University of Zurich[UZH]

Digital Society Initiative

Position Paper

A Legal Framework for Artificial Intelligence

Florent Thouvenin, Markus Christen, Abraham Bernstein, Nadja Braun Binder, Thomas Burri, Karsten Donnay, Lena Jäger, Mariela Jaffé, Michael Krauthammer, Melinda Lohmann, Anna Mätzener, Sophie Mützel, Liliane Obrecht, Nicole Ritter, Matthias Spielkamp, Stephanie Volz

This position paper was developed during a workshop held in Balsthal from 26–28 August 2021 and funded by the Strategy Lab of the Digital Society Initiative (DSI) at the University of Zurich. In addition to the authors of this paper, three representatives of the federal administration also participated in this workshop, namely Monique Cossali Sauvain (FOJ), Roger Dubach (FDFA) and Thomas Schneider (OFCOM). They represent Switzerland in the Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI).
Further information: dsi.uzh.ch/strategy-lab

The great technical advances in **artificial intelligence (AI)** and the use of these technologies in various areas raise fundamental questions about their impact on individuals and society. The term artificial intelligence sometimes evokes misleading associations and diffuse fears. From a technical perspective, it is an established collective term that encompasses a **range of technologies** that make automated decisions, recommendations, conclusions or predictions. AI includes knowledge-based systems, statistical methods and machine learning approaches (e.g., using neural networks). The high performance of these technologies is mainly based on the combination of a large number of mathematical optimizations that extract structures from significant amounts of data using large computing capacities.

To avoid misleading associations, we do not use the term AI in this position paper but rather speak of **"algorithmic systems"**. This term does not refer to specific current or future technologies but to applying **these technologies in a social context**. The need for legal coverage only arises when technologies are used and affect individuals and/or society. The term "algorithmic systems" also allows us to cover applications with the same effects as artificial intelligence but based on other technologies.

When considering the need for regulation, it should be noted that using algorithmic systems does **not generally lead to entirely new challenges**. That is, some of them exist even if no algorithmic systems are used. Decisions are made by people, and the challenges only become more visible when using these systems. However, other challenges take on a new quality and dimension by using such systems. For example, certain forms of behavioural influence can be used much more efficiently—both in terms of precision (e.g., personalization) and quantity (scaling).

The **European Commission** published a proposal for a Regulation on Artificial Intelligence ("AI Act") on April 21, 2021[1], which will now be submitted to the Parliament and the Council of Ministers. The Council of Europe has adopted the first recommendation on AI[2]

1 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI Act) and amending certain Union acts, COM(2021) 206 final.

2 Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies) https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1154

and established an Ad hoc Committee on Artificial Intelligence (CAHAI) to study the feasibility and possible elements of a legal framework for AI development, design, and application. Switzerland is not bound by the EU's requirements, and it is currently still open as to whether it will sign a possible Council of Europe convention. Nevertheless, any Council of Europe requirements will give member states discretion to design their national solutions, and **Switzerland should use this discretion to develop its own approach**. In the process, Switzerland will decide in detail which aspects of EU law will be adopted and where it will deliberately deviate from EU law to benefit its individuals, economy and society.

This position paper sets out **the approaches that should be taken to the legal coverage of algorithmic systems** in Switzerland, the issues that require particular attention, and how Switzerland should position itself in the context of European regulatory trends.

The discussion has a practical and strategic urgency because algorithmic systems have an increasing influence on private and public life, infrastructures for algorithmic systems are increasingly being created in Switzerland and abroad, and the European and international environment is increasingly turning to the regulation of these systems, which will inevitably have an impact on Switzerland.

Regulatory Goals

Regulatory coverage of the challenges of using algorithmic systems serves two equally important goals. First, the regulation should leave **as much room as possible for developing and using algorithmic systems** that benefit individuals and society. Second, it must also ensure that the individuals affected by the use of algorithmic systems and society as a whole do **not suffer any disadvantages** from these uses (i.e., affected individuals are not discriminated against, referendums are not manipulated and principles of the rule of law are not undermined).

Regulatory Approach

The use of algorithmic systems leads to various **challenges** that must be addressed using the law; the focus is on **five areas**: recognisability and comprehensibility, discrimination, manipulation, liability, and data protection and data security.

The challenges posed by algorithmic systems are manifold and often have a new dimension or quality, but they are not unique to such systems. Therefore, these challenges should not be covered by a general "AI law" or an "algorithm law". Instead, a **combination of general and sector-specific standards is appropriate. The focus here is on the selective adaptation of existing laws**. After all, the legal system already contains standards that can address many of the challenges associated with algorithmic systems. However, in quite a few cases, it will probably be necessary to **adapt the interpretation and application of existing standards** to meet the new challenges appropriately.

Given the multitude of manifestations of algorithmic systems, a **technology-neutral approach** that can grasp the challenges independent of a specific technology should be chosen. Due to the rapid pace of technological development, a regulation can only survive if it is not geared to a specific technology. This principle applies without restriction to the design of general standards. However, it does not exclude regulation focusing on a particular technology in specific sectors (e.g., medical devices, vehicles).

Regulatory Need

The use of algorithmic systems is generally associated with data processing. If this involves personal data, **data protection law** applies. However, the processing of personal data by algorithmic systems does not raise any fundamentally new questions. It, therefore, seems possible in principle to solve the challenges for the protection of privacy and data protection using existing data protection law.

However, the use of algorithmic systems also leads to further challenges. For example, such systems are often not **recognizable** to those affected, and their mode of operation is not **comprehensible**. In addition, such systems can **discriminate** against people and **manipulate** their thoughts and actions. Furthermore, algorithmic systems raise new **liability issues**. In all these areas, there is a need for regulation. This also applies to **ensuring the safety of autonomous systems** and to specific

approval procedures. Finally, the question arises about whether the use of certain, particularly problematic autonomous systems should be prohibited (at least for the time being).

Recognisability and comprehensibility

The use and functioning of algorithmic systems must be recognizable and comprehensible to affected persons. This transparency has several dimensions:

(1) Persons interacting with algorithmic systems must be able to recognize that they are doing so with such a system and not with a human being. This can be achieved by introducing **an obligation to label when using algorithmic systems**. Since the interaction of an algorithmic system with a person generally involves the processing of personal data, such an obligation to label could be provided for in the Data Protection Act.

(2) Persons who are affected in a relevant way by the decision of an algorithmic system must be able to **understand this decision**. This does not mean that the persons must understand the technical functioning of the systems in detail; rather, the comprehensibility must be appropriate to the addressee. The extent of comprehensibility also depends on the significance of the decision for the person concerned and the legal requirements (e.g., justification of court rulings or orders by authorities) in the specific context. Therefore, it must be ensured that the data subjects can understand the logic underlying an automated decision (particularly, the data used and the criteria relevant to the decision) and obtain the required information to challenge the decision if necessary. This information must be made easily accessible and understandable for laypersons.

(3) In addition to individual recognisability, **recognisability for the interested public** must be ensured in the case of government use of algorithmic systems. For this purpose, it would be conceivable to create a publicly accessible register showing the areas in which the public administration uses algorithmic systems. Such a register should, among other things, provide information on the type and origin of the data processed, the legal basis, the purpose and means of processing, the body responsible, the logic of the algorithmic system and the actors who have participated in the development of the system. This information should be easily accessible and prepared in a standardized format.

Discrimination

The task of algorithmic systems is often to make distinctions. These distinctions are problematic when people are **treated differently based on protected characteristics** such as origin, race, gender, age, language, social status, lifestyle, religious, ideological or political convictions, or physical, mental or psychological disabilities, without any objective reason, which can lead to discrimination. In algorithmic systems, discrimination can occur because they directly or indirectly use protected characteristics as decision parameters or they are trained with data that exhibit a bias. Thus, certain socially existing biases can be reproduced in predictions or decisions in algorithmic systems. In many cases, however, algorithmic systems make the discrimination visible in the first place. Thus, the use of such systems also opens up the possibility of taking action against discrimination.

The problem of discrimination goes far beyond algorithmic systems but becomes particularly evident through their use. Therefore, discrimination should be covered by rules that apply **regardless** of whether a human or a machine makes the discriminatory decision or action. In most cases, the current legal situation in Switzerland only prohibits discrimination by state actors. However, many algorithmic systems are used by private parties, for example, in granting loans or selecting job applications. These discriminations could be prevented by a **general equal treatment law** that covers and sanctions discrimination by private parties, especially companies, based on specific protected characteristics.

It is often difficult to prove discrimination, and this problem could be solved by **reversing the burden of proof**. The person allegedly discriminated against would only have to provide sufficient prima facie evi-

dence of discrimination. The company would then have to prove that the decision was not based on a protected characteristic. The use of algorithmic systems may also prove advantageous in this context because—unlike in the case of human decisions—it is generally possible to identify the criteria used for the decision and prove that a decision is not based on protected characteristics.

Manipulation

Algorithmic systems can influence the thoughts and actions of people who interact with such systems. Typical examples are displaying particular targeted content, suppressing other relevant content and personalizing offers or prices on social media. However, the targeted influencing of a person's thoughts and actions by a third party (manipulation) is a widespread phenomenon, for example, in advertising. Influence by third parties is **always a restriction on the autonomy** of the person concerned. However, the nature and extent of the influence are highly variable, and in many cases, influence is unproblematic. This applies, for example, if the influence is unspecific and recognizable to the person concerned, as in the case of traditional forms of political and commercial advertising.

In the legal identification of problematic forms of manipulation, a distinction must be made between the decisions and actions of individuals in their roles as consumers and as citizens:

(1) In **manipulating citizens** in the context of democratic processes, the protection of **democratic will formation** is paramount. Algorithmic systems can endanger this because they allow particularly efficient and hardly recognizable forms of dissemination of one-sided information, exaggeration and lies. In addition, it is possible to display individualized content to individuals (or small groups) to influence their thinking, opinion-forming and voting behaviour specifically. This individualization of content can mean that certain statements do not even become the subject of public debate where they can be questioned and possibly refuted. **Freedom of information and expression** is of central importance in democratic decision-making. Ensuring that political actors and the population have a great deal of freedom in perceiving and disseminating information is central to the formation of public opinion. It should only be restricted with great restraint. Accordingly, the regulation of algorithmic systems should first and foremost aim to create transparency about the nature and extent of the dissemination of potentially questionable content (e.g., making known the criteria according to which Facebook displays content, suppresses it or identifies it as problematic), without evaluating the statements themselves. This evaluation must be left to the open-ended process of public opinion-forming. Users should also be able to recognize through appropriate measures how algorithmic systems individualize content to develop a sensitivity for how this influences them.

(2) In **manipulating consumers**, the **protection of individual freedom of choice** and the protection of **functioning competition** are of equal importance. Manipulation of consumers through the dissemination of false or misleading information is also of central importance. However, this type of manipulation can be covered by the applicable competition law (UWG). The situation is different for other forms of manipulation, such as the ongoing display of new content on social media platforms to keep consumers on the platform for as long as possible to show them as much advertising as possible. It should be examined here whether there is a need for action. In particular, this could be the case with vulnerable persons (e.g., addictive social media consumption by minors).

For both groups, manipulation does not necessarily have to be legally recorded as a process. Rather, it may be sufficient to create possibilities that allow **decisions to be reversed** if they have been made because of manipulation. For consumers, the introduction of rights of withdrawal would be conceivable, as they already exist today for door-to-door sales and telephone sales and—in the EU—also generally for so-called distance sales (especially e-commerce). In the case of votes, there is already the possibility of a challenge if the result has been

significantly influenced, for example, by the dissemination of false information.

Liability

A central challenge in the use of algorithmic systems is liability in the case of damage. Although the norms of general liability law also apply to such systems, proving that the prerequisites for **operators' liability** are associated with difficulties, especially in the case of fault. In certain sectors, strict liability rules that apply to algorithmic systems (e.g., for vehicles in the Road Traffic Act or drones in the Air Traffic Act) are already available. The introduction of general operator liability in the form of strict liability should be avoided. However, it should be examined whether **strict operator liability should be introduced for operators of algorithmic systems in other sectors**. A sector-specific approach would enable careful coordination with security regulations to be fulfilled ex ante.

The **liability of manufacturers** will then come to the fore. It is problematic that the Product Liability Act is tailored to conventional products and thus basically to physical objects placed on the market after their manufacture and can no longer be influenced by the manufacturers. The coverage of algorithmic systems by the **Product Liability Act** presupposes that such systems are recognized as products at all. Then the manufacturers should be liable for safe (further) developments of their products. At the same time, however, they must be able to exonerate themselves in the event of improper influence by other parties. The Swiss Product Liability Act must be updated accordingly.

Safety

Algorithmic systems must meet **common safety standards**, and they must be sufficiently robust and protected against harmful environmental influences and operating errors. In addition, sufficient protection against attacks must be ensured, whereby newer forms of attacks (e.g., manipulation of training data) must also be considered. The stringency of the requirements depends on the areas of application; for example, algorithmic systems that control processes in critical infrastructures (e.g., power supply) must meet stricter criteria than those that control a vacuum cleaner robot, for example.

Insofar as algorithmic systems process personal data, the provisions of data protection law are applicable, which require appropriate data security. However, these provisions are primarily aimed at protecting personal data and only indirectly cover the systems. Moreover, they do not apply if algorithmic systems do not process personal data, which may be the case, especially in critical infrastructures. It should therefore be examined whether the introduction of a **general IT security law** is necessary. As an alternative to state regulation of specific security requirements, the general binding nature of standards developed by standardization organizations could be considered.

Approval procedures

Already today, some products may only be brought to market after approval by a government authority (e.g., vehicles or medical devices). These approval procedures must also be followed when products use algorithmic systems.

In the **existing approval procedures**, the relevant prerequisites and procedures must be adapted to guarantee the required safety and quality of the products, even if they are based on the use of algorithmic systems. It should be noted that algorithmic systems can be further developed after approval or can even develop themselves further (through machine learning). In these cases, it must be ensured that the approval is reviewed again at each appropriate development step (life cycle regulation).

It should also be examined whether **new approval procedures** need to be created to ensure the safety of risky products or services that use algorithmic systems. The focus here is on systems that interact with their environment (e.g., care or cleaning robots and toys). On the other hand, predictive instruments used in sensitive areas, such as law enforcement or crime prevention, could also be subject to approval. For less risky products, certification could also be envisaged.

Prohibited applications

Finally, it should be examined whether specific applications of algorithmic systems should be banned because

they lead (or can lead) to restrictions on fundamental rights that should not be accepted. As an alternative to a **ban**, a **moratorium** on using specific algorithmic systems could also be enacted. Such a moratorium would make it possible to examine more closely the medium- and long-term consequences of algorithmic systems in critical areas and decide only later whether the use of such systems should be permitted. From today's perspective, the following applications are in the foreground:

- The use of **facial recognition and other remote biometric recognition procedures** in public spaces, insofar as there is a risk that these algorithmic systems will be used for mass surveillance;
- The use of **social scoring** to regulate access to basic resources (government services, credit, social security, etc.).

Given rapid technological developments, it should also be regularly evaluated whether new forms of algorithmic systems (e.g., for the autonomous exercise of lethal force in the security sector) should also be prohibited.

Switzerland's position in the international context

Work is currently underway in various jurisdictions (EU, USA, China) on the regulation of algorithmic systems. The developments in the EU and the Council of Europe are particularly relevant for Switzerland. Switzerland should **not strive for a passive adoption of these regulatory approaches**. Instead, it should develop its own position based on the principles formulated in this position paper and actively introduce it into the international and, in particular, European discourse together with international partners with similar ideas. In doing so, the coherence of domestic and foreign policy should be maintained, and the active discourse should be reflected in domestic policy, too.

Swiss companies that want to offer or use autonomous systems on the **European market** will have to comply with the future requirements of EU law. However, this does not mean that Switzerland should adopt these requirements in its national law. Rather, it seems sensible to create room to manoeuvre for those Swiss companies that do not (yet) want to offer their products on the European market by providing a sufficiently open legal framework (e.g., by a general prohibition of discrimination instead of specific requirements on risk management and data quality).

Next steps

This position paper shows that there is a need for action in Switzerland. The challenges associated with the use of algorithmic systems by companies and the state are sufficiently clear. Against this background and with a view to developments abroad, **Switzerland should promptly begin to develop norms** that can adequately address the challenges outlined. This work should be undertaken by a broad-based, **interdisciplinary commission of experts**. In many areas, there is still a **need for research**, for example, in the field of manipulation. The necessary research work should be continued with high intensity parallel to the work of a commission of experts to ensure that Switzerland's regulation can be based on secure scientific foundations.

University of Zurich[UZH]

Digital Society Initiative

Positionspapier

Ein Rechtsrahmen für Künstliche Intelligenz

Florent Thouvenin, Markus Christen, Abraham Bernstein, Nadja Braun Binder, Thomas Burri, Karsten Donnay, Lena Jäger, Mariela Jaffé, Michael Krauthammer, Melinda Lohmann, Anna Mätzener, Sophie Mützel, Liliane Obrecht, Nicole Ritter, Matthias Spielkamp, Stephanie Volz

Dieses Positionspapier wurde im Rahmen eines Workshops erarbeitet, der vom 26.–28. August 2021 in Balsthal durchgeführt und vom Strategy Lab der Digital Society Initiative (DSI) der Universität Zürich finanziert wurde. Neben den Autor*innen dieses Papiers haben auch drei Vertreter*innen der Bundesverwaltung an diesem Workshop teilgenommen, nämlich Monique Cossali Sauvain (BJ), Roger Dubach (EDA) und Thomas Schneider (BAKOM). Sie vertreten die Schweiz im Ad Hoc Komitee des Europarates zu Künstlicher Intelligenz (CAHAI). Weitere Informationen: dsi.uzh.ch/strategy-lab

Die grossen technischen Fortschritte im Bereich der **Künstlichen Intelligenz (KI)** und der Einsatz dieser Technologien in einer Vielzahl von Bereichen werfen grundlegende Fragen zu den Auswirkungen auf Individuen und die Gesellschaft auf. Der Begriff der Künstlichen Intelligenz weckt bisweilen irreführende Assoziationen und diffuse Ängste. Aus technischer Perspektive handelt es sich um einen etablierten Sammelbegriff, der **eine Reihe von Technologien** umfasst, die automatisierte Entscheidungen fällen, Empfehlungen machen, Schlussfolgerungen ziehen oder Vorhersagen treffen. Dazu gehören wissensbasierte Systeme und statistische Methoden ebenso wie Ansätze des maschinellen Lernens (z.B. unter Einsatz neuronaler Netze). Die grosse Leistungsfähigkeit dieser Technologien basiert meist auf der Aneinanderreihung einer Vielzahl von mathematischen Optimierungen, die unter Nutzung grosser Rechnerkapazitäten Strukturen aus grossen Datenmengen extrahieren.

Um irreführende Assoziationen zu vermeiden, verwenden wir in diesem Positionspapier nicht den Begriff der Künstlichen Intelligenz (KI), sondern sprechen von «**algorithmischen Systemen**». Damit werden nicht bestimmte heutige oder künftige Technologien bezeichnet, sondern es wird auf die **Anwendung dieser Technologien in einem sozialen Kontext** verwiesen. Denn Bedarf nach einer rechtlichen Erfassung entsteht erst, wenn Technologien eingesetzt werden und Wirkung für Individuen und/oder die Gesellschaft entfalten. Der Begriff der algorithmischen Systeme erlaubt zudem, auch Anwendungen zu erfassen, die gleiche Wirkungen entfalten wie Künstliche Intelligenz, aber auf anderen Technologien beruhen.

Bei der Frage nach dem Regelungsbedarf ist zu beachten, dass der Einsatz von algorithmischen Systemen in der Regel **nicht zu völlig neuen Herausforderungen führt**. Einige davon bestehen auch, wenn keine algorithmischen Systeme verwendet werden, sondern Entscheide von Menschen getroffen werden – sie werden durch die Nutzung dieser Systeme nur besser sichtbar. Andere Herausforderungen wiederum erhalten durch die Nutzung solcher Systeme eine neue Qualität und Dimension, weil bspw. bestimmte Formen der Verhaltensbeeinflussung viel effizienter genutzt werden können – sowohl bezüglich der Präzision (z.B. zur Personalisierung) als auch hinsichtlich der Quantität (Skalierung).

Die **Europäische Kommission** hat am 21. April 2021 einen Vorschlag für eine Verordnung über Künst-

liche Intelligenz («AI Act») veröffentlicht[1], der nun Parlament und Ministerrat vorgelegt wird. Der **Europarat** hat eine erste Empfehlung zu Künstlicher Intelligenz verabschiedet[2] und einen Expert*innenausschuss (Ad hoc Committee on Artificial Intelligence, CAHAI) eingesetzt, der die Machbarkeit und mögliche Elemente eines Rechtsrahmens für die Entwicklung, Gestaltung und Anwendung von KI untersucht. Die Schweiz ist nicht an die Vorgaben der EU gebunden und es ist derzeit noch offen, ob sie eine allfällige Konvention des Europarates unterzeichnen wird. Absehbar ist immerhin, dass allfällige Vorgaben des Europarates den Mitgliedstaaten viel Freiraum bei der Ausgestaltung ihrer nationalen Lösungen lassen werden. **Die Schweiz sollte diesen Freiraum nutzen, um einen eigenen Ansatz zu entwickeln**. Dabei wird im Einzelnen zu entscheiden sein, welche Ansätze des EU-Rechts übernommen werden und wo die Schweiz zum Nutzen von betroffenen Personen, Wirtschaft und Gesellschaft vom EU-Recht bewusst abweichen sollte.

Dieses Positionspapier legt dar, welche **Ansätze zur rechtlichen Erfassung algorithmischer Systeme** in der Schweiz verfolgt werden sollten, welche Fragen besondere Beachtung erfordern und wie sich die Schweiz im Umfeld der europäischen Regulierungstendenzen positionieren soll.

Die Diskussion hat eine praktische und strategische Dringlichkeit, weil algorithmische Systeme zunehmend Einfluss auf das private und öffentliche Leben haben, in der Schweiz und im Ausland vermehrt Infrastrukturen für algorithmische Systeme geschaffen werden und sich das europäische und internationale Umfeld zunehmend der Regulierung dieser Systeme zuwendet, was unvermeidlich einen Einfluss auf die Schweiz nach sich ziehen wird.

1 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final.

2 Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies) https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1154

Regelungsziele

Die rechtliche Erfassung der Herausforderungen des Einsatzes von algorithmischen Systemen dient zwei gleichwertigen Zielen: Zum einen soll die Regelung möglichst viel **Raum für die Entwicklung und Verwendung von algorithmischen Systemen** belassen, die für Einzelne und die Gesellschaft einen Nutzen schaffen. Zum andern ist sicherzustellen, dass die von der Verwendung von algorithmischen Systemen betroffenen Personen und die Gesellschaft als Ganzes aus diesen Verwendungen **keine Nachteile** erleiden, dass also bspw. betroffene Personen nicht diskriminiert, Volksabstimmungen nicht manipuliert und rechtsstaatliche Prinzipien nicht untergraben werden.

Regelungsansatz

Der Einsatz von algorithmischen Systemen führt zu vielfältigen **Herausforderungen**, die mit den Mitteln des Rechts zu erfassen sind; im Vordergrund stehen **fünf Bereiche**: Erkennbarkeit und Nachvollziehbarkeit, Diskriminierung, Manipulation, Haftung sowie Datenschutz und Datensicherheit.

Die Herausforderungen, die durch algorithmische Systeme entstehen, sind vielfältig und weisen oft eine neue Dimension oder Qualität auf, sie bestehen aber nicht nur beim Einsatz solcher Systeme. Diese Herausforderungen sollten deshalb nicht durch ein generelles «KI-Gesetz» oder ein «Algorithmen-Gesetz» erfasst werden. Angezeigt ist vielmehr eine **Kombination von allgemeinen und sektorspezifischen Normen**. Dabei steht die **punktuelle Anpassung bestehender Gesetze** im Vordergrund. Denn die Rechtsordnung enthält bereits Normen, die in der Lage sind, viele der Herausforderungen zu erfassen, die mit dem Einsatz von algorithmischen Systemen verbunden sind. Allerdings dürfte es in nicht wenigen Fällen erforderlich sein, **die Auslegung und Anwendung bestehender Normen anzupassen**, um den neuen Herausforderungen angemessen begegnen zu können.

Angesichts der Vielzahl der Erscheinungsformen algorithmischer Systeme ist grundsätzlich ein **technologieneutraler Ansatz** zu wählen, der die Herausforderungen unabhängig von der Verwendung einer bestimmten Technologie zu erfassen vermag. Wegen der

raschen technischen Entwicklung kann eine Regelung nur Bestand haben, wenn sie nicht auf eine bestimmte Technologie ausgerichtet ist. Dieser Grundsatz gilt ohne Einschränkung für die Ausgestaltung allgemeiner Normen. Er schliesst aber die Fokussierung der Regulierung auf eine konkrete Technologie in spezifischen Sektoren (z.B. Medizinprodukte, Fahrzeuge) nicht aus.

Regelungsbedarf

Der Einsatz von algorithmischen Systemen ist in der Regel mit der Verarbeitung von Daten verbunden. Handelt es sich dabei um Personendaten, greift das **Datenschutzrecht**. Die Bearbeitung von Personendaten durch algorithmische Systeme wirft allerdings keine grundlegend neuen Fragen auf. Es erscheint deshalb grundsätzlich möglich, die Herausforderungen für den Schutz der Privatsphäre und den Datenschutz mit den Mitteln des bestehenden Datenschutzrechts zu lösen.

Die Nutzung algorithmischer Systeme führt allerdings auch zu weiteren Herausforderungen. So ist der Einsatz solcher Systeme für die Betroffenen oft nicht **erkennbar** und ihre Funktionsweise nicht **nachvollziehbar**. Zudem können solche Systeme Menschen **diskriminieren** und in ihrem Denken und Handeln **manipulieren**. Ausserdem wirft der Einsatz algorithmischer Systeme neue **Haftungsfragen** auf. In allen diesen Bereichen besteht Regelungsbedarf. Das gilt auch für die **Gewährleistung der Sicherheit autonomer Systeme** und für bestimmte **Zulassungsverfahren**. Schliesslich fragt sich, ob der Einsatz von bestimmten, besonders problematischen autonomen Systemen (zumindest einstweilen) verboten werden sollte.

Erkennbarkeit und Nachvollziehbarkeit

Der Einsatz von algorithmischen Systemen und deren Funktionsweise muss für die betroffenen Personen erkennbar und verständlich sein. Diese Transparenz hat mehrere Dimensionen:

(1) Personen, die mit algorithmischen Systemen interagieren, müssen erkennen können, dass sie dies mit einem solchen System und nicht mit einem Menschen tun. Dies kann durch Einführung einer **Kennzeichnungspflicht beim Einsatz von algorithmischen Systemen** erreicht werden. Da die Interaktion eines algorithmischen Systems mit einer Person in aller Regel mit der Bearbeitung von Personendaten verbunden ist, könnte eine solche Kennzeichnungspflicht im Datenschutzgesetz vorgesehen werden.

(2) Personen, die von der Entscheidung eines algorithmischen Systems in relevanter Weise betroffen sind, müssen diese **Entscheidung nachvollziehen** können. Das bedeutet nicht, dass die Personen die technische Funktionsweise der Systeme im Einzelnen verstehen müssen; die Nachvollziehbarkeit muss vielmehr adressatengerecht sein. Der Umfang der Nachvollziehbarkeit hängt zudem von der Bedeutung der Entscheidung für die betroffene Person und den rechtlichen Anforderungen (z.B. Begründung von Urteilen von Gerichten oder Verfügungen von Behörden) im konkreten Kontext ab. Sicherzustellen ist deshalb, dass die betroffenen Personen in der Lage sind, die einer automatisierten Entscheidung zugrundeliegende Logik (insb. verwendete Daten und für die Entscheidung relevante Kriterien) zu verstehen und die notwendigen Informationen zu erhalten, um die Entscheidung gegebenenfalls anzufechten. Diese Informationen müssen leicht zugänglich und für Laien verständlich verfügbar gemacht werden.

(3) Zusätzlich zur individuellen **Erkennbarkeit** ist beim staatlichen Einsatz algorithmischer Systeme die Erkennbarkeit für die **interessierte Öffentlichkeit** sicherzustellen. Denkbar wäre, hierzu ein öffentlich zugängliches Register zu schaffen, aus dem ersichtlich wird, in welchen Bereichen die öffentliche Verwaltung algorithmische Systeme einsetzt. Ein solches Register sollte unter anderem Auskunft geben über die Art und Herkunft der bearbeiteten Daten, die Rechtsgrundlage, den Zweck und die Mittel der Bearbeitung, das verantwortliche Organ, die Logik des algorithmischen Systems und die Akteure, die an der Entwicklung des Systems mitgewirkt haben. Diese Informationen sollten leicht zugänglich sein und in einem standardisierten Format aufbereitet werden.

Diskriminierung

Die Aufgabe von algorithmischen Systemen besteht oft darin, Unterscheidungen zu treffen. Diese Unterscheidungen sind dann problematisch, wenn **Personen aufgrund von geschützten Merkmalen** wie Herkunft, Rasse, Geschlecht, Alter, Sprache, soziale Stellung, Lebensform, religiöse, weltanschauliche oder politische Überzeugung oder körperliche, geistige oder psychische Behinderung **unterschiedlich behandelt** werden, ohne dass dafür ein sachlicher Grund besteht. In diesem Fall liegt eine Diskriminierung vor. Bei algorithmischen Systemen können Diskriminierungen namentlich vorkommen, weil sie direkt oder indirekt geschützte Merkmale als Entscheidungsparameter verwenden oder weil sie mit Daten trainiert werden, die einen «bias» aufweisen. So können bestimmte gesellschaftlich existierende Vorurteile in Prognosen oder Entscheidungen solcher Systeme reproduziert werden. In vielen Fällen macht der Einsatz von algorithmischen Systemen die Diskriminierung aber erst sichtbar. Damit eröffnet der Einsatz solcher Systeme auch die Möglichkeit, gegen Diskriminierungen vorzugehen.

Die Problematik der Diskriminierung geht weit über den Einsatz algorithmischer Systeme hinaus, wird durch deren Einsatz aber besonders deutlich. Das Problem der Diskriminierung sollte deshalb durch Regeln erfasst werden, die **unabhängig** davon greifen, ob die diskriminierende Entscheidung oder Handlung von einem Menschen oder einer Maschine vorgenommen wird. Die aktuelle Rechtslage in der Schweiz verbietet in den meisten Fällen nur die Diskriminierung durch staatliche Akteure. Doch viele algorithmische Systeme werden von Privaten eingesetzt, etwa bei der Kreditvergabe oder bei der Selektion von Bewerbungen. Diese Diskriminierungen könnten durch ein **allgemeines Gleichbehandlungsgesetz** verhindert werden, das Diskriminierungen durch Private, insb. Unternehmen, aufgrund bestimmter, geschützter Merkmale erfasst und sanktioniert.

Der Nachweis einer Diskriminierung ist oft schwer zu erbringen. Dieses Problem könnte durch eine **Beweislastumkehr** gelöst werden. Die angeblich diskriminierte Person müsste das Vorliegen einer Diskriminierung nur hinreichend glaubhaft machen und das Unternehmen müsste dann den Nachweis erbringen, dass die Entscheidung nicht auf einem geschützten Merkmal beruht. Der Einsatz algorithmischer Systeme kann sich dabei auch als vorteilhaft erweisen, weil es – anders als bei menschlichen Entscheidungen – grundsätzlich möglich ist, die für die Entscheidung genutzten Kriterien zu erkennen und den Nachweis zu erbringen, dass eine Entscheidung nicht auf geschützte Merkmale abstellt.

Manipulation

Algorithmische Systeme können das Denken und Handeln von Personen beeinflussen, die mit solchen Systemen interagieren. Typische Beispiele sind das Anzeigen bestimmter und das Unterdrücken anderer relevanter Inhalte auf Social Media und die Personalisierung von Angeboten oder Preisen. Die zielgerichtete Beeinflussung des Denkens und Handelns einer Person durch einen Dritten (Manipulation) ist allerdings ein weit verbreitetes Phänomen, etwa bei Werbung. Die Beeinflussung durch Dritte ist zwar stets ein **Eingriff in die Autonomie** der betroffenen Person. Art und Ausmass der Beeinflussung sind aber höchst unterschiedlich und in vielen Fällen ist eine Beeinflussung unproblematisch. Das gilt beispielsweise dann, wenn eine Beeinflussung unspezifisch und für die betroffene Person erkennbar ist, wie etwa bei den traditionellen Formen der politischen und kommerziellen Werbung.

Bei der rechtlichen Erfassung von problematischen Formen der Manipulation ist insbesondere zwischen Entscheidungen bzw. Handlungen von Individuen in ihren Rollen als Konsument*innen und als Staatsbürger*innen zu unterscheiden:

(1) **Bei der Manipulation von Staatsbürger*innen** im Kontext demokratischer Prozesse steht der Schutz der **demokratischen Willensbildung** im Vordergrund. Diese kann beim Einsatz von algorithmischen Systemen gefährdet sein, weil solche Systeme besonders effiziente und kaum erkennbare Formen der Verbreitung von einseitiger Information, Übertreibung und Lüge erlauben. Zudem ist es möglich, einzelnen Personen (oder kleinen Gruppen) individualisierte Inhalte anzuzeigen, um ihr Denken, ihre Meinungsbildung und ihr Stimmverhalten gezielt zu beeinflussen. Diese Individualisierung der In-

halte kann dazu führen, dass bestimmte Aussagen gar nicht zum Gegenstand der öffentlichen Debatte werden, in der sie in Frage gestellt und gegebenenfalls widerlegt werden können. Bei der demokratischen Willensbildung kommt der **Informations- und Meinungsfreiheit** zentrale Bedeutung zu. Diese sichert politischen Akteuren und der Bevölkerung einen grossen Freiraum beim Wahrnehmen und Verbreiten von Informationen, der für die öffentliche Meinungsbildung zentral ist und nur sehr zurückhaltend eingeschränkt werden darf. Entsprechend sollte die Regulierung algorithmischer Systeme vorab das Schaffen von Transparenz über Art und Ausmass der Verbreitung von allfällig fragwürdigen Inhalten zum Ziel haben (z.B. das Bekanntmachen der Kriterien, nach denen Facebook Inhalte anzeigt, unterdrückt oder als problematisch kenntlich macht), ohne die Aussagen selbst zu bewerten. Diese Bewertung muss dem ergebnisoffenen Prozess der öffentlichen Meinungsbildung überlassen bleiben. Nutzer*innen sollen zudem durch geeignete Massnahmen erkennen können, wie Inhalte durch algorithmische Systeme individualisiert werden, um eine Sensibilität dafür zu entwickeln, wie sie dadurch beeinflusst werden.

(2) Bei der **Manipulation von Konsument*innen** stehen der Schutz der **individuellen Entscheidungsfreiheit** und der Schutz des **funktionierenden Wettbewerbs** gleichgeordnet nebeneinander. Auch bei Konsument*innen kommt der Manipulation durch die Verbreitung von falschen oder irreführenden Informationen zentrale Bedeutung zu. Diese Art der Manipulation kann allerdings mit dem geltenden Wettbewerbsrecht (UWG) erfasst werden. Anderes gilt bei anderen Formen der Manipulation, bspw. beim laufenden Anzeigen neuer Inhalte auf Social-Media-Plattformen mit dem Ziel, die Konsument*innen möglichst lange auf der Plattform zu halten, um ihnen möglichst viel Werbung anzeigen zu können. Hier ist zu prüfen, ob Handlungsbedarf besteht. Dies könnte insbesondere bei vulnerablen Personen der Fall sein (z.B. bei suchtartigem Social Media Konsum von Minderjährigen).

Für beide Gruppen muss Manipulation nicht notwendigerweise als Vorgang rechtlich erfasst werden. Vielmehr kann es ausreichen, Möglichkeiten zu schaffen, die es erlauben, **Entscheidungen rückgängig zu machen**, wenn sie aufgrund einer Manipulation erfolgt sind. Für Konsument*innen wäre insb. die Einführung von Widerrufsrechten denkbar, wie sie schon heute bei Haustürgeschäften und Telefonverkäufen und – in der EU – auch allgemein beim sog. Fernabsatz (insb. E-Commerce) bestehen. Bei Abstimmungen besteht zudem schon heute die Möglichkeit der Anfechtung, wenn das Ergebnis bspw. durch die Verbreitung von falschen Informationen massgeblich beeinflusst wurde.

Haftung

Eine zentrale Herausforderung beim Einsatz algorithmischer Systeme ist die Haftung im Fall eines Schadens. Zwar finden die Normen des allgemeinen Haftpflichtrechts auch auf solche Systeme Anwendung; der Nachweis der Voraussetzungen für die **Haftung von Betreiber*innen** ist allerdings mit Schwierigkeiten verbunden, insbesondere beim Verschulden. In bestimmten Sektoren stehen bereits verschuldensunabhängige Haftungsregeln zur Verfügung, die auch bei algorithmischen Systemen greifen (z.B. für Fahrzeuge im Strassenverkehrsgesetz oder für Drohnen im Luftverkehrsgesetz). Von der Einführung einer allgemeinen Betreiberhaftung in Form einer Gefährdungshaftung ist zwar abzusehen. Zu prüfen ist aber, ob in **weiteren Sektoren für Betreiber*innen algorithmischer Systeme eine verschuldensunabhängige Betreiberhaftung** eingeführt werden sollte. Ein sektorspezifisches Vorgehen würde eine behutsame Abstimmung mit ex ante zu erfüllenden Sicherheitsvorschriften ermöglichen.

In den Vordergrund rücken wird sodann die **Haftung der Hersteller*innen**. Als problematisch erweist sich, dass das Produktehaftungsgesetz auf herkömmliche Produkte und damit grundsätzlich auf physische Gegenstände zugeschnitten ist, die nach ihrer Herstellung in Verkehr gebracht werden und von den Hersteller*innen nicht mehr beeinflusst werden können. Die Erfassung von algorithmischen Systemen durch das **Produktehaftungsgesetz** setzt voraus, dass solche Systeme überhaupt als Produkte anerkannt werden.

Sodann sollten die Hersteller*innen für sichere (Weiter-)Entwicklungen ihrer Produkte haften. Gleichzeitig müssen sie sich jedoch bei unsachgemässer Einflussnahme anderer Beteiligter entlasten können. Das Schweizer Produktehaftungsgesetz ist entsprechend zu aktualisieren.

Sicherheit

Algorithmische Systeme müssen **gängigen Sicherheitsstandards** genügen. Sie müssen also ausreichend robust sowie vor schädlichen Umwelteinflüssen und Bedienungsfehlern geschützt sein. Zudem muss ein ausreichender Schutz gegen Angriffe gewährleistet sein, wobei auch neuere Formen von Angriffen (z.B. Manipulation von Trainingsdaten) zu beachten sind. Die Strenge der Anforderungen hängt von den Anwendungsbereichen ab; so müssen etwa algorithmische Systeme, die Prozesse in kritischen Infrastrukturen kontrollieren (z.B. Stromversorgung), strengeren Kriterien genügen als solche, die beispielsweise einen Staubsaugerroboter steuern.

Soweit algorithmische Systeme Personendaten bearbeiten, sind die Bestimmungen des Datenschutzrechts anwendbar, die eine angemessene Datensicherheit verlangen. Diese Bestimmungen zielen allerdings in erster Linie auf den Schutz von Personendaten und erfassen die Systeme nur indirekt. Zudem finden sie keine Anwendung, wenn algorithmische Systeme keine Personendaten bearbeiten, was gerade bei kritischen Infrastrukturen der Fall sein kann. Es ist deshalb zu prüfen, ob die Einführung eines allgemeinen **IT-Sicherheitsgesetzes** erforderlich ist. Als Alternative zu einer staatlichen Regulierung konkreter Sicherheitsanforderungen könnte sich die Allgemeinverbindlicherklärung von Standards aufdrängen, die von Standardisierungsorganisationen entwickelt werden.

Zulassungsverfahren

Bereits heute gibt es Produkte, die nur nach Zulassung durch eine staatliche Behörde auf den Markt gebracht werden dürfen (z.B. Fahrzeuge oder Medizinprodukte). Diese Zulassungsverfahren müssen auch dann durchlaufen werden, wenn Produkte algorithmische Systeme verwenden.

Bei den **bestehenden Zulassungsverfahren** sind die relevanten Voraussetzungen und Verfahren so anzupassen, dass sie die erforderliche Sicherheit und Qualität der Produkte auch dann gewährleisten, wenn diese auf dem Einsatz algorithmischer Systeme beruhen. Dabei ist zu beachten, dass algorithmische Systeme nach der Zulassung weiterentwickelt werden können oder sich gar selbst weiterentwickeln können (durch maschinelles Lernen). In diesen Fällen muss sichergestellt werden, dass die Zulassung bei jedem relevanten Entwicklungsschritt erneut überprüft wird («*life cycle regulation*»).

Zudem ist zu prüfen, ob **neue Zulassungsverfahren** geschaffen werden müssen, um die Sicherheit von risikobehafteten Produkten oder Diensten zu gewährleisten, die algorithmische Systeme verwenden. Im Vordergrund stehen dabei Systeme, die mit ihrer Umwelt interagieren, bspw. Pflege- oder Putzroboter, aber auch Spielzeuge. Zum anderen könnten auch Prognoseinstrumente, die in sensiblen Bereichen, etwa in der Strafverfolgung oder bei der Kriminalprävention, eingesetzt werden, einer Zulassung unterstellt werden. Bei weniger risikobehafteten Produkten könnte auch eine Zertifizierung vorgesehen werden.

Verbotene Anwendungen

Schliesslich ist zu prüfen, ob bestimmte Anwendungen von algorithmischen Systemen zu verbieten sind, weil sie zu Eingriffen in Grundrechte führen (oder führen können), die nicht hingenommen werden sollten. Als Alternative zu einem **Verbot** könnte auch ein **Moratorium** für den Einsatz bestimmter algorithmischer Systeme erlassen werden. Ein solches Moratorium würde es ermöglichen, die mittel- und langfristigen Folgen des Einsatzes von algorithmischen Systemen in kritischen Bereichen näher zu untersuchen und erst später zu entscheiden, ob der Einsatz solcher Systeme zugelassen werden soll. Im Vordergrund stehen aus heutiger Sicht die folgenden Anwendungen:

– Der Einsatz von **Gesichtserkennung und anderen biometrischen Fernerkennungsverfahren** im öffentlichen Raum, sofern die Gefahr besteht, dass diese algorithmischen Systeme für eine Massenüberwachung eingesetzt werden;

– Der Einsatz von **Social Scoring** mit dem Ziel, den Zugang zu grundlegenden Ressourcen (staatliche Dienstleistungen, Kredite, soziale Sicherheit etc.) zu regulieren.

Mit Blick auf die rasche technische Entwicklung ist zudem regelmässig zu evaluieren, ob neue Formen der Nutzung von algorithmischen Systemen (z.B. zur autonomen Ausübung tödlicher Gewalt im Sicherheitsbereich) ebenfalls verboten werden sollten.

Position der Schweiz im internationalen Umfeld

Aktuell wird in verschiedenen Rechtsräumen (EU, USA, China) an der Regulierung von algorithmischen Systemen gearbeitet. Relevant für die Schweiz sind insbesondere die Entwicklungen in der EU und im Europarat. Die Schweiz sollte **keine passive Übernahme dieser Regulierungsansätze** anstreben. Vielmehr sollte sie – basierend auf den in diesem Positionspapier formulierten Grundsätzen – eine eigene Position erarbeiten und diese aktiv zusammen mit internationalen Partner*innen mit ähnlichen Vorstellungen in den internationalen und insbesondere europäischen Diskurs einbringen. Dabei sollte die Kohärenz von Innen- und Aussenpolitik gewahrt bleiben und der aktive Diskurs auch innenpolitisch gespiegelt werden.

Schweizer Unternehmen, die autonome Systeme auf dem **europäischen Markt** anbieten oder einsetzen wollen, werden sich an die künftigen Vorgaben des EU-Rechts halten müssen. Das bedeutet aber nicht, dass die Schweiz diese Vorgaben in ihr nationales Recht übernehmen sollte. Vielmehr scheint es sinnvoll, durch einen hinreichend offenen rechtlichen Rahmen (bspw. durch ein allgemeines Verbot von Diskriminierung statt durch spezifische Vorgaben zu Risikomanagement und Datenqualität) Spielraum für diejenigen Schweizer Unternehmen zu schaffen, die ihre Produkte (noch) nicht auf dem europäischen Markt anbieten wollen.

Weiteres Vorgehen

Dieses Positionspapier zeigt, dass in der Schweiz Handlungsbedarf besteht. Die mit dem Einsatz von algorithmischen Systemen durch Unternehmen und den Staat verbundenen Herausforderungen sind hinreichend deutlich erkennbar. Vor diesem Hintergrund und mit Blick auf die Entwicklungen im Ausland sollte die Schweiz **zeitnah mit der Erarbeitung von Normen** beginnen, welche die skizzierten Herausforderungen angemessen erfassen können. Diese Arbeit sollte von einer breit aufgestellten, interdisziplinär zusammengesetzten **Expert*innenkommission** übernommen werden. In vielen Bereichen besteht zudem noch **Forschungsbedarf**, bspw. im Bereich der Manipulation. Die erforderlichen Forschungsarbeiten sollten mit hoher Intensität parallel zur Arbeit einer Expert*innenkommission fortgeführt werden, um sicherzustellen, dass die Regelung der Schweiz auf gesicherten wissenschaftlichen Grundlagen aufbauen kann.

University of Zurich[UZH]

Digital Society Initiative

Prise de position

Un cadre juridique pour l'intelligence artificielle

Florent Thouvenin, Markus Christen, Abraham Bernstein, Nadja Braun Binder, Thomas Burri, Karsten Donnay, Lena Jäger, Mariela Jaffé, Michael Krauthammer, Melinda Lohmann, Anna Mätzener, Sophie Mützel, Liliane Obrecht, Nicole Ritter, Matthias Spielkamp, Stephanie Volz

Le présent document de prise de position a été élaboré dans le cadre d'un atelier qui s'est déroulé du 26 au 28 août 2021 à Balsthal et qui a été financé par le Strategy Lab de la Digital Society Initiative (DSI) de l'Université de Zurich. Outre les auteurs/-es du présent document, trois représentants/-es de l'administration fédérale ont également participé à cet atelier, à savoir Monique Cossali Sauvain (OFJ), Roger Dubach (DFAE) et Thomas Schneider (OFCOM). Ils représentent la Suisse dans le Comité ad hoc du Conseil de l'Europe sur l'intelligence artificielle (CAHAI).
Pour plus d'informations : dsi.uzh.ch/strategy-lab

Les importantes avancées techniques dans le domaine de **l'intelligence artificielle (IA)** et la mise en œuvre de ces technologies dans une multitude de domaines soulèvent des questions fondamentales sur leur impact sur les individus et la société. La notion d'intelligence artificielle suscite parfois des associations trompeuses et des peurs diffuses. D'un point de vue technique, l'intelligence artificielle est un terme générique établi englobant une **série de technologies** qui prennent des décisions automatiques, émettent des recommandations, tirent des conclusions ou réalisent des prévisions. Il s'agit notamment de systèmes à base de connaissances, de méthodes statistiques, ainsi que d'approches issues de l'apprentissage automatique (par ex. à l'aide de réseaux neuronaux). La performance considérable de ces technologies se fonde le plus souvent sur la succession d'une multitude d'optimisations mathématiques qui, à l'aide de grandes capacités de calcul, permettent d'extraire des structures à partir d'importantes quantités de données.

Afin d'éviter les associations trompeuses, nous n'utilisons pas la notion d'intelligence artificielle (IA) dans la présente prise de position, mais parlons de « **systèmes algorithmiques** ». Ainsi, nous ne désignons pas des technologies spécifiques actuelles ou futures, mais l'**application de ces technologies dans un contexte social**. En effet, le besoin d'une réglementation n'apparaît que lorsque les technologies sont mises en œuvre et ont un effet sur les individus et/ou la société. En outre, la notion de « systèmes algorithmiques » permet également de prendre en compte des applications qui ont le même effet que l'intelligence artificielle, mais qui reposent sur d'autres technologies.

En ce qui concerne la réglementation nécessaire dans ce contexte, il convient de noter que la mise en œuvre de systèmes algorithmiques **n'entraîne** généralement **pas de nouveaux défis**. En effet, certains d'entre eux se présentent même lorsque des décisions sont prises par l'homme, sans qu'un système algorithmique ne soit utilisé. Cependant, le recours à ces systèmes rend simplement ces décisions plus visibles. En revanche, d'autres défis acquièrent une nouvelle qualité et une nouvelle dimension lors de l'utilisation de ces systèmes, par exemple parce que certaines formes d'influence sur les comportements peuvent être utilisées de manière plus efficace, tant au niveau de la précision (par ex. avec la personnalisation) que de la quantité (mise à l'échelle).

Le 21 avril 2021, la **Commission européenne** a publié une proposition de règlement sur l'intelligence artificielle (« AI Act »)[1], qui sera désormais transmise au Parlement et au Comité des Ministres. Le **Conseil de l'Europe** a adopté une première recommandation sur l'intelligence artificielle[2] et a mis en place un comité d'experts/-es (Comité ad hoc sur l'intelligence artificielle, CAHAI) qui examine la faisabilité et les éléments possibles d'un cadre juridique pour le développement, la conception et l'application de l'IA. La Suisse n'est pas liée aux règles de l'UE et, actuellement, aucune décision n'a encore été prise quant à la signature d'une éventuelle convention du Conseil de l'Europe. On peut toutefois s'attendre à ce que les éventuelles règles du Conseil de l'Europe laissent une grande marge de manœuvre aux États membres lors de la conception de leurs solutions nationales. **La Suisse devrait profiter de cette marge de manœuvre pour développer sa propre approche**. À cet égard, il conviendra notamment de décider quelles approches du droit européen seront reprises et à quel niveau la Suisse devrait volontairement s'en écarter au profit des personnes concernées, de l'économie et de la société.

La présente prise de position explique quelles **approches relatives à la prise en compte juridique des systèmes algorithmiques** devraient être poursuivies en Suisse, à quelles questions il faut accorder une attention particulière et comment la Suisse doit se positionner dans l'environnement des tendances réglementaires européennes.

La discussion revêt une urgence pratique et stratégique, car les systèmes algorithmiques ont une influence croissante sur la vie privée et publique, davantage d'infrastructures sont créées pour des systèmes algorithmiques en Suisse et à l'étranger, et l'environnement européen et international se tourne toujours plus vers la réglementation de ces systèmes, ce qui entraîne inévitablement une influence sur la Suisse.

Objectifs législatifs

La prise en compte juridique des défis relatifs à la mise en œuvre de systèmes algorithmiques répond à deux objectifs similaires : d'une part, la réglementation doit laisser le plus d'**espace possible au développement et à l'utilisation des systèmes algorithmiques**, qui offrent des avantages pour les individus et la société. D'autre part, il convient de garantir que l'utilisation de systèmes algorithmiques **ne porte pas préjudice** aux personnes concernées et à la société dans son ensemble; il convient par exemple de veiller à ce que les personnes concernées ne soient pas discriminées, à ce que les votations populaires ne soient pas manipulées et à ce que les principes de l'État de droit ne soient pas renversés.

Approche réglementaire

La mise en œuvre de systèmes algorithmiques conduit à de multiples **défis** qui doivent être appréhendés par des moyens juridiques ; **cinq domaines** figurent au premier plan : identification et compréhension, discrimination, manipulation, responsabilité ainsi que protection et sécurité des données.

Les défis engendrés par les systèmes algorithmiques sont multiples et présentent souvent une dimension ou qualité inédite ; mais ceux-ci ne se présentent toutefois pas seulement lors de l'utilisation de tels systèmes. Aussi, ces défis ne devraient-ils pas être régis par une « loi IA » générale ou par une « loi sur les algorithmes ». Une **combinaison de normes générales et spécifiques au secteur** se révèle bien plus appropriée. À cet égard, **l'adaptation ponctuelle des lois existantes** reste l'objectif prioritaire. En effet, le système juridique comporte déjà des normes en mesure d'appréhender de nombreux défis liés à la mise en œuvre de systèmes algorithmiques. Toutefois, dans un certain nombre de cas, il sera probablement nécessaire **d'adapter l'interprétation et l'application de normes existantes** afin de pouvoir relever les nouveaux défis de manière appropriée.

Au vu de la multitude de formes de systèmes algorithmiques, il convient d'adopter une **approche tech-**

1 Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union européenne, COM(2021) 206 final.

2 Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies) https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1154

nologiquement neutre capable d'appréhender les défis indépendamment de la technologie spécifique employée. En raison de l'évolution technologique rapide, une réglementation ne peut perdurer que si elle n'est pas destinée à une technologie spécifique. Ce principe s'applique sans restriction à la conception de normes générales. Toutefois, il n'exclut pas une réglementation centrée sur une technologie concrète dans des secteurs spécifiques (par ex. produits médicaux, véhicules).

Nécessité d'une réglementation

Généralement, l'utilisation de systèmes algorithmiques est associée au traitement des données. Le **droit sur la protection des données** s'applique s'il s'agit de données à caractère personnel. Toutefois, le traitement de données à caractère personnel par le biais de systèmes algorithmiques ne suscite aucune question fondamentalement nouvelle. Ainsi, il paraît en principe possible de résoudre les défis liés à la protection de la sphère privée et des données avec les moyens du droit existant en matière de protection des données.

Néanmoins, l'utilisation de systèmes algorithmiques conduit aussi à d'autres enjeux. Ainsi, la mise en œuvre de ces systèmes n'est souvent pas **identifiable** pour les personnes concernées et leur fonctionnement n'est pas **compréhensible**. En outre, de tels systèmes peuvent mener à la **discrimination** d'individus et à la **manipulation** de leurs pensées et actions. De plus, la mise en œuvre de systèmes algorithmiques soulève de nouvelles **questions relatives à la responsabilité**. Le besoin d'une réglementation subsiste dans tous ces domaines, et c'est également le cas pour la **garantie de la sécurité des systèmes autonomes et certaines procédures d'autorisation**. Enfin, il convient de déterminer si l'utilisation de certains systèmes autonomes particulièrement problématiques doit être interdite (du moins temporairement).

Identification et compréhension

Les personnes concernées doivent identifier et comprendre l'utilisation de systèmes algorithmiques ainsi que leur fonctionnement. Cette transparence revêt plusieurs dimensions :

(1) Les personnes qui interagissent avec des systèmes algorithmiques doivent être en mesure de reconnaître qu'elles le font avec un tel système et non avec un être humain. Cette dimension peut être respectée en introduisant une **obligation de marquage lors de la mise en œuvre de systèmes algorithmiques**. L'interaction d'un système algorithmique avec une personne étant généralement liée au traitement de données à caractère personnel, cette obligation de marquage pourrait être prévue dans la loi sur la protection des données.

(2) Les personnes qui sont affectées de manière significative par une décision prise par un système algorithmique doivent être en mesure de **comprendre cette décision**. Cela ne signifie pas qu'elles doivent comprendre le fonctionnement technique des systèmes dans le détail, mais plutôt que leur intelligibilité doit être adaptée au destinataire. En outre, l'étendue de l'intelligibilité dépend de l'importance de la décision pour les personnes concernées et des exigences juridiques (par ex. justification de jugements émis par les tribunaux ou ordonnances édictées par les autorités) dans un contexte concret. Il convient donc de s'assurer que les personnes concernées sont en mesure de comprendre la logique sous-jacente à une décision automatisée (notamment les données utilisées et les critères pertinents à la prise de décision) et d'obtenir les informations nécessaires afin, le cas échéant, de contester la décision. Ces informations doivent être rendues facilement accessibles aux profanes et compréhensibles par ceux-ci.

(3) En parallèle à **l'identification** individuelle, il convient de garantir l'identification par le **public intéressé** dans le cas d'une mise en œuvre étatique de systèmes algorithmiques. À cet égard, il serait envisageable de créer un registre accessible au public qui indique les domaines où l'administration publique met en œuvre des systèmes algorithmiques. Entre autres, un tel registre devrait fournir des informations relatives au type et à l'origine des données traitées, à la législation, à la finalité et aux moyens de traitement, à l'organisme responsable, à

la logique du système algorithmique et aux acteurs qui ont participé au développement du système. Ces informations devraient être facilement accessibles et présentées selon un format standardisé.

Discrimination

Les systèmes algorithmiques ont souvent pour tâche de prendre des décisions. Ces dernières sont problématiques lorsque des **personnes** sont **traitées différemment en raison de caractéristiques protégées**, telles que l'origine, l'ethnie, le sexe, l'âge, la langue, la position sociale, le mode de vie, les convictions religieuses, idéologiques ou politiques, ou les invalidités d'ordre physique, mental ou psychique, sans qu'il n'existe une raison objective à cet égard. Dans ce cas de figure, il y a discrimination. Dans le cas de systèmes algorithmiques, des discriminations peuvent notamment survenir du fait qu'ils utilisent des caractéristiques protégées directement ou indirectement comme paramètres décisionnels ou parce qu'ils sont formés avec des données qui présentent un « biais ». Ainsi, certains préjugés existant dans la société peuvent être reproduits dans les prévisions et les décisions de tels systèmes. Dans de nombreux cas toutefois, la mise en œuvre de systèmes algorithmiques révèle une discrimination visible. Ainsi, l'utilisation de tels systèmes donne la possibilité d'agir à l'encontre des discriminations.

La problématique de la discrimination va bien au-delà de la mise en œuvre de systèmes algorithmiques, mais elle est particulièrement visible lors de son utilisation. Ainsi, le problème de la discrimination devrait être abordé par le biais de règles s'appliquant **indépendamment** du fait que la décision ou l'action discriminante ait été effectuée par un être humain ou par une machine. Dans la plupart des cas, la situation juridique actuelle en Suisse n'interdit que la discrimination par des acteurs étatiques. Cependant, de nombreux systèmes algorithmiques sont mis en œuvre par des organismes privés, par ex. lors de l'attribution de crédits ou du recrutement d'employés. Ces discriminations pourraient être prévenues par une **loi générale sur l'égalité de traitement**, qui enregistre et sanctionne les discriminations infligées par des organismes privés, en particulier des entreprises, en raison de caractéristiques spécifiques et protégées.

Il est souvent difficile de fournir une preuve de discrimination. Ce problème pourrait être résolu par un **renversement de la charge de la preuve**. La personne se disant discriminée devrait seulement établir l'existence d'une discrimination avec un degré de crédibilité suffisant, et l'entreprise devrait alors prouver que la décision ne se fonde pas sur une caractéristique protégée. À cet égard, l'usage de systèmes algorithmiques peut également s'avérer avantageux, puisqu'il est possible – contrairement aux décisions prises par l'être humain – de reconnaître les critères utilisés pour la prise de décision et de prouver que cette dernière ne repose pas sur des caractéristiques protégées.

Manipulation

Les systèmes algorithmiques peuvent influencer la pensée et les actions des personnes qui interagissent avec de tels systèmes. Comme exemples typiques, citons l'affichage de contenus spécifiques et la suppression d'autres contenus pertinents sur les réseaux sociaux, ainsi que la personnalisation d'offres ou de prix. L'influence ciblée sur la pensée et sur les agissements d'une personne par un tiers (manipulation) constitue toutefois un phénomène largement répandu, par ex. dans le cas de la publicité. Certes, l'influence par des tiers constitue toujours **une intrusion dans l'autonomie** de la personne concernée. Toutefois, le type et l'ampleur de l'influence sont extrêmement variés et, dans de nombreux cas, une influence ne pose pas de problème. Un tel cas de figure s'applique par exemple lorsqu'une influence est non spécifique et identifiable par la personne concernée, par ex. dans le cas des formes traditionnelles de publicité politique et commerciale.

Dans le cas de la prise en compte juridique de formes problématiques de manipulation, il faut notamment différencier les décisions et les actions d'individus dans leurs rôles de consommateurs/-trices et de citoyens/-nes :

(1) Dans le cas de la **manipulation de citoyens/-nes** dans le contexte de processus démocratiques, la protection de la **formation démocratique de la volonté** est essentielle. Cette dernière peut être mise

en danger dans le cas de l'utilisation de systèmes algorithmiques, car ceux-ci permettent des formes particulièrement efficaces et à peine identifiables de diffusion unilatérale d'information, d'exagération et de mensonge. Par ailleurs, il est possible de présenter à des personnes individuelles (ou à des petits groupes) des contenus individualisés, afin d'influencer de manière ciblée leurs pensées, leur formation d'opinion et leur comportement de vote. Cette individualisation des contenus peut introduire des déclarations qui ne sont pas abordées dans les débats publics et qui ne peuvent par conséquent pas être remises en cause ni, le cas échéant, être réfutées. Dans le cas de la formation démocratique de la volonté, la **liberté d'information et d'opinion** revêt une importance clé. Cette liberté assure aux acteurs politiques et à la population une importante marge de manœuvre dans la perception et la diffusion d'informations. Cette large marge de manœuvre est essentielle à la formation de l'opinion publique et ne doit être limitée qu'avec circonspection. Par conséquent, la réglementation des systèmes algorithmiques devrait avant tout viser à créer une transparence sur le type et l'ampleur de la diffusion de contenus éventuellement discutables (par ex. en divulguant les critères selon lesquels Facebook affiche des contenus, les supprime ou les identifie comme problématiques), sans évaluer les déclarations elles-mêmes. Cette évaluation doit rester du ressort du processus ouvert de formation de l'opinion publique. En outre, les utilisateurs/-trices doivent être en mesure de reconnaître, par le biais de mesures appropriées, la manière dont les contenus sont individualisés par des systèmes algorithmiques, afin de développer une sensibilité à la manière dont ils sont influencés.

(2) Dans le cas de la **manipulation de consommateurs/-trices**, la protection de la **liberté individuelle de décision** et la protection de la **concurrence effective** se positionnent au même niveau. Chez les consommateurs/-trices également, la manipulation par le biais de la diffusion d'informations erronées ou trompeuses revêt une importance clé. Toutefois, ce type de manipulation peut être couvert par le droit en vigueur sur la concurrence (LCD). Il en va autrement dans le cas d'autres formes de manipulation, comme l'affichage continuel de nouveaux contenus sur les plates-formes de médias sociaux dans le but de maintenir les consommateurs/-trices aussi longtemps que possible sur la plate-forme afin de pouvoir leur présenter le plus de publicité possible. Il s'agit ici de vérifier s'il existe une nécessité d'action. Cela pourrait notamment être le cas chez les personnes vulnérables (par ex. dans le cas d'une utilisation addictive des réseaux sociaux chez les mineurs).

Pour ces deux groupes, la manipulation ne doit pas nécessairement être enregistrée juridiquement comme une procédure. Au contraire, il peut être suffisant de créer des **possibilités qui permettent d'annuler la décision** si cette dernière est prise à la suite d'une manipulation. Pour les consommateurs/-trices, l'introduction de droits de rétractation serait notamment envisageable, comme il en existe déjà aujourd'hui pour le démarchage à domicile, les ventes par téléphone, et, de manière générale, – dans l'UE – la conclusion de contrats à distance (notamment dans le cadre du commerce électronique). En outre, lors des votations, la possibilité de contestation existe aujourd'hui déjà lorsque le résultat a par ex. été massivement influencé par la divulgation de fausses informations.

Responsabilité

La responsabilité en cas de dommage constitue un défi central lors de l'utilisation de systèmes algorithmiques. Bien que les normes du droit général de la responsabilité civile s'appliquent également à de tels systèmes, la preuve des conditions préalables de la **responsabilité des exploitants/-es** est difficile, en particulier en cas de faute. Dans certains secteurs, il existe déjà des règles de responsabilité objective, qui s'appliquent également dans le cas de systèmes algorithmiques (par ex. aux véhicules dans la loi sur la circulation routière ou aux drones dans la loi sur le trafic aérien). L'introduction d'une responsabilité générale des exploitants/-es sous la forme d'une responsabilité objective est à proscrire. Il

convient toutefois de vérifier si une **responsabilité objective des exploitants/-es** devrait être introduite **dans d'autres secteurs pour les exploitants/-es de systèmes algorithmiques**. Une procédure spécifique au secteur permettrait une coordination prudente avec des prescriptions de sécurité à remplir *ex ante*.

La **responsabilité des fabricants/-es** est donc mise en avant. Il s'avère problématique que la loi sur la responsabilité du fait des produits soit adaptée aux produits conventionnels et, d'une manière générale, aux objets physiques qui sont mis sur le marché après leur fabrication et qui ne peuvent plus être influencés par les fabricants/-es. L'enregistrement de systèmes algorithmiques par le biais de la **loi sur la responsabilité du fait des produits** suppose que de tels systèmes soient reconnus comme produits. Les fabricants/-es devraient donc être tenus responsables de l'évolution (ultérieure) sûre de leurs produits. Simultanément, ils doivent toutefois pouvoir se décharger en cas d'influence inappropriée de la part d'autres parties prenantes. Par conséquent, la loi suisse sur la responsabilité du fait des produits doit être actualisée.

Sécurité

Les systèmes algorithmiques doivent satisfaire les **normes habituelles de sécurité**. Par ailleurs, ils doivent être suffisamment robustes et protégés contre les influences environnementales néfastes, ainsi que contre les erreurs de manipulation. En outre, une protection suffisante contre les attaques doit être garantie, de nouvelles formes d'attaques devant être également prises en considération (par ex. manipulation de données de formation). La sévérité des exigences dépend des domaines d'application ; ainsi, par ex., les systèmes algorithmiques qui contrôlent les processus dans des infrastructures critiques (par ex. l'approvisionnement en électricité) doivent satisfaire des critères plus stricts que ceux qui, par exemple, pilotent un robot aspirateur.

Pour autant que les systèmes algorithmiques traitent des données à caractère personnel, les dispositions de la loi sur la protection des données sont applicables, ce qui implique une sécurité adéquate des données. Toutefois, ces dispositions visent en premier lieu la protection des données à caractère personnel et ne couvrent qu'indirectement les systèmes. En outre, elles ne s'appliquent pas lorsque les systèmes algorithmiques ne traitent pas de données à caractère personnel, ce qui peut justement être le cas dans les infrastructures critiques. Il convient donc d'examiner si l'introduction d'une **loi générale de sécurité informatique** est nécessaire. En guise d'alternative à la réglementation étatique d'exigences concrètes de sécurité, la déclaration d'application générale des normes développées par des organismes de normalisation pourrait s'imposer.

Procédure d'autorisation

Aujourd'hui déjà, il existe des produits qui ne peuvent être mis sur le marché qu'après lune autorisation par une autorité étatique (par ex. véhicules ou produits médicaux). Les produits doivent également être soumis à des procédures d'autorisation lorsqu'ils font appel à des systèmes algorithmiques.

Dans le cas des **procédures d'autorisation existantes**, les conditions préalables et les procédures pertinentes doivent être adaptées de telle manière à ce qu'elles garantissent également la sécurité et la qualité requises des produits, même si ces derniers reposent sur l'utilisation de systèmes algorithmiques. À cet égard, il convient de noter que les systèmes algorithmiques peuvent être perfectionnés après l'autorisation, voire même être en mesure de se perfectionner par eux-mêmes (via l'apprentissage automatique). Dans ces cas de figure, il convient de garantir que l'autorisation sera à nouveau examinée à chaque étape pertinente de développement (« *life cycle regulation* »).

Par ailleurs, il convient de vérifier si de **nouvelles procédures d'autorisation** doivent être créées afin de garantir la sécurité de produits ou de services à risque qui utilisent des systèmes algorithmiques. Les systèmes interagissant avec l'environnement, comme les robots nettoyeurs et les robots de soins, mais aussi les jouets, figurent au premier plan. D'autre part, les instruments prévisionnels utilisés dans des domaines sensibles, comme dans les procédures pénales ou la prévention de la criminalité, pourraient aussi être soumis à autorisation. Une certification pourrait aussi être envisagée pour des produits moins risqués.

Applications interdites

Enfin, il convient d'examiner si certaines applications de systèmes algorithmiques doivent être interdites, parce qu'elles conduisent (ou peuvent conduire) à des atteintes aux droits fondamentaux qui ne devraient pas être acceptées. En guise d'alternative à une **interdiction**, un **moratoire** pourrait également être décrété pour l'utilisation de systèmes algorithmiques spécifiques. Un tel moratoire permettrait d'examiner de plus près les conséquences à moyen et à long terme de l'utilisation de systèmes algorithmiques dans des domaines critiques et décider seulement ultérieurement si une mise en œuvre de tels systèmes doit être autorisée. Actuellement, l'accent est mis sur les applications suivantes :

- L'usage d'**outils de reconnaissance faciale et d'autres procédés biométriques de reconnaissance à distance** dans un espace public, dans la mesure où il existe un risque que ces systèmes algorithmiques soient utilisés pour une surveillance de masse ;
- L'utilisation du **social scoring** dans le but de réguler l'accès aux ressources fondamentales (prestations étatiques, crédits, sécurité sociale, etc.).

Au vu de l'évolution technologique rapide, il convient toutefois d'évaluer régulièrement si de nouvelles formes d'utilisation de systèmes algorithmiques (par ex. la pratique autonome de force létale dans le domaine de la sécurité) devraient également être interdites.

Position de la Suisse dans un environnement international

À l'heure actuelle, divers espaces juridiques (UE, USA, Chine) travaillent à la régulation des systèmes algorithmiques. Les développements en UE et au Conseil de l'Europe sont notamment pertinents pour la Suisse. La Suisse ne devrait pas chercher à **reprendre de manière passive ces approches réglementaires**. Au contraire, elle devrait – en se basant sur les principes formulés dans la présente prise de position – élaborer une position qui lui est propre et la présenter activement dans le cadre du discours international et en particulier européen, avec les partenaires internationaux qui partagent les mêmes opinions. À cet égard, la cohérence entre la politique intérieure et extérieure devrait être maintenue et le discours actif devrait également se refléter au niveau de la politique intérieure.

Les **entreprises suisses** qui proposent des systèmes autonomes sur le marché européen ou qui souhaitent les mettre en œuvre devront se soumettre aux prescriptions à venir du droit européen. Cela ne signifie toutefois pas que la Suisse doit reprendre ces prescriptions telles quelles dans son droit national. Il semble plutôt judicieux de créer une marge de manœuvre pour les entreprises suisses qui ne souhaitent pas (encore) proposer leurs produits sur le **marché européen** au moyen d'un cadre juridique suffisamment ouvert (par ex. via une interdiction générale de discriminer au lieu de prescriptions spécifiques relatives à la gestion des risques et à la qualité des données).

Procédure ultérieure

La présente prise de position montre qu'il est nécessaire d'agir en Suisse. Les défis liés à l'utilisation de systèmes algorithmiques par les entreprises et l'État sont suffisamment évidents. Dans ce contexte et compte tenu des développements à l'étranger, la Suisse devrait **rapidement commencer à élaborer des normes** qui peuvent appréhender de manière adéquate les défis esquissés. Ce travail devrait être assumé par une **commission interdisciplinaire et diversifiée composée d'experts/-es**. En outre, **davantage de recherches** sont nécessaires dans de nombreux domaines, par ex. dans celui de la manipulation. Les travaux de recherche requis devront être poursuivis à un rythme soutenu, parallèlement aux travaux d'une commission d'experts/-es, afin de garantir que la réglementation suisse repose sur des fondements scientifiques avérés.