



Guidelines on Artificial Intelligence for the Confederation

General frame of reference on the use of
artificial intelligence within the
Federal Administration

1 Context and basics

Based on the ‘Challenges of Artificial Intelligence’ published on 13 December 2019¹, the Federal Council issued a mandate to the Federal Department of Economic Affairs, Education and Research (EAER) to work with the Federal Department of the Environment, Transport, Energy and Communications (DETEC) and the Interdepartmental Working Group on Artificial Intelligence to prepare **strategic guidelines to address the challenges of artificial intelligence (AI) at federal level**.

As a **basic technology**, AI is a key component of the ongoing digitalisation process taking place at governmental, economic and social levels. The Federal Council considers artificial intelligence to be a pivotal factor in the digital transformation, offering considerable potential for **innovation and growth**. Accordingly, favourable regulatory conditions for AI applications are crucial. However, AI also creates **specific challenges**, which include the risk of data-based discrimination in AI decisions and the inadequate traceability of results. The protection of privacy in AI must also be ensured.

AI is an important means of furthering the Federal Council's digital transformation objectives and is at the heart of development of digital infrastructures for the Federal Administration.² These guidelines serve as a **general frame of reference** on the use of AI within the Federal Administration and help to **ensure a coherent policy**. They are intended to provide guidance on the following specific contexts to federal agencies and external partners entrusted with governmental tasks:

- When developing sectoral AI strategies aimed at achieving coherent AI-relevant federal policy;
- When introducing or adapting specific regulations for all sectors affected by AI applications;
- When developing and using AI systems within the Federal Administration;
- When helping to shape the international regulatory framework on AI.

The Digital Switzerland Strategy adopted in September 2020 includes **overarching coordinated action** achieved through adoption of guidelines on government action in the area of digitalisation. The strategy is intended to facilitate structural change through the digital transformation, to actively exploit opportunities and adequately address risks. In addition, the strategy will be updated at regular intervals. The principles laid down in this strategy also apply as a frame of reference for the use of AI. Another document that was used to develop the AI guidelines is the Digital Foreign Policy Strategy for 2021-2024. Adopted in November 2020, this strategy paper presents the conceptual foundations for helping to shape *international* governance in the area of digitalisation.

Chapter 2 of the present document contains **seven basic guidelines** on the use of AI. Chapter 3 sets out how subsequent development of AI should be monitored.

The areas of education, science and innovation are very important when it comes to ensuring the availability of the expertise needed to harness the innovation and growth potential of AI and to address environmental and societal challenges. Specific guidelines for these policy areas are therefore presented in Annex 1. In addition, it is worth mentioning that the existing legal framework fully applies to AI applications having an impact on society. Explanations are given in Annex 2.

¹ Report of the interdepartmental working group on artificial intelligence to the Federal Council (2019): ‘Challenges of Artificial Intelligence’.

² Federal Council (2019) ‘Vision of the Digital Transformation within the Federal Administration and the Development of Digital Infrastructures’.

2 Guidelines on the use of AI

In Switzerland, the legal basis applying to the use of AI is provided both at national and international level, in particular the Federal Constitution of 18 April 1999 of the Swiss Confederation (Cst, SR 101) and the standards set forth in the European Convention on Human Rights (ECHR). In addition, the following guidelines must be observed when approaching AI. These guidelines apply to varying degrees depending on the context in which the AI application is to be used:³

Guideline 1: Putting people first

When developing and using AI, **the dignity and well-being of the individual as well as the common good** should be paramount. Self-determination should be preserved: people should be able to take part in political and social life in a self-responsible and autonomous manner. The use of AI should help to improve the quality of people's life and to sustainably develop our society from a social, political, economic and environmental standpoint. In this context, the use of AI should create equal opportunities for people and encourage and facilitate their access to education, goods, services and technology.

The protection of **fundamental rights** is particularly important in the use of AI. Fundamental rights and ethical considerations must also be an inherent feature in the design and use of AI ('ethics by design'). AI applications that are likely to affect fundamental rights should be accompanied by an impact assessment, continuous monitoring and appropriate safeguards and controls, especially in the case of self-learning systems.⁴

Thus, individuals, groups and genders must be protected from **discrimination** and stigmatisation. To this end, adequate technical and organisational safeguards and controls must be provided. Moreover, well-balanced, high-quality data must be used or suitable compensatory safeguards must be put in place.

AI technologies used by the Confederation must be designed in such a way that **privacy** is protected by default and **data protection provisions** are always complied with (see also Annex 2).

Guideline 2: Regulatory conditions for the development and application of AI

The Confederation must continue to ensure the best possible regulatory conditions so that the opportunities offered by AI can be exploited. Switzerland should continue its efforts to become an innovative leader in the area of AI research, development and application and remain an appealing location for companies specialised in this field. AI should help to ensure high-quality living standards based on sustainable development principles. Here, the 2030 Agenda should serve as the general frame of reference. In order to be able to achieve all of these different objectives, a balanced regulatory approach is needed.

³ The guidelines draw on the work of the OECD, the Council of Europe and the EU, see:

- 'OECD Principles on AI', May 2019 ([LINK](#))
- 'Recommendation of the Committee of Ministers of the Council of Europe on the human rights impacts of algorithmic systems', April 2020 ([LINK](#))
- 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust', February 2020 ([LINK](#)).
- 'Ethics Guidelines for Trustworthy AI' by the Independent and High-Level Expert Group on AI, established by the European Commission, June 2018 ([LINK](#)).

⁴ For the impact assessment, existing instruments such as the Regulatory Impact Assessment (RIA) ([LINK](#)), the Equality Impact Assessment ([LINK](#)) and the Data Protection Impact Assessment provided for in the new Data Protection Act (FADP) should all be used as the frame of reference.

The continuation of **technology-neutral regulation** by the federal government provides leeway for innovation. At the same time, it also ensures a democratically legitimised and responsible application of new technologies, maintains a high level of legal certainty and bolsters Switzerland's good reputation as a location for research and innovation. All of these are success factors for Switzerland's positioning in the area of AI. Such regulatory conditions have proven effective in the past and allow Switzerland to seize the opportunities offered by this enabling technology. Moreover, they enable Switzerland to use its good starting position to successfully use technologies.

With AI, too, the decision on the **choice of specific technologies** should be left to business and science stakeholders. The role of policymakers is to create the necessary freedom of action for these stakeholders while at the same time establishing the necessary framework and limitations.

Education, research and innovation are a key foundation of Switzerland's high level of competitiveness. Competences in these areas must be strengthened with a view to exploiting the application and innovation potential of AI. Where possible, obstacles to research and innovation should be lowered to pave the way for responsible development of AI.⁵ Innovations should be encouraged within the framework of the established innovation promotion bodies and be driven by the proven 'bottom-up' principle.

AI offers the potential for **economic growth, prosperity, security and employment** as well as for **reduction of the ecological footprint and energy consumption** and this potential should be optimally exploited. The economy should have sufficient room for efficiency gains and for the development of new business models. Moreover, regulatory conditions for digital solutions should be designed in such a way that does not hinder innovations and contributes both to greater value creation and sustainable development.

Guideline 3: Transparency, traceability and explainability

Transparency, traceability and explainability are basic requirements for trustworthy AI. AI-based decision-making and interaction with AI systems should be clearly recognisable as such. In order to guarantee compliance with other principles and fundamental human rights, the functioning of AI and its purpose should be disclosed in a responsible and legally compliant manner. In addition, the data sets used for the training or development of AI should be disclosed within the framework of legal obligations in order to enable monitoring. Decision-making processes based on AI should be designed in such a way that ensures that they are traceable for directly and indirectly affected persons and the mode of action must also be verifiable for experts. This applies in particular to processes that may lead to ethically questionable AI decisions. It must be taken into account that for some AI methods, traceability is a fundamental challenge.

The **data policy** must strike a balance between the protection of privacy and the use of data. It should work towards ensuring that data used for AI applications are of sufficient quality and are documented. This includes the collection and use of data for a specific purpose ('fit for purpose') in accordance with ethical standards and, in particular, ensuring the interoperability of data systems. The Data Protection Act also includes a provision requiring transparency in connection with automated decisions (cf. Annex 2 II).

⁵ The specific guidelines for the policy areas of education and science serve as a basis for improving social responsibility in those areas. These are presented in Annex 1.

Guideline 4: Accountability

In order to be able to clarify and establish responsibilities in the event of damage, an accident or a violation of the law, **liability must be clearly defined** when using AI. It must not be possible to delegate responsibility to machines.

Guideline 5: Safety

AI systems must be designed to be **safe, robust and resilient** in order to have a positive impact on people and the environment. They must not be vulnerable to misuse or misapplication. Appropriate safeguards must also be in place to prevent serious misuse. Where appropriate, decentralised networked AI systems should be relied upon. By adequately tracking and assessing the impact of the use of AI, the risks for individuals, society, the economy and the environment can be identified at an early stage and either eliminated or mitigated.⁶

Guideline 6: Actively shape AI governance

As a highly developed and interconnected country, Switzerland feels that it is important to **actively shape global AI governance**. It should therefore continue to be involved in the relevant international organisations and processes, such as the UN, OECD, ITU, UNESCO, Council of Europe and Partnership for Peace (PfP). This is especially important when it comes to developing global standards and norms on the use of AI. At the same time, Switzerland must also monitor developments taking place within the EU and NATO. Switzerland should contribute to such developments in line with its own interests and values. In particular, it should work to ensure that existing obligations and standards - namely in the areas of human rights and responsible corporate governance⁷ - are also taken into account at international level in the development and use of AI.

Guideline 7: Involve all relevant national and international stakeholders

In the debates on AI governance, Switzerland should work to ensure that **all relevant stakeholders are included in the political decision-making processes**. This means not just national governments but also the private sector, civil society and technical and scientific experts from all over the world (including from developing countries). These stakeholders must also effectively be held accountable for how they use AI. In line with its foreign policy priorities, Switzerland also supports networking and cross-sectoral cooperation between stakeholders on to topic of AI. The aim is to strengthen Geneva's position as the centre for digital AI governance.

⁶ In particular, national strategies for critical infrastructure protection (CIP) ([LINK](#)) and for the protection of Switzerland against cyber risks (NCS) ([LINK](#)) must be taken into account in risk and vulnerability analyses.

⁷ The main international standards on responsible corporate governance are the United Nations Guiding Principles on Business and Human Rights (UNGPs), the OECD Guidelines for Multinational Enterprises and the ILO Tripartite Declaration of Principles concerning Multinational Enterprises and Social Policy.

3 Monitoring advances in AI

Given the strong technological dynamics, advances in AI should be closely and continuously monitored and intensified where necessary.⁸ This should be done within the framework of sectoral policy under the responsibility of the respective federal agencies. To ensure coherence, the present guidelines should serve as the frame of reference. It is also important for all stakeholders **to maintain dialogue and to share information and knowledge** on the challenges that specifically affect the given sector as well as on the various AI-related measures taken.

Given the global dimension of AI, dialogue is conducted at both national and international level, particularly in international forums in which Switzerland is also represented. These forums provide an important framework when it comes to shaping and guiding fundamental issues surrounding the subsequent development and use of AI and ensuring appropriate monitoring.

These guidelines must also remain relevant and applicable to the latest developments. Regular assessment of application of these guidelines and any adjustments made must be in alignment with the Digital Switzerland Strategy and the Digital Foreign Policy Strategy for 2021-2024.

⁸ The 'Challenges of Artificial Intelligence' report published in December 2019 provides a detailed overview of the policy areas where the various federal agencies are already doing this. It also presents the specific challenges and technical issues that they are already addressing.

Annex 1: Specific Guidelines for the policy area Education, Research and Innovation

The essential conditions for the successful use of AI include the acquisition of digital skills within the context of **education, science and innovation**.

AI offers tremendous opportunities when it comes to improving **teaching and learning processes**. At the same time, there are certain skills that citizens must have in order to live and work in an AI-driven digitalised society. The education system must ensure that everyone in society acquires basic digital skills while offering scope for more advanced AI specialist training in such things as AI system production.

Science has a key role to play in ensuring that AI applications develop in a manner that best serves the interests of society and sustainable development. First of all, science provides the **foundations for the further development of AI technologies**, for which there is intense global competition. In addition, scientists conduct research on AI applications, thereby enabling AI to be used to address major social challenges. And finally, any risks and problems associated with AI can be prevented or mitigated on the basis of new scientific findings.

In order to optimally exploit the opportunities offered by AI and to establish Switzerland as a leading location for AI research, development and applications, **competences in education, research and innovation must on par with developments and must be further improved**. This will require close adherence to the federal policy on technology and innovation, which has thus far proven effective for the dynamic technological development of AI:

- **Basically a technology-neutral approach**

The Confederation does not specify which technologies should be used in education, science and industry and largely refrains from promoting specific technologies. Policymakers should ensure that regulatory conditions enable new technologies to develop in an optimal and innovation-friendly manner. Such openness on the part of the state towards new technologies allows the potential of new ideas and innovations to be optimally exploited. In order to encourage transparency, traceability and explainability, preference should be given to open AI systems wherever possible; this applies in particular to AI applications in education and science.

- **Bottom-up approach**

The decision on the choice of specific technologies is left to individual stakeholders in the education, science and business spheres. The role of policymakers is to provide these stakeholders with the necessary freedom of action and favourable conditions. A high degree of autonomy of the stakeholders combined with competition between them ensures stakeholder responsibility.

In order for science and education to make the greatest possible contribution to societal well-being, research into AI applications should also take into account the full range of **societal and environmental impacts**. Here are some of the main considerations to be taken into account:

- **Basics of artificial intelligence**

The core methodological elements of AI must be open as this fundamentally changes the scope and possible applications of these technologies both from a quantitative and qualitative standpoint. It is therefore essential that the methodological foundations also be expanded in parallel to AI applications.

- **Ethical, legal and social norms and values**

Science should help to ensure that AI research takes into account societal and environmental impacts and that the design and application of AI systems take into account legal norms as well as ethical and societal demands and values.

- **Cooperation between humans and AI**

Science and education should improve our understanding of AI in a way that allows humans to successfully use AI in a way that complements and enhances their activities and skills.

- **Acceptance of AI**

Science should encourage greater public awareness of AI, including opportunities and risks associated with the application of AI. This should be achieved through active dialogue with society and through training (particularly in the area of data literacy).

AI-related challenges that exist in this context should be addressed by research and education institutions in their area of responsibility. In particular, this means that the relevant bodies and higher education institutions should establish guidelines on how AI should be used in the given area of research and education. Moreover, research priorities should be set and adequate funding should be allocated to corresponding research projects.

Annex 2: Applicability of the applicable legal system

I. Generally applicable standards

AI develops within a pre-existing legal framework that also applies to it. This legal framework is comprised of both international and national standards. This appendix presents some important legal instruments whose technology-neutral provisions also apply to AI. One section is specifically dedicated to the new Data Protection Act, which provides for special measures for AI. The following list is not exhaustive.

- **Fundamental rights and human rights:**

Fundamental rights must be respected throughout the legal system. They are listed in Articles 7 ff. of the Federal Constitution (Cst, SR 101). International law such as the European Convention on Human Rights (ECHR; SR 0.101) and the International Covenant on Civil and Political Rights (SR 0.103.2) provide analogous guarantees. Fundamental rights must be applied throughout the legal system (Art. 35 Cst). Furthermore, any restriction of fundamental rights must comply with the requirements of Article 36 of the Federal Constitution. Compliance with the principles of international humanitarian law must be ensured at all times. Thus, conformity with international humanitarian law must be ensured in the development of new weapons.

- **Protection of intellectual property:**

The area of intellectual property is also important for AI, especially in the processing or use of data protected by copyright. Here it is important to take into account the Patent Act (PatA; SR 232.14), the Copyright Act (CopA; SR 231.1) and various legal provisions on trade secrecy.

- **Civil and criminal liability:**

Since robots do not have legal personality, a natural or legal person is liable for damages caused by AI if the conditions for liability are met. For example, traffic accidents or medical errors can constitute liability cases that can be attributed to the use of AI-based software. The Swiss Code of Obligations (CO; SR 220), the Swiss Criminal Code (SCC; SR 311.0) or the Product Liability Act (PLA; SR 221.112.944), among others, apply in this regard.

- **Prohibition of discrimination:**

The provisions on the prohibition of discrimination and the promotion of equality naturally also apply to AI. In particular, Articles 8 paragraphs 2 and 3 of the Federal Constitution, the Gender Equality Act (GEA; SR 151.1), the Disability Discrimination Act (DDA; SR 151.3), Articles 28 and seq. of the Swiss Civil Code (SCC; SR 210) on the protection of personal privacy, the International Covenant on Economic, Social and Cultural Rights (SR 0.103.1), the International Convention on the Elimination of All Forms of Racial Discrimination (SR 0.104) and the Convention on the Rights of Persons with Disabilities (SR 0.109) all apply to AI.

- **Product safety:**

The Product Safety Act (ProdSA; SR 930.11) provides for product safety and the commercial or professional placing of products on the market. ProdSA protects the safety and health of users and third parties, but does not protect privacy. Networked toys are subject to the provisions of FDHA's Toy Safety Ordinance (ToySO; SR 817.023.11) as the main legislative instrument. ToySO only contains requirements regarding the health protection of consumers. The risks associated with the digital networking of a toy

are not covered by this ordinance. For risks relating to data protection or privacy in toys, the Data Protection Act applies.

II. New Data Protection Act

Data protection legislation and in particular the revised Data Protection Act (FADP; SR 235.1) play a key role in the AI sector. The revised FADP, which was adopted in the 2020 autumn session, provides for various measures that could have a specific impact on AI:

- **The term ‘profiling’:**

The new Data Protection Act defines profiling as ‘any automated processing of personal data to evaluate certain characteristic traits of a natural person, in particular for the purpose of analysing or predicting aspects of that natural person’s work performance, economic situation, health, personal preferences, interests, reliability, behaviour, location or change of location’ (Art. 5 let. f of the revised FADP). Profiling requires certain safeguards to be put in place (in particular the need for a formal legal basis, see Art. 34 para. 2 let. b of the revised FADP). The Swiss Parliament has also introduced a new legal definition for ‘high-risk profiling’ in Art. 5(g). However, this term is mostly relevant within the context of private law. The requirement of that a formal legal basis exist on profiling by federal bodies has remained unchanged in the Federal Council’s draft for the revised FADP.

- **Genetic and biometric data:**

Genetic and biometric data that uniquely identify a natural person are now part of the catalogue of particularly sensitive personal data. This has consequences for AI applications that use technologies such as facial recognition. Certain requirements also apply to the processing of particularly sensitive personal data (in particular the need for a formal legal basis, cf. Art. 34 para. 2 let. a of the revised FADP).

- **Use of technology to protect privacy and default settings aimed at protecting privacy (‘privacy by design’ / ‘privacy by default’):**

The party responsible is required to design data processing in such a way that technical and organisation aspects are in compliance with data protection regulations. This must be done from the planning stage (Art. 7 of the revised FADP).

- **Requirement to conduct an impact assessment:**

The party responsible must prepare a data protection impact assessment in advance if a planned processing operation poses considerable risk to the personal privacy or fundamental rights of a natural person. Considerable risk exists in particular in the case of extensive processing of sensitive personal data or if extensive public areas are systematically monitored (Art. 22 of the revised FADP).

- **Duty to provide information in the case of an automated decision (Art. 21 and 25 para. 2 let. f of the revised FADP):**

The party responsible shall inform the data subject of any decision based solely on automated processing when said decisions has a legal consequence for the data subject or significantly affects him or her. In addition, the data subject has the right to express his or her stance on the decision and may request that the decision be reviewed by a natural person. These measures do not apply if the data subject has expressly consented to the decision being automated or if the decision is directly related to the awarding or performance of a contract between the party responsible and the data subject and the data subject’s request has been granted. If the automated decision is

made by a federal agency, it must mark the decision accordingly. The right of the data subject to express his or her stance on the automated decision and to request that the given automated decision be reviewed by a natural person does not apply if the person is not required to be heard prior to the decision pursuant to Article 30 paragraph 2 of the Administrative Procedure Act (APA, SR 172.021) or pursuant to another federal act. When exercising his or her right to information, the data subject shall in particular be notified of the existence of any automated decision and the logic on which that decision is based.

- **Need for a formal legal basis:**

Federal agencies may only process personal data if there is a legal basis to do so (Art. 34 of revised FADP). A legal basis in the formal sense is required in the following cases:

- When particularly sensitive personal data is to be processed;
- When the activity at hand qualifies as profiling;
- When the purpose of data processing or data processing method used may seriously undermine the fundamental rights of the data subject. This could be the case, for instance of an automated decision. In this latter case, a legal basis would still be required even if no profiling takes place and even if no particularly sensitive data are processed.

Article 34(3) regulates the exceptions.