



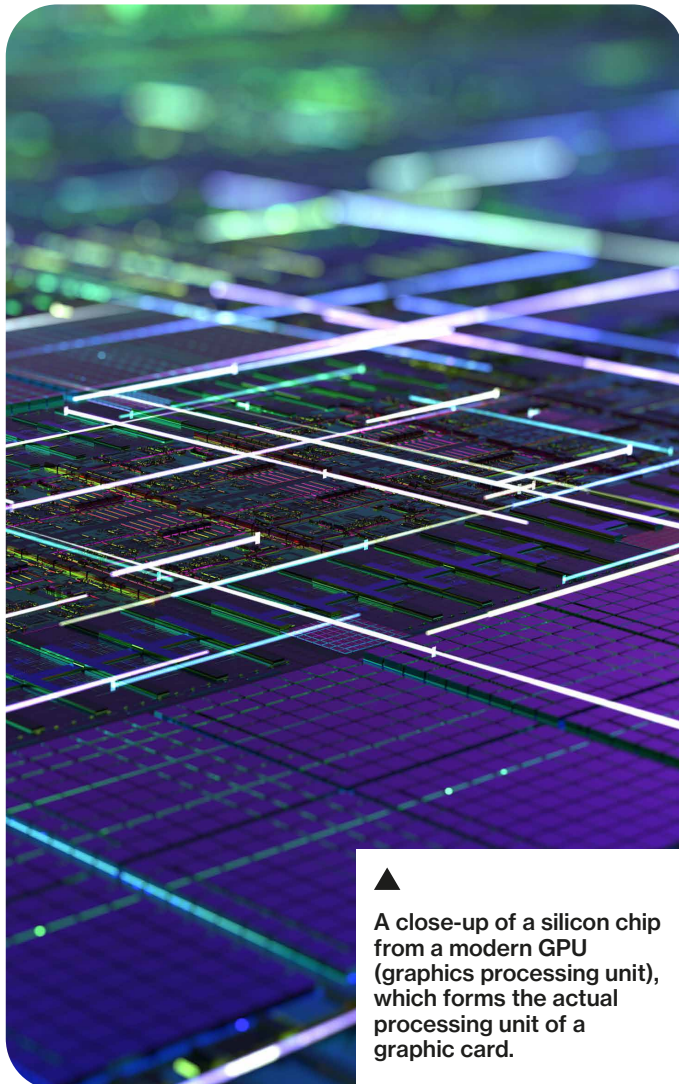
Local AI and Cybersecurity

How Switzerland can use edge AI in
a secure and trustworthy manner

Implications of edge AI for cybersecurity

AI models can increasingly be deployed directly on devices and are no longer reliant on cloud computing power. These so-called edge AI models also influence cybersecurity. Specific courses of action are outlined to ensure that Switzerland does not miss out on developments in this area and can utilize edge AI in a secure and trustworthy manner.

This factsheet is based on the work of the Cybersecurity Topical Platform of the Swiss Academy of Engineering Sciences SATW.



▲
A close-up of a silicon chip from a modern GPU (graphics processing unit), which forms the actual processing unit of a graphic card.

Developments in the field of artificial intelligence (AI), which here refers to data-driven learning methods such as machine learning, in which computer models derive patterns from data rather than following explicit rules, are characterized by rapid scaling in many respects. This is accompanied by exponential growth in volumes of data and a strong demand for computing power. Consequently, the demand for cloud infrastructures is also rising. These infrastructures are used both for training AI models and their deployment. In this context, the term “cloud AI” is also used. The cloud is the central element in this context, where data from distributed sources is aggregated and complex models can be trained or deployed.

A new trend toward local deployment

The sheer scale of this growth in cloud-based AI is overshadowing the fact that a second interesting development is taking place in parallel: AI systems are increasingly being deployed locally at the periphery of the communications network, i.e. at the edge, and on devices with limited storage space and computing power. The term “edge” does not refer to an abstract technical field, but quite simply to the place where data is generated. This could be, for example, in digital devices, autonomous vehicles, industrial robots and sensors, or a smart camera. Edge AI therefore means that the AI system operates locally and, consequently, close to the data source (see arrow in Figure 1), rather than in a remote central data center to which the data must first be transmitted (cloud AI).

From cloud to edge AI

The distinction between cloud AI and edge AI is not entirely clear. Between the two extremes of training and deployment in the cloud and on the device, respectively, there are various possible combinations (see Figure 1). In the literature, these combinations are still generally grouped under the rather inconsistent term “edge AI” [1] – even when, in fact, only the device is used (Level 7). The combination currently predominant for edge AI applications remains the one in which resource-intensive training takes place in the cloud and deployment occurs at the edge (Levels 1 to 3) [2], [3].

Depending on the combination of edge AI and cloud AI, the advantages and disadvantages of the two approaches become more apparent (see the gray arrows in Figure 1). In particular, when deployed, edge AI enables quick response times, reduced bandwidth requirements, offline availability and protection of sensitive data inherent in the system – in other words, adherence to the principle of “privacy by design” – and resource-efficient data processing [4]. These advantages stem mainly from the fact that the data is processed locally rather than in the clouds of AI service providers, and because edge AI models tend to be significantly smaller than those in the cloud. However, edge computing has less storage capacity and computing power and is exposed to the risk of an incomplete data set, as only local data is accessed [2].



Box 1: Examples of edge AI applications

Use cases stem primarily from the benefits offered by edge AI. Edge AI is already widely used in industry, one example being in the field of predictive maintenance, where sensors monitor production processes and can raise the alarm in real time if any anomalies are detected, without sensitive data having to leave the production site.

Moreover, in autonomous mobility where decisions are extremely time-sensitive, edge AI makes it possible to analyze information from various sensors in real time, without the need for a permanent internet connection or guaranteed bandwidth.

Furthermore, the system-level protection of sensitive data offers great potential for applications involving personal data, such as those in the healthcare sector, for example in real-time prognoses carried out locally in emergency situations, or in the education sector in the context of personalized learning apps [2], [5], [6].

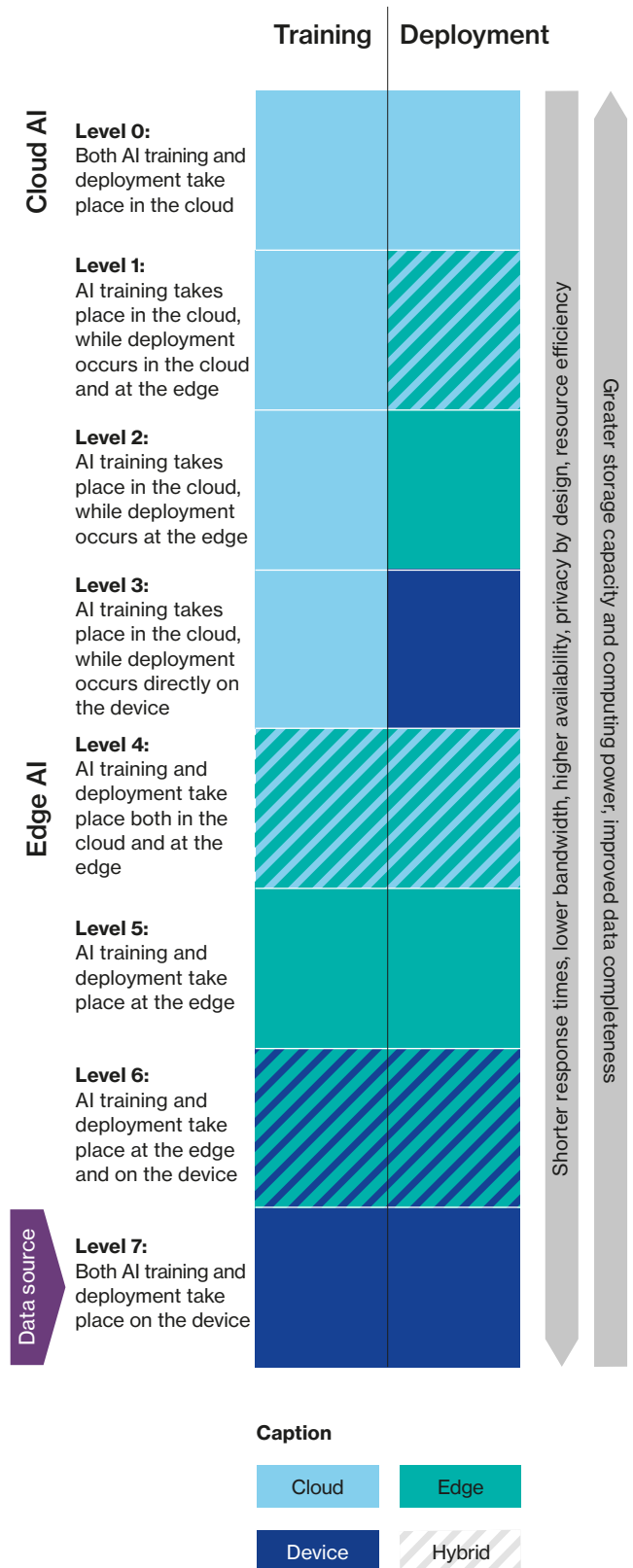


Figure 1: The figure illustrates the various ways in which AI can be trained and deployed in the cloud or at the edge. Whenever this does not take place entirely in the cloud (Level 0), it represents various forms of edge AI, right up to the extreme variant, in which both the training and deployment of AI take place on the device (Level 7). Illustration adapted from: [2], [7].

Edge AI creates cybersecurity trade-offs

The use of edge AI offers both advantages and disadvantages in relation to the three key principles of cybersecurity, namely confidentiality, integrity and availability (see Figure 2), although these often conflict with one another. When it comes to availability, edge AI generally improves it by reducing reliance on functioning network connections. At the same time, the use of edge AI stands or falls with the availability of the relevant device. If this is faulty, it is no longer possible to access the AI model in the case of very sophisticated forms of edge AI, because the device constitutes the single point of failure and there is no backup.

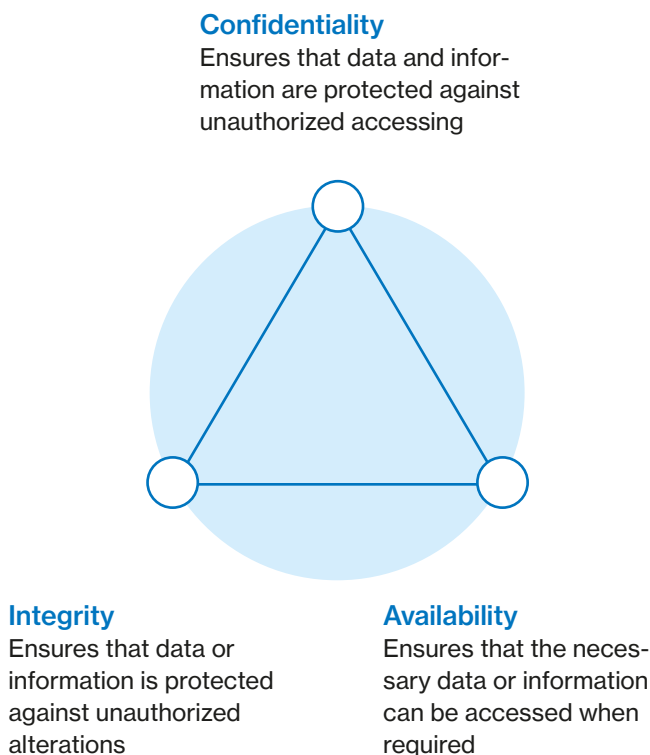


Figure 2: The figure illustrates the “trinity” of confidentiality, integrity and availability, which is well known in the field of cybersecurity. All three elements must be met in order to ensure a system’s cybersecurity.

Confidentiality and integrity, in turn, can be enhanced because, when edge AI is used, data no longer needs to be processed in cloud infrastructures. As data processing is distributed, less data is concentrated in one place. Successful attacks therefore tend to have a less drastic impact on confidentiality and integrity overall, with less data being affected [1]. Individual privacy can also be promoted through edge AI and the strengthening of digital sovereignty. After all, a 2025 survey identifies dependence on cloud services as one of Switzerland’s top three

weaknesses [8]. Reducing this dependency can therefore help to build trust, particularly for edge AI applications that process personal data. Finally, edge AI eliminates the need for data transfers between the edge and the cloud, thereby reducing the risk of integrity breaches caused by intercepted data.

However, edge operators are also responsible for providing adequate local protection for their applications and data in order to ensure confidentiality and integrity. This protection becomes significantly more complex, precisely because of the proximity of data, AI models and hardware (e.g. sensors). The proximity increases both the attack surface (the number of entry points) and the attack vectors, meaning the possible methods by which edge AI can be attacked at the local level [9], shifting the responsibility for cybersecurity from the cloud to the edge. This challenge must be addressed with appropriate measures (see courses of action).



Box 2: Swiss players in the fields of edge AI and cybersecurity

A great deal of research into edge AI is being conducted in Switzerland, whether in the field of automation (e.g. [CSEM](#)), physical AI (e.g. [ETH Zurich](#)) or nanoelectronics (e.g. [EPFL](#)). In the field of cybersecurity, too, Switzerland boasts a vibrant research landscape [11] and an [active start-up scene](#).

By contrast, at the intersection of edge AI and cybersecurity, there are fewer players in Switzerland – but they do exist and they operate on an international scale. Examples of such organizations in the field of edge AI and cybersecurity include the following:

- The Europe-wide [EdgeAI-Trust consortium](#), which alongside the ZHAW also includes a local electronics enterprise and a firm specializing in the Internet of Things (IoT), is focusing on the further development of edge AI systems to improve their explainability and comprehensibility.
- The [CONVOLVE](#) initiative funded by Horizon Europe, in which ETH Zurich is involved alongside foreign companies and universities, focuses on the development and design of hardware for edge AI. The initiative’s aim is to make Europe one of the world leaders in edge AI hardware.

Courses of action: Becoming a pioneer in cybersecurity for edge AI

To ensure that the implications of edge AI for cybersecurity are taken into account, the cybersecurity topical platform of the Swiss Academy of Engineering Sciences SATW sets out the following courses of action.

These would enable the Swiss economy to play a pioneering role in ensuring the security of edge AI applications by proactively implementing industrial standards and recognized guidelines. The following measures are suitable:

- **Implementation of zero-trust architectural principles (in accordance with [NIST SP 800-207](#)):** Zero trust, based on the principle of “never trust, always verify”, encompasses a set of principles aimed at ensuring that no entity, such as an individual or a machine, is simply trusted. Accordingly, a separate access token is generated for each access to each resource, such as an edge AI application, a dataset or a service, and this access token can also be revoked at any time should the security situation change. This architecture makes it possible to address the cybersecurity risks posed by edge AI, given its larger attack surface and the many attack vectors.
- **Implementation of international standards:** When edge AI is used in an industrial context, implementing the series of standards for industrial automation ([IEC-62443](#)) is a logical choice. The approach taken by this series of standards, which involves defining zones and communication channels relevant to security, makes it possible to reflect the distributed nature of edge AI applications and to operate these applications within a segregated zone.
- **Safety standards for AI models:** New or revised standards that address specific security issues relating to AI models, such as data poisoning (manipulated training data), model poisoning (manipulated model weights) or adversarial attacks (manipulated input data), should be introduced. Edge AI exacerbates these risks because the systems are distributed, hardware and software are physically closer to one another, the attack surface is larger, and the attack vectors are more numerous. Security issues of this nature could be addressed before edge AI systems are actually deployed, thanks to mandatory red team tests (i.e. targeted attack simulations).

Where voluntary initiatives to ensure the cybersecurity of edge AI applications are insufficient, regulations should require providers to adhere to a binding cybersecurity framework. High-risk AI systems, meaning edge AI systems used, for example, in critical infrastructure (such as transport or electricity and water networks), in biometric identification or in educational assessment, should be required to adhere to the principle of security by design. This principle requires that the accuracy, robustness and cybersecurity of systems are not merely ensured retrospectively, but are embedded in the architecture and development process from the outset. Such an obligation is also laid down in Article 15 of the [EU AI Act](#).

As edge AI models process data and execute code locally, the protection of hardware components during operation is also becoming increasingly important. To ensure this protection, providers must increasingly rely on trusted execution environments (TEEs). These TEEs involve a hardware-based isolation of the area in which data is processed and the code of the edge AI model is executed [\[10\]](#).

The TEEs mentioned are a good example of how many research questions remain unanswered in the area of edge AI and cybersecurity, ranging from the specific requirements for edge AI chips to the quality assurance of edge AI models (for an overview, see [\[2\]](#)). Although Switzerland already has a certain research and business ecosystem in this area (see Box 2), research in this field should be further expanded and the transfer of findings into practice should be specifically encouraged.

Conclusion

Rapid development in the field of AI is increasingly occurring at a local level. As a highly industrialized country, Switzerland is well placed to assume a leading role in the field of edge AI. However, to ensure that this is done in a safe and trustworthy manner, the proposed courses of action should be implemented without delay.



Modern edge AI computer in an industrial environment with a tablet and smart-phone. A symbolic image representing secure, connected and mobile AI applications.

Swiss Academy of Engineering Sciences SATW

SATW identifies industrially relevant technological developments on behalf of the federal government and informs politicians and society about their significance and ramifications. As a specialist organization with a high level of credibility, it provides independent, objective and comprehensive information on technology as a basis for the formation of sound opinions. SATW also promotes an interest in and understanding of technology among the general public, especially young people. It is a politically independent, non-commercial organization.



satw

Swiss Academy of
Engineering Sciences SATW

St. Annagasse 18
8001 Zurich

044 226 50 11
info@satw.ch
www.satw.ch

Imprint

Project management: Tobias Schlegel

Authors: Umberto Annino, Matthias Bossardt,
Daniel Caduff, Stefan Frei, Roger Halbheer,
Pascal Lamia, Martin Leuthold, Hannes Lubich,
Raphael Reischuk, Bernhard Tellenbach,
Daniel Walther, Andreas Wespi.

Editorial: Esther Lombardini

Graphic design: Brandlook GmbH

Images: iStockphoto (Quardia), OpenAI (ChatGPT)

Illustrations: Brandlook GmbH

Translation: Translingua AG

Printing: Egger AG

2026