebenenfalls die Auswahl der relevanten technischen Normen.

Wir weisen darauf hin, dass das Eidgenössische Departement für Umwelt, Verkehr, Energie und Kommunikation (**UVEK**) den Auftrag erhalten hat, einen **Umsetzungsplan für nicht verbindliche Massnahmen** als Ergänzung zum Vorentwurf des Eidgenössischen Justiz- und Polizeidepartements (EJPD) zu erarbeiten. Die Frage nach der rechtlichen Tragweite des *soft law* muss jedoch im Rahmen der legislativen Arbeiten des EJPD geklärt werden, da es sich hierbei **um eine auf normativer Ebene zu regelnde Frage** handelt.

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.3.2.7	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 84 – 89 **SECO, Förderung der Normierungsorganisationen im Bereich der Digitalisierung:** Bericht an den Bundesrat, 16. August 2022, (Deutsch)

[7] Vgl. Insbesondere die Artikel 6, 7, 8, 13 und 24 des Bundesgesetzes über den Datenschutz, und die Artikel 1 bis 6 der Datenschutzverordnung.

[8] Vgl. Insbesondere das Bundesgesetz zur Informationssicherheit und die dazugehörenden Verordnungen.

6. Normes techniques et codes de conduite

De nombreuses normes techniques et codes de conduite existent déjà ou sont en cours d'élaboration dans le domaine de l'IA. Quel rôle peuvent jouer ces instruments dans la réglementation suisse de l'IA ?

De quoi s'agit-il ?

La réglementation de certains aspects de l'intelligence artificielle peut se faire par le biais de droit dit souple (« **soft law** »). Divers instruments entrent en ligne de compte. Il convient dans ce contexte de distinguer entre l'**autorégulation**, dans laquelle les milieux privés sont leurs propres régulateurs, et la **corégulation**, où l'État intervient activement dans l'activité d'autorégulation du secteur privé. Sans prétendre à l'exhaustivité, le tableau ci-dessous illustre quelques options :

Autorégulation	Corégulation
• Adoption spontanée de codes de conduite sectoriels sans intervention étatique (ex. codes de bonne pratique des associations professionnelles). • Élaboration de normes techniques par des organismes privés (ISO, IETF, W3C, etc.) sans mandat ni approbation du législateur. • Création de labels ou certifications privés destinés à renforcer la confiance du marché, sans reconnaissance juridique obligatoire.	• Le législateur établit des principes généraux et délègue aux organismes de normalisation le soin de les concrétiser par des normes techniques (p. ex. normes techniques en matière de sécurité des produits). • Le législateur se réfère explicitement à des normes techniques dans la loi, leur conférant une valeur juridique contraignante. • Le législateur renvoie indirectement et dynamiquement à « l'état de la technique », obligeant les acteurs à suivre l'évolution des normes privées pour être en conformité légale. • Certification ou codes de conduite élaborés par des acteurs privés mais validés/agréés par une autorité publique (ex. certifications en matière de protection des données).

Dans ce contexte, les « **normes techniques** » constituent des **règles, lignes directrices ou spécifications non contraignantes** élaborées par des experts au sein d'organismes de normalisation. Les normes techniques régissent divers objets tels que les produits, les procédures, les méthodes de mesure, les processus et les services, et sont utilisées dans presque tous les secteurs et domaines spécialisés. Les normes techniques peuvent **compléter et/ou soutenir la mise en œuvre de dispositions juridiquement contraignantes**. Ces normes jouent un **rôle important** en matière de numérisation. Dans le domaine de l'IA, le paysage international et européen de la normalisation est déjà très développé (cf. p. ex. normes ISO, CEI, UIT, CEN, CENELEC, ETSI).

Un « **code de conduite** » est en revanche un cadre de référence établi par des acteurs d'un secteur plutôt que par un organisme officiel (cf. par exemple les codes de conduite selon l'art. 11 de la loi fédérale sur la protection des données).

Que prévoit la convention du Conseil de l'Europe ?

L'art. 12 (« Fiabilité ») de la convention prévoit que chaque État prend, le cas échéant, des mesures pour promouvoir la fiabilité des systèmes d'intelligence artificielle et la confiance en leurs résultats, ce qui pourrait inclure des exigences en matière de qualité et de sécurité adéquates tout au long du cycle de vie des systèmes d'intelligence artificielle.

Selon le rapport explicatif, cette disposition souligne le rôle que peuvent jouer les normes techniques, notamment, dans l'évaluation et la vérification de la **fiabilité** des systèmes d'IA, ainsi que dans la documentation et la communication transparentes des éléments de preuve de ce processus. Les normes techniques pourraient par exemple être utilisées afin de fournir des orientations concrètes aux acteurs concernés sur la manière de réaliser une **analyse des impacts** des systèmes d'IA.

Le rapport explicatif de la convention relève également qu'il convient de veiller à ce que les normes techniques soient élaborées dans le cadre d'un **processus transparent et inclusif**, qui encourage la cohérence avec les instruments internationaux et nationaux en matière de droits de l'homme.

La convention laisse une ample marge de manœuvre aux États Parties quant au choix des mesures appropriées à prendre.

Quelle est la situation actuelle en Suisse ?

La Suisse applique déjà certaines règles qui, bien qu'elles ne soient pas spécifiques à l'IA, sont tout de même pertinentes dans ce domaine. Il s'agit de normes légales en matière de sécurité des données personnelles[9] et de sécurité de l'information[10] qui, en présence de systèmes d'IA, permettent d'améliorer leur fiabilité, ainsi que leurs conditions fonctionnelles préalables, telles que la qualité, l'intégrité, la sécurité des données ainsi que la cybersécurité. La loi fédérale sur la protection des données prévoit la possibilité d'élaborer des codes de conduite (art. 11), ainsi que des mécanismes de certification (art. 13), se rapportant au traitement de données personnelles.

Actuellement, le droit suisse ne prévoit pas de **mécanisme juridique** régissant **l'interaction** entre droit souple et normes légales dans le domaine de l'IA. De tels mécanismes existent en revanche dans le droit de la protection des données.

Quels sont les enjeux ?

En Suisse, certains secteurs se sont déjà dotés de codes de conduite en matière d'IA (cf. par exemple les lignes directrices en matière d'IA du Conseil suisse de la presse). En outre, le recours à des normes techniques en matière d'IA est déjà **une réalité** en pratique : ces dernières constituent un moyen important pour les acteurs concernés de créer des conditions-cadre favorables en fonction de leur domaine d'activité. Les normes techniques sur l'IA développées à l'échelle internationale peuvent aussi avoir un impact sur les acteurs suisses.

Il convient dès lors de **clarifier *comment* la future législation suisse en matière d'IA pourra interagir avec le droit souple**. Cette question soulève plusieurs enjeux, notamment :

- Le rôle normatif à attribuer au droit souple dans le cadre de la future réglementation suisse en matière d'IA ;
- L'intégration éventuelle des normes techniques dans la future législation ou le renvoi à celle-ci ;
- Les thématiques par rapport auxquelles il serait opportun de renvoyer aux normes techniques (p. ex. exigences en matière d'analyse d'impact) et la sélection des normes techniques pertinentes, le cas échéant.

Nous relevons que le Département fédéral de l'environnement, des transports, de l'énergie et de la communication (**DETEC**) a reçu un mandat pour l'élaboration d'un **plan de mise en œuvre de mesures non-contraignantes** en complément à l'avant-projet législatif du Département fédéral de justice et police (DFJP). La problématique de la portée juridique du droit souple doit pourtant être tranchée dans le cadre des travaux législatifs du DFJP, car il s'agit d'une **question à régler sur le plan normatif**.

Informations supplémentaires :	
Analyse jurisdique de base de l'OFJ (Allemand / Français) : chapitre 4.3.2.7	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 84 – 89 **SECO, Förderung der Normungsorganisationen im Bereich der Digitalisierung: (**Bericht an den Bundesrat, 16. August 2022, Deutsch)

[9] Cf. notamment les articles 6, 7, 8, 13 et 24 de la loi fédérale sur la protection des données, et les articles 1 à 6 de l'ordonnance sur la protection des données.

[10] Cf. en particulier la loi fédérale sur la sécurité de l'information et ses ordonnances.

7. Aufsicht

Die Schaffung einer Aufsichtsbehörde für KI wirft Fragen hinsichtlich des Anwendungsbereichs, der Zuständigkeiten und der Koordinierung mit anderen bestehenden Aufsichtsbehörden auf. Welche Behörde(n) soll(en) die Nutzung von KI in der Schweiz überwachen und mit welchen Zuständigkeiten?

Worum geht es?

Die zunehmende Verbreitung von KI-Systemen wirft die Frage auf, wie **Kontrollmechanismen** gestaltet sein müssen, damit ihre Entwicklung und Nutzung die Grundrechte, die Demokratie und die Rechtsstaatlichkeit wahren. Anders als in anderen Bereichen (Datenschutz, Wettbewerb, Finanzmärkte) gibt es in der Schweiz bislang **keine spezialisierte Behörde**, die den Einsatz von KI bereichsübergreifend im Hinblick auf den Schutz der Menschenrechte, den Rechtsstaat und die Demokratie beaufsichtigt. Allerdings existieren **zahlreiche sektorielle Behörden**, in deren Aufsichtsbereich auch Fragen der KI-Nutzung fallen können.

Was sieht die Konvention des Europarats vor?

Die Konvention hält verschiedene Garantien zur Einrichtung von **Kontrollmechanismen** im innerstaatlichen Recht fest:

- Artikel 14 verpflichtet die Staaten, betroffenen Personen bei der zuständigen Behörde ein **wirksames Rechtsmittel** gegen Menschenrechtsverletzungen, die innerhalb des Lebenszyklus von KI-Systemen auftreten, **zugänglich zu machen**.
- Artikel 26 verpflichtet die Vertragsparteien, **einen oder mehrere wirksame Mechanismen** zur Aufsicht über die Einhaltung der Verpflichtungen aus der Konvention einzurichten oder zu bestimmen. Diese Mechanismen müssen ihre Aufgaben **unabhängig** und unparteiisch wahrnehmen und über die notwendigen Kompetenzen, Fachkenntnisse und Ressourcen verfügen, um ihren Auftrag wirksam zu erfüllen. Bestehen mehrere Mechanismen nebeneinander, muss der Staat Massnahmen treffen, um deren **Zusammenarbeit** sicherzustellen. Weichen die eingerichteten Mechanismen von den bereits bestehenden nationalen Strukturen im Bereich der Menschenrechte ab, muss der Staat zudem für eine **enge** und wirksame **Zusammenarbeit** zwischen ihnen sorgen.

Was ist die aktuelle Situation in der Schweiz?

In der Schweiz verfügen bereits mehrere Behörden über Teilkompetenzen im Zusammenhang mit KI, zum Beispiel:

- **Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB)** für alle Aspekte im Zusammenhang mit der Bearbeitung von Personendaten.
- **Eidgenössische Finanzmarktaufsicht (FINMA)** für Fragen der Governance und des Risikomanagements von KI bei Beaufsichtigten (z. B. Banken, Versicherungen).
- **Bundesamt für Kommunikation (BAKOM)** für bestimmte sektorielle Aspekte im Bereich Telekommunikation und Medien.
- **Verschiedene kantonale Behörden** für die Aufsicht über kantonales Recht in verschiedenen kantonal geregelten Bereichen (Datenschutz, Gewerbepolizei, Raumplanung usw.).
- **Gerichte**: Verwaltungs-, Zivil- und Strafgerichte können in Gerichtsfällen im Zusammenhang mit der Nutzung von KI in unterschiedlichen Kontexten urteilen.

Bislang gibt es jedoch in der Schweiz **keine** Aufsichtsbehörde, die den gesamten Anwendungsbereich der Konvention abdeckt. Zudem erfüllen nicht alle bestehenden Aufsichtsbehörden die in der Konvention festgelegten **Anforderungen an die Unabhängigkeit**.

Welche Herausforderungen stellen sich?

Es bedarf eines Eingreifens des Gesetzgebers, um eine oder mehrere bestehende Behörden mit der Überwachung der Einhaltung der Konvention zu beauftragen. Der Gesetzgeber müsste den **Anwendungsbereich, die Organisation und die Eingriffsbefugnisse** der Aufsichtsbehörde(n) definieren. Sofern mehrere Behörden zuständig sind, müssen **Koordinationsmechanismen** vorgesehen werden, um unerwünschte Überschneidungen und widersprüchliche Entscheide möglichst zu vermeiden. Die Schaffung einer **neuen eidgenössischen KI-Aufsichtsbehörde** ist eine denkbare Option, über die politisch jedoch noch nicht entschieden wurde. Eine Alternative bestünde darin, einer bestehenden Behörde **ein erweitertes Mandat** zu erteilen – mit dem Risiko einer Zersplitterung der Zuständigkeiten. Schliesslich ist auch das Verhältnis zwischen der Rolle der Aufsichtsbehörden und jener der **Gerichte** gegenüber den betroffenen Personen zu klären.

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.4.5	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 141 bis 144

7. Surveillance

La mise en place d'un organe de surveillance de l'IA pose des questions de champ d'application, de compétences et de coordination avec les autres organes de surveillance existant. Quelle(s) autorité(s) pour surveiller l'usage de l'IA en Suisse et avec quelles compétences ?

De quoi s'agit-il ?

L'essor des systèmes d'intelligence artificielle soulève la question des **mécanismes de contrôle** permettant d'assurer que leur développement et leur utilisation respectent les droits fondamentaux, la démocratie et l'État de droit. Contrairement à d'autres domaines (protection des données, concurrence, marchés financiers), il n'existe aujourd'hui en Suisse pas d'**autorité spécialisée** pour encadrer l'usage de l'IA de manière transversale autour des thèmes de la protection des droits de l'homme, de l'État de droit et de la démocratie. Il existe toutefois de **nombreuses autorités sectorielles** dont le domaine de surveillance peut également toucher des questions liées à l'usage de l'IA.

Que prévoit la convention du Conseil de l'Europe ?

La convention consacre différentes garanties concernant l'institution en droit interne de **mécanismes de contrôle** :

- Selon l'article 14 chaque État est tenu d'offrir aux personnes concernées une **possibilité effective de former un recours** auprès des autorités compétentes contre les violations des droits de l'homme résultant des activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle.
- Selon l'article 26, chaque État doit mettre en place ou désigner **un ou plusieurs mécanismes effectifs** chargés de contrôler le respect des obligations issues de la convention. Ces mécanismes doivent exercer leurs fonctions de manière **indépendante** et impartiale, tout en disposant des compétences, de l'expertise et des ressources nécessaires pour accomplir efficacement leur mission. Lorsque plusieurs mécanismes coexistent, l'État doit prendre des mesures pour assurer leur **coopération**. Enfin, si les mécanismes institués diffèrent des structures nationales déjà existantes en matière de droits de l'homme, l'État doit également veiller à promouvoir une **collaboration étroite** et efficace entre eux.

Quelle est la situation actuelle en Suisse ?

En Suisse, plusieurs autorités disposent déjà de compétences partielles touchant à l'IA, par exemple :

- **Préposé fédéral à la protection des données et à la transparence (PFPDT)** pour toute la dimension concernant le traitement de données personnelles.
- **Autorité fédérale de surveillances des marchés financiers (FINMA)** pour tout ce qui se rapporte à la gouvernance et à la gestion des risques de l'IA auprès des assujettis (p. ex. banques, assurances).
- **Office fédéral de la communication (OFCOM)** pour certains aspects sectoriels liés aux télécommunications et aux médias.
- **Différentes autorités cantonales** pour la surveillance du droit cantonal dans divers secteurs soumis au droit cantonal (protection des données, police du commerce, aménagement du territoire, etc.).

- **Tribunaux** : les tribunaux administratifs, civils et pénaux peuvent être amenés à trancher des litiges portant sur des questions relatives à l'usage de l'IA dans différents contextes.

À ce jour, il n'existe **pas** en Suisse d'autorité de surveillance qui couvre tout le champ de la convention. En outre, toutes les autorités de surveillance existantes ne satisfont pas **aux exigences d'indépendance** requises par la convention.

Quels sont les enjeux ?

Une intervention du législateur est nécessaire pour confier à une ou plusieurs autorités existantes la tâche de contrôler le respect de la convention. Le législateur devrait définir le **champ d'application, l'organisation et les pouvoirs d'intervention** de l'autorité ou des autorités de surveillance. Si plusieurs autorités sont compétentes, il convient de prévoir des **mécanismes de coordination** afin d'éviter au maximum les chevauchements indésirables et les décisions contradictoires. L'option consistant à créer **une nouvelle autorité fédérale** de surveillance de l'IA est envisageable, mais n'a pas encore fait l'objet de décision politique. Une alternative serait d'attribuer **un mandat élargi** à une autorité existante, avec toutefois le risque de dispersion des compétences. Enfin, il importe de vérifier l'articulation entre le rôle des autorités de surveillance et celui des **tribunaux** vis-à-vis des particuliers.

Informations supplémentaires :	
Analyse jurisdique de base de l'OFJ (Allemand / Français) : chapitre 4.4.5	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 141 à 144

8. Transparenz

Transparenz ist der Eckpfeiler einer vertrauenswürdigen KI: Sie ermöglicht es, Entscheidungen zu verstehen und gegebenenfalls anzufechten, Bias zu erkennen, Vertrauen zu stärken und die Zuverlässigkeit der Systeme zu gewährleisten. Welche konkreten Massnahmen müssen ergriffen werden, damit sie während des gesamten Lebenszyklus der KI eingehalten wird?

Worum geht es?

Transparenz bei KI-Systemen ist kein Selbstzweck, sondern eine Grundvoraussetzung für die Entwicklung und den Einsatz **vertrauenswürdiger KI, welche die Grundrechte wahrt.** Sie sollte den gesamten Lebenszyklus eines KI-Systems umfassen. Nur dank Transparenz lässt sich beispielsweise die Qualität der Trainingsdaten überprüfen, Verzerrungen erkennen und das korrekte Funktionieren der programmierten Algorithmen sicherstellen. Zudem sollte die Anwendung des Transparenzprinzips die Nachvollziehbarkeit der Systeme fördern und den Nutzerinnen und Nutzern vermitteln, wie diese zu ihren Entscheidungen gelangen. Eine erhöhte Transparenz trägt letztlich dazu bei, die Achtung aller Grundrechte zu stärken, insbesondere das Recht auf informationelle Selbstbestimmung, das Diskriminierungsverbot sowie das Recht auf rechtliches Gehör.

Was sieht die KI-Konvention des Europarates vor?

Die Konvention enthält mehrere Bestimmungen zu Transparenz, so zum Beispiel:

- **Artikel 8** verpflichtet die Staaten, Massnahmen zu ergreifen, um Transparenz und Kontrolle der Tätigkeiten im Lebenszyklus von KI-Systemen sicherzustellen. Dazu gehören die **Kennzeichnung** von KI-generierten Inhalten, die **Erklärbarkeit** und **Interpretierbarkeit** von KI-Systemen sowie die Einführung von **Kontrollmechanismen** zur Überwachung, Bewertung und Steuerung der Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen.
- **Artikel 15** verpflichtet die Staaten, sicherzustellen, dass betroffene Personen **wirksame Garantien, Schutz und Verfahrensrechte** geniessen, wenn ein KI-System erhebliche Auswirkungen hat. Je nach Kontext müssen **Personen, die mit KI-Systemen interagieren, darüber informiert werden**.

Hinweis: Fragen zur Transparenz von KI in Entscheidungsprozessen, wie auch automatisierte Entscheidungen im öffentlichen und Privatrecht werden in einer anderen thematischen Gruppe behandelt.

Wie stellt sich die aktuelle Situation in der Schweiz dar

Im schweizerischen Recht bestehen bereits verschiedene Mechanismen, die eine gewisse Transparenz im Zusammenhang mit KI-Systemen gewährleisten:

- **Zugang zum Quellcode**: Artikel 9 des **Bundesgesetzes über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben (EMBAG)** verpflichtet die Bundesverwaltung, den Quellcode von Software, die sie entwickelt oder entwickeln lässt, aktiv zu veröffentlichen, wobei es Ausnahmen gibt. Auch das **Bundesgesetz über das Öffentlichkeitsprinzip in der Verwaltung (BGÖ)** könnte so ausgelegt werden, dass der Quellcode einer Software ein amtliches Dokument darstellt, das dem Öffentlichkeitsprinzip unterliegt. Für KI-Systeme, die auf *machine learning* beruhen, ermöglicht der Zugang

zum Quellcode bzw. zu den Modellparametern allerdings nicht ohne Weiteres ein Verständnis ihres Funktionsmechanismus (sogenannter Black-Box-Effekt).

- **Transparenz im Datenschutzgesetz (DSG)**: Das DSG **enthält mehrere Bestimmungen, die direkt oder indirekt Transparenz bei der Bearbeitung** von Personendaten gewährleisten sollen. Dazu gehören die allgemeinen Grundsätze (insbesondere Erkennbarkeit und Zweckbindung, Datenschutz durch Technik und durch datenschutzfreundliche Voreinstellungen, Sicherheit (vgl. Artikel 6, 7 und 8 DSG), die Pflichten zur Meldung automatisierter Bearbeitungen und Registerführung (Art. 12, 56 und 31 DSV[11]), die Informationspflichten bei der Beschaffung von Daten (Art. 19 DSG), das Auskunftsrecht (Art. 25 Abs. 2 DSG) sowie Dokumentations- und Protokollierungspflichten (Art. 3–5 DSV). Im öffentlichen Sektor trägt zudem das Erfordernis einer gesetzlichen Grundlage (Art. 34 DSG) bis zu einem gewissen Grad zur Transparenz bei.
- **Kompetenznetzwerk Künstliche Intelligenz (CNAI** – Competence Network for Artifical Intelligence): Dieses Netzwerk führt eine **Liste der KI-Projekte der Bundesverwaltung**. Die Meldung der Projekte erfolgt auf **freiwilliger** Basis.

Welche Herausforderungen stellen sich?

Das schweizerische Recht enthält nur punktuell Regelungen zur Sicherstellung von Transparenz im Zusammenhang mit KI-Systemen. **Angesichts der zunehmenden Verbreitung dieser Technologien** stellt sich jedoch die **Frage nach ergänzenden Massnahmen** im öffentlichen wie auch im privaten Sektor. So sieht das geltende Recht derzeit weder eine allgemeine Informationspflicht bei Interaktionen mit KI-Systemen noch Massnahmen zur Kennzeichnung KI-generierter Inhalte oder eine allgemeine Meldepflicht für KI-Projekte in einem Register vor. Auch die Transparenzpflichten des DSG sind nicht spezifisch auf KI-Systeme ausgerichtet und beschränken sich auf Fälle der Bearbeitung von Personendaten. Zwar ist die Pflicht zur Veröffentlichung von Quellcodes grundsätzlich positiv zu bewerten, ihre Wirkung bleibt im Kontext von KI jedoch häufig begrenzt.

Weitere Informationen: **Rechtliche Basisanalyse des BJ** (Deutsch / Französisch): Kapitel 4.3.2.3, 4.3.3.2, 4.3.3.2	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 56 ff.

[11] Datenschutzverordnung

8. Transparence

La transparence est la pierre angulaire d'une IA vertueuse : elle permet de comprendre et, si nécessaire, contester ses décisions, de détecter les biais, de renforcer la confiance et de garantir la fiabilité des systèmes. Quelles mesures concrètes faut-il adopter pour qu'elle soit respectée tout au long du cycle de vie de l'IA ?

De quoi s'agit-il ?

La transparence des systèmes d'IA n'est pas une fin en soi, mais un prérequis pour le développement et l'utilisation d'une IA **digne de confiance et respectueuse de droits fondamentaux**. Elle devrait s'appliquer à l'entier du cycle de vie des systèmes d'IA. C'est notamment grâce à elle que l'on pourra s'assurer de la qualité de l'entraînement des systèmes d'IA, de l'absence de biais, ou encore du bon fonctionnement des algorithmes programmés. Par ailleurs, l'application de ce principe devrait aussi permettre de garantir l'explicabilité des systèmes et le fait que les utilisateurs puissent comprendre comment ils prennent leurs décisions. Une plus grande transparence permet finalement de renforcer et de garantir le respect de l'ensemble des droits fondamentaux, en particulier l'autodétermination informationnelle, l'interdiction de la discrimination ou encore le droit d'être entendu.

Que prévoit la convention du Conseil de l'Europe ?

La convention contient plusieurs dispositions en lien avec la transparence, telle que :

- **L'article 8**, impose aux États de mettre en place des mesures pour assurer la transparence et le contrôle des activités menées dans le cadre du cycle de vie des systèmes d'IA. Cela inclut **l'identification** des contenus générés par des systèmes d'IA, **l'explicabilité** et **l'interprétabilité** des systèmes d'IA ainsi que la mise en place de **mécanismes de contrôle** pour surveiller, évaluer et orienter les activités menées dans le cadre du cycle de vie des systèmes d'IA.
- **L'article 15** enjoint les États à s'assurer que les personnes concernées **bénéficient de garanties, de protection et de droits procéduraux effectifs** lorsqu'un système d'IA a un impact significatif. En fonction du contexte, les personnes qui **interagissent avec des systèmes d'IA doivent en être informées.**

À noter que les questions en lien avec la transparence de l'IA dans les processus décisionnels, comme les décisions automatisées en droit privé et public, font l'objet d'un autre groupe thématique.

Quelle est la situation actuelle en Suisse ?

En droit suisse, plusieurs mécanismes permettent déjà d'assurer une certaine transparence en lien avec les systèmes d'IA :

- **Accès aux codes source** : l'article. **9 de la loi fédérale sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités (LMETA)** oblige l'administration fédérale à publier activement le code source des logiciels qu'elle développe ou fait développer, sauf exception. La **loi fédérale sur le principe de la transparence dans l'administration (LTrans)** pourrait aussi être interprétée de telle manière que le code source d'un logiciel constitue un document officiel soumis au principe de la transparence. Néanmoins, pour des systèmes d'IA reposant sur le *machine learning*, l'accès au code source, resp. aux paramètres du modèle, ne permet pas d'emblée d'en comprendre le fonctionnement (effet « boîte noire »).

- **Transparence dans la loi sur la protection des données (LPD) :** la LPD **prévoit plusieurs articles visant à garantir, directement ou non, la transparence des traitements** de données personnelles des personnes physiques. Il s'agit des principes généraux (en particulier reconnaissabilité et finalité, 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, 56 et 31 OPDo[12]), des devoirs d'information en cas de collecte de données (art. 19 LPD), du droit d'accès (art. 25, al. 2 LPD) ou encore des obligations en matière de documentation et de journalisation (art. 3 à 5 OPDo). Dans le secteur public, l'exigence de base légale (art. 34 LPD) permet aussi d'assurer une certaine transparence.
- Le réseau de compétences en intelligence artificielle (**CNAI** – Competence Network for Artificial Intelligence) tient une **liste des projets d'IA de l'administration fédérale**. Les projets sont annoncés sur une base **volontaire**.

Quels sont les enjeux ?

De manière générale, le droit suisse prévoit des dispositions permettant d'assurer ponctuellement la transparence en lien avec les systèmes d'IA. Néanmoins, **au vu de l'utilisation banalisée de ces systèmes,** il convient d'**examiner le besoin de prendre des mesures supplémentaires** dans le secteur public et dans le secteur privé. Le droit suisse ne prévoit en effet pas d'obligation générale d'informer en cas d'interaction avec un système d'IA, de mesures permettant l'identification des contenus générés par l'IA, ou encore d'obligation générale d'annoncer les projets d'IA dans un registre. Par ailleurs, les obligations de transparence qui découlent de la LPD ne renseignent pas spécifiquement sur les systèmes d'IA et sont limitées aux cas dans lesquels des données personnelles de particuliers sont concernées. Enfin, l'obligation de mise à disposition des codes source est certes louable, mais ses effets sont souvent limités dans le contexte de l'IA.

Informations supplémentaires : **Analyse jurisdique de base de l'OFJ** Allemand / Français) : chapitres 4.3.2.3, 4.3.3.2, 4.3.3.2	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 56ss

[12] Ordonnance sur la protection des données

9. Innovation und Forschung

Die Schweiz verfügt über ein innovationsfreundliches Umfeld im Bereich der KI, das von Forschungsinstituten und Unternehmen voll ausgeschöpft wird. Mit welchen Massnahmen kann die Forschung weiter gefördert und gleichzeitig die Grundrechte und das öffentliche Interesse geschützt werden?

Worum geht es?

Künstliche Intelligenz muss sich in einem rechtlichen Rahmen entwickeln können, der zugleich flexibel und schützend ist. Fehlen geeignete Schutzvorkehrungen, drohen **Rechtsunsicherheit, Missbrauch** sowie eine Gefährdung von Grundrechten, Rechtsstaatlichkeit und Demokratie. Umgekehrt kann eine zu strenge und nicht ausreichend flexible Regulierung die Forschung hemmen und verhindern, dass **Gesellschaft und Wirtschaft** von den Chancen der Innovation **profitieren**.

Was sieht die KI-Konvention des Europarats vor?

Unter Vorbehalt von Artikel 13 und 25 Absatz 2 **gilt die Konvention nicht für Forschungs- und Entwicklungstätigkeiten** im Zusammenhang mit KI-Systemen, die noch nicht zur Nutzung bereitgestellt wurden, **sofern die Versuche nicht geeignet sind**, die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit zu **beeinträchtigen**. Mit anderen Worten: Rein experimentelle KI-Systeme fallen nicht unter die in der Konvention festgelegten Grundsätze. KI-Systeme, die im Anschluss an Forschungs- und Entwicklungstätigkeiten **zur Nutzung bereitgestellt werden**, müssen hingegen grundsätzlich die Bestimmungen der Konvention einhalten, auch hinsichtlich ihrer Konzeption und Entwicklung (Art. 3 Abs.2).

Artikel 13 fordert die Staaten auf, gegebenenfalls **die Einrichtung kontrollierter Umgebungen für die Entwicklung**, Erprobung und Testung von KI-Systemen unter Aufsicht der zuständigen Behörden zu ermöglichen, **um Innovation zu fördern und zugleich negative Auswirkungen auf die Menschenrechte zu vermeiden**. Die Konvention überlässt es den Staaten, die Modalitäten der Umsetzung dieser Bestimmung festzulegen. Dem erläuternden Bericht zufolge kann es sich dabei insbesondere um sogenannte «regulatorische Sandboxen» (*regulatory sandboxes*[13]), um Empfehlungen oder um Garantien der Verfahrensaussetzung handeln, die klarstellen, wie die Regulierungsbehörden die verschiedenen Lebenszyklen von KI handhaben.

Wie stellt sich die aktuelle Situation in der Schweiz dar?

In der Schweiz sind Unternehmen und Forschungsinstitute sehr aktiv in der KI-Forschung. In unterschiedlichen Bereichen wie Finanzwesen, Forschung, Gesundheit oder Industrie sind **zahlreiche Innovationszentren ins Leben gerufen worden**.

Die **Schweizer Gesetzgebung enthält mehrere Bestimmungen, welche Pilotprojekte auch unter Einsatz von KI ermöglichen**. Beispiele sind etwa:

[13] Nach der hier in der juristischen Analyse des BJ zugrunde gelegten Definition umfassen regulatorische Sandboxen einerseits Pilotprojekte und andererseits *Sandboxen im engeren Sinn*: *Pilotprojekte* dienen dazu, neue Technologien oder neue Verfahren unter realen Bedingungen während einer begrenzten Zeitspanne zu erproben, wobei in einzelnen Aspekten von der bestehenden Regulierung abgewichen wird. Ihre Ergebnisse zeigen auf, inwieweit die Regulierung überarbeitet werden muss. *Sandboxen im engeren Sinn* hingegen ermöglichen es Unternehmen, Verfahren, Produkte und Dienstleistungen auf dem Markt mithilfe befristeter Ausnahmen von bestimmten Regulierungen zu testen. Verstösse gegen diese Regulierungen werden zwar korrigiert, jedoch nicht sanktioniert. Auf diese Weise können Unternehmen herausfinden, ob ihre innovativen Geschäftsmodelle im geltenden Rechtsrahmen bestehen können. Sie können sich so ein Bild davon machen, was mit der geltenden Regulierung möglich ist – und was nicht. Die Erprobungen im Rahmen von Sandboxen können entweder unter realen Bedingungen oder in einem simulierten und/oder kontrollierten Umfeld stattfinden.

- Artikel 35 des Bundesgesetzes über den Datenschutz sieht beispielsweise vor, dass der Bundesrat unter bestimmten Bedingungen die automatisierte Bearbeitung besonders schützenswerter Daten oder andere Bearbeitungen im Sinne von Artikel 34 Absatz 2 Buchstaben b und c bewilligen kann, noch bevor ein formelles Gesetz in Kraft tritt.
- Artikel 15 des Bundesgesetzes über den Einsatz elektronischer Mittel zur Erfüllung von Behördenaufgaben erlaubt die Durchführung von Pilotprojekten unter Einsatz von KI.
- Artikel 25h des Strassenverkehrsgesetzes ermächtigt das ASTRA, Ausnahmebewilligungen für Tests mit automatisierten Fahrzeugen zu erteilen.
- Artikel 23a des Bundesgesetzes über die Elektrizitätsversorgung sieht vor, dass das UVEK Pilotprojekte bewilligen kann, die auf die Entwicklung innovativer Technologien, Geschäftsmodelle oder Produkte im Energiesektor abzielen, soweit damit Erfahrungen für eine mögliche Gesetzesänderung gesammelt werden können.
- Darüber hinaus bestehen Pilotprojekte in der Krankenversicherung zur Eindämmung der Kostensteigerung, zur Verbesserung der Qualität oder zur Förderung der Digitalisierung, in der Invalidenversicherung zur Unterstützung der Wiedereingliederung sowie im Bereich der Berufsbildung.

Im Gegensatz dazu sind **Sandboxen im engeren Sinne**, die es Unternehmen erlauben würden, innovative Systeme unter realen oder simulierten Bedingungen ohne Sanktionen zu testen, **bislang weniger weit entwickelt**.

Welche Herausforderungen stellen sich?

Die Schweiz ist für ihre Kompetenzen in den Bereichen Forschung und Innovation, auch im Bereich der KI, anerkannt. Dank weltweit renommierter akademischer Institutionen und einem dynamischen wirtschaftlichen Geflecht aus Start-ups und Technologieunternehmen wird das Potenzial der KI gut ausgeschöpft und stärkt die Wettbewerbsfähigkeit und Attraktivität des Standorts Schweiz in zahlreichen Sektoren. Gleichwohl stellt sich die Frage, ob der schweizerische Rechtsrahmen ausreicht, um Innovationen im Bereich der KI zu begleiten, zu fördern und zu regulieren. Die Begriffe «regulatorische Sandboxen», «Pilotprojekte» und «Sandboxen im engeren Sinne» sind unscharf und schwer voneinander abzugrenzen. Auch die gesetzlichen Voraussetzungen für die Einrichtung der einen oder anderen Massnahme ist wenig klar. Diese Situation belässt den betroffenen Akteuren zwar einen gewissen Handlungsspielraum, der freieres Experimentieren ermöglicht, sie schafft jedoch zugleich rechtliche Unsicherheit, die Innovation hemmen und zu Missbräuchen führen kann.

Weitere Informationen:	
Rechtliche Basisanalyse des BJ (Deutsch / Französisch): Kapitel 4.2.3.3 und 4.3.2.8	**Erläuternder Bericht zur Konvention** (Englisch / Französisch): N 33 ff. und 90 ff.

9. Innovation et recherche

La Suisse dispose d'un environnement propice à l'innovation en matière d'IA, dont font pleinement usage les instituts de recherche et les entreprises. Quelles mesures pour encourager encore la recherche, tout en protégeant les droits fondamentaux et l'intérêt public ?

De quoi s'agit-il ?

L'IA doit pouvoir se développer dans un cadre juridique à la fois souple et protecteur. L'absence de cautèles peut conduire à une **insécurité juridique et à des abus** et mettre en péril les droits fondamentaux, l'État de droit et la démocratie. À l'inverse, une législation trop stricte et pas suffisamment agile ne permettra pas de développer la recherche et de **tirer parti des avantages sociétaux** et **économiques** liés à l'innovation.

Que prévoit la convention du Conseil de l'Europe ?

Sous réserve des articles 13 et 25, paragraphe 2, **la convention ne s'applique pas aux activités de recherche et de développement** relatives aux systèmes d'IA **qui n'ont pas encore été rendus disponibles à l'utilisation, si les essais ne sont pas susceptibles d'interférer** avec les droits de l'homme, la démocratie et l'État de droit. En d'autres termes, un système d'IA purement expérimental n'est pas concerné par les principes énoncés dans la convention. Les systèmes d'IA qui sont **rendus disponibles à l'utilisation** à la suite des activités de recherche et de développement devront en revanche en principe se conformer à la convention, y compris en ce qui concerne leur conception et leur développement (art. 3, par. 2).

L'article 13 appelle les États à permettre, le cas échéant, **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, afin de **favoriser l'innovation tout en évitant les impacts négatifs sur les droits de l'homme**. La convention laisse aux États 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* »)[14], de recommandations, ou de lettres de non-intervention pour clarifier comment les régulateurs aborderont les différents cycles de vie de l'IA.

Quelle est la situation actuelle en Suisse ?

Les **entreprises suisses et les instituts de recherches en Suisse sont très actifs** dans la recherche en matière d'IA. **De nombreux pôles d'innovation ont été créés** dans des secteurs variés tels que la finance, la recherche, la santé ou l'industrie.

La **législation suisse contient plusieurs dispositions légales permettant la mise en place de projets pilotes susceptibles de s'appliquer à l'IA**. On citera par exemple :

[14] D'après la définition retenue ici dans l'analyse juridique de l'OFJ, les bacs à sable réglementaires contiennent d'une part les projets-pilotes et d'autre part les bacs à sable au sens strict : *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. *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. Les essais dans le cadre de bacs à sables peuvent se faire en conditions réelles ou dans un cadre simulé et/ou contrôlé.

- l'article 35 de la loi fédérale sur la protection des données qui 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 à certaines conditions.
- l'article 15 de la loi sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités, qui permet aussi de réaliser des projets pilotes susceptibles de s'appliquer à l'IA.
- l'article 25h de la loi sur la circulation routière, qui autorise l'OFROU à octroyer des autorisations exceptionnelles pour effectuer des tests avec des véhicules automatisés.
- l'article 23a de loi fédérale sur l'approvisionnement en électricité, qui prévoit que le DETEC peut autoriser des projets pilotes visant le développement de technologies, de modèles d'affaires ou de produits innovants dans le secteur de l'énergie dans la mesure où ils permettent de recueillir des expériences en vue d'une modification de la loi.
- On trouve aussi des projets pilotes dans les domaines de l'assurance-maladie afin de freiner l'augmentation des coûts de la santé et de renforcer la qualité ou de promouvoir la numérisation, de l'assurance-invalidité pour faciliter la réadaptation ou encore dans la formation professionnelle.

En revanche, **les bacs à sable au sens strict**, qui permettraient à des entreprises de tester des systèmes innovants dans des conditions réelles ou simulées sans être sanctionnées, **sont peu développés**.

Quels sont les enjeux ?

La Suisse est reconnue pour ses compétences en matière d'innovation et de recherche, aussi en matière d'IA. Grâce à des institutions académiques de renommée mondiale et à un tissu économique dynamique de start-ups et d'entreprises technologiques, le potentiel de l'IA est bien exploité et renforce la compétitivité et l'attractivité de la place suisse dans de nombreux secteurs. Néanmoins, la question de savoir si le cadre législatif suisse est suffisant pour accompagner, encourager et encadrer l'innovation dans le domaine de l'IA se pose. Les notions de « bacs à sable réglementaires », de « projets-pilotes » et de « bacs à sable au sens strict » sont floues et difficiles à délimiter les unes des autres. Les exigences légales pour l'institution de telle ou telle mesure sont peu claires. Cette situation laisse certes une marge de manœuvre aux acteurs concernés qui peuvent expérimenter plus librement, mais elle créé aussi une insécurité juridique susceptible de freiner l'innovation et de conduire à des abus.

Informations supplémentaires : **Analyse juridique de base de l'OFJ** (Allemand / Français) : chapitres 4.2.3.3 et 4.3.2.8	**Rapport explicatif de la convention** (Anglais / Français) : paragraphes 33 ss et 90 ss