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.