



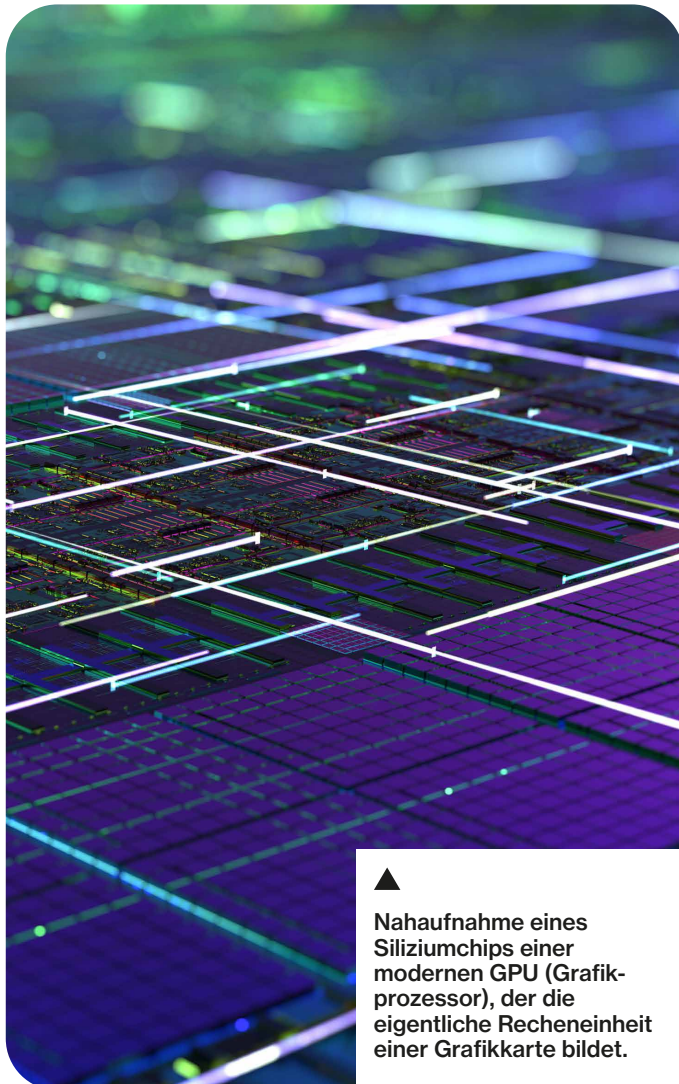
Lokale KI und Cybersicherheit

Wie die Schweiz Edge-KI sicher
und vertrauenswürdig nutzen kann

Implikationen von Edge-KI für die Cybersicherheit

KI-Modelle können zunehmend direkt auf Endgeräten eingesetzt werden und sind nicht mehr auf Cloud-Rechenleistung angewiesen. Diese sogenannten Edge-KI-Modelle haben dabei auch einen Einfluss auf die Cybersicherheit. Damit die Schweiz diesbezüglich keine Entwicklungen verpasst und Edge-KI auf sichere und vertrauenswürdige Art einsetzen kann, werden konkrete Handlungsoptionen aufgezeigt.

Dieses Faktenblatt basiert auf den Arbeiten der Themenplattform Cybersecurity der Schweizerischen Akademie der Technischen Wissenschaften SATW.



▲
Nahaufnahme eines Siliziumchips einer modernen GPU (Grafikprozessor), der die eigentliche Recheneinheit einer Grafikkarte bildet.

Die Entwicklungen im Bereich künstliche Intelligenz (KI) – gemeint sind hier datengetriebene Lernverfahren wie maschinelles Lernen, bei denen Computermodelle aus Daten Muster ableiten, statt expliziten Regeln zu folgen – sind in vielen Belangen geprägt durch rasche Skalierung. Damit einher gehen exponentiell wachsende Datenmengen und eine starke Nachfrage nach Rechenleistung. Entsprechend steigt auch der Bedarf nach Cloud-Infrastrukturen. Diese Infrastrukturen werden sowohl für das Training von KI-Modellen als auch für deren Einsatz verwendet. In einem solchen Kontext spricht man auch von Cloud-KI. Die Cloud stellt hierbei das zentrale Element dar, in dem Daten aus verteilten Quellen aggregiert und komplexe Modelle trainiert oder eingesetzt werden.

Neuer Trend zum Lokalen

Durch das schiere Ausmass dieses Cloud-KI-Wachstums wird in den Hintergrund gedrängt, dass parallel dazu eine zweite interessante Entwicklung stattfindet: Vermehrt werden KI-Systeme in der Peripherie des Kommunikationsnetzwerks – dem sogenannten «Edge» – und auf Geräten mit limitiertem Speicherplatz und begrenzter Rechenleistung lokal eingesetzt. Der Begriff «Edge» bezeichnet dabei nicht einen abstrakten technischen Bereich, sondern ganz einfach den Ort, an dem Daten entstehen. Das kann beispielsweise in digitalen Endgeräten (Devices), autonomen Fahrzeugen, industriellen Robotern und Sensoren oder einer intelligenten Kamera sein. Edge-KI bedeutet also, dass das KI-System lokal und damit nahe an der Datenquelle arbeitet (vgl. Pfeil in Abbildung 1), statt in einem weit entfernten zentralen Rechenzentrum, an das die Daten zuerst übermittelt werden müssen (Cloud-KI).

Von Cloud- bis Edge-KI

Die Unterscheidung zwischen Cloud-KI und Edge-KI ist dabei nicht messerscharf. Zwischen den beiden Extremen des Trainings und des Einsatzes in der Cloud bzw. auf dem Endgerät liegen verschiedene Kombinationsmöglichkeiten (vgl. Abbildung 1). In der Literatur werden diese Kombinationen noch eher uneinheitlich unter dem Begriff Edge-KI zusammengefasst [1] – selbst dann, wenn eigentlich nur noch das Endgerät zum Einsatz kommt (Level 7). Die zurzeit dominierende Kombination für Edge-KI-Anwendungen ist nach wie vor jene, bei der das ressourcenintensive Training in der Cloud und der Einsatz im Edge (Level 1 bis 3) stattfinden [2], [3].

Je nach Kombination von Edge-KI bzw. Cloud-KI treten auch die Vor- und Nachteile der beiden Ausprägungen stärker zu Tage (vgl. graue Pfeile in Abbildung 1). Insbesondere beim Einsatz erlaubt Edge-KI schnelle Reaktionszeiten, reduzierte Bandbreiten, Offline-Verfügbarkeit, einen systemimmanenten Schutz sensibler Daten – also die Einhaltung des Grundsatzes der Privacy by Design – und ressourcenschonende Datenverarbeitung [4]. Diese Vorteile ergeben sich mehrheitlich, weil die Daten lokal und nicht in den Clouds von KI-Service-Anbietern verarbeitet werden und weil Edge-KI-Modelle in der Tendenz deutlich kleiner sind als jene in der Cloud. Gleichzeitig muss der Einsatz im Edge mit weniger Speicherkapazität und Rechenleistung auskommen und setzt sich der Gefahr einer nicht vollständigen Datenlage aus, weil nur auf lokale Daten zugegriffen wird [2].



Box 1: Anwendungsbeispiele von Edge-KI

Die Anwendungsfälle ergeben sich im Wesentlichen aus den Vorteilen, die Edge-KI bietet. Bereits verbreitet im Einsatz ist Edge-KI in der Industrie wie etwa im Bereich der vorausschauenden Instandhaltung, bei der Sensoren Produktionsprozesse überwachen und bei Ungereimtheiten in Echtzeit Alarm schlagen können, ohne dass dabei sensible Daten die Produktionsstätte verlassen müssen.

Auch in der autonomen Mobilität – wo Entscheidungen enorm zeitkritisch sind – ermöglicht es Edge-KI, Informationen aus verschiedenen Sensoren in Echtzeit auszuwerten und dies ohne Notwendigkeit einer permanenten Internetverbindung bzw. garantierter Bandbreiten.

Der systemimmanente Schutz sensibler Daten eröffnet wiederum grosses Potenzial bei Anwendungen mit Personendaten, wie sie im Gesundheitsbereich beispielsweise bei lokal durchgeführten Echtzeitprognosen in Notfällen oder im Bildungsbereich im Rahmen von individualisierten Lern-Apps vorkommen [2], [5], [6].

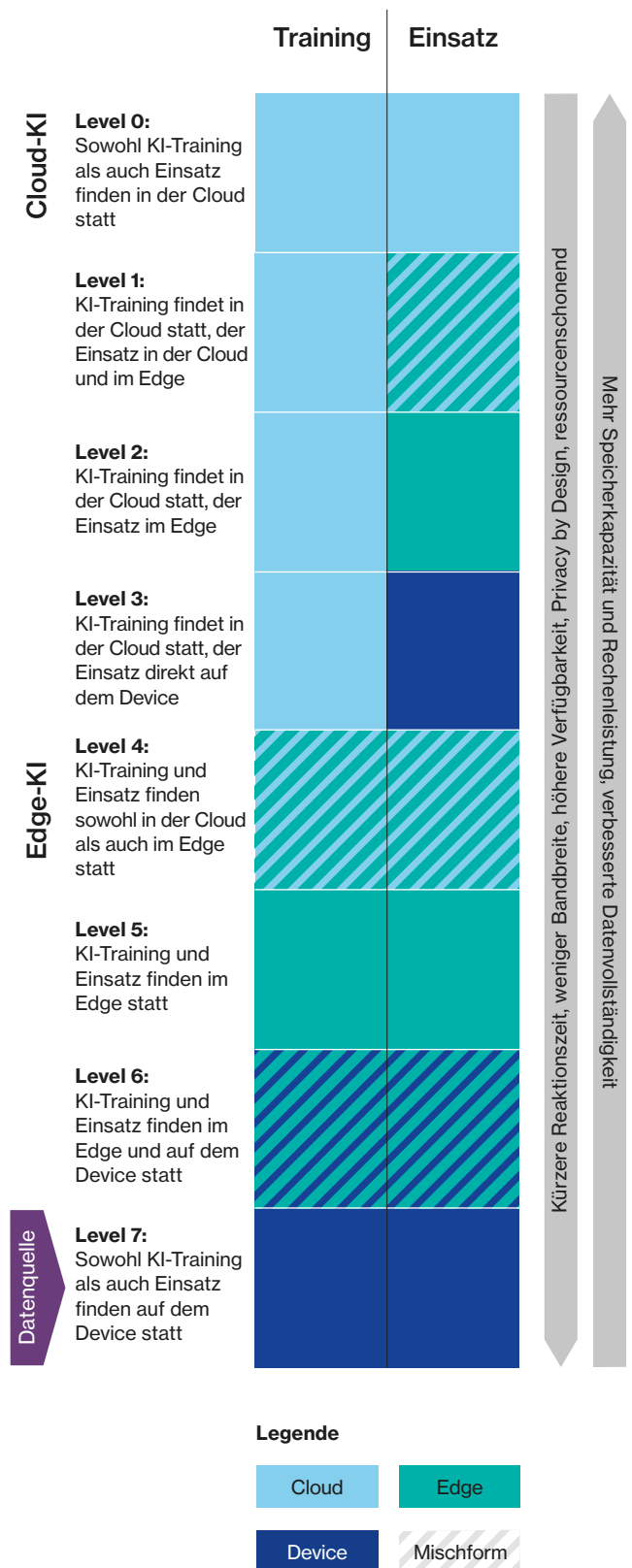


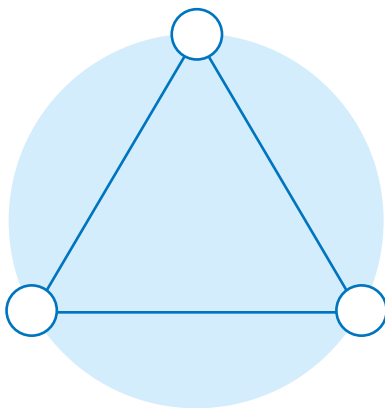
Abbildung 1: Die Abbildung zeigt die verschiedenen Möglichkeiten, mit welcher KI in der Cloud bzw. im Edge trainiert und eingesetzt werden kann. Wann immer dies nicht vollständig in der Cloud geschieht (Level 0), handelt es sich um unterschiedliche Ausprägungen von Edge-KI, bis hin zur Extremvariante, bei der sowohl Training als auch Einsatz der KI auf dem Device stattfinden (Level 7). Abbildung in Anlehnung an: [2], [7].

Edge-KI schafft Zielkonflikte bezüglich Cybersicherheit

Der Einsatz von Edge-KI bietet entlang des in der Cybersicherheit gängigen Dreiklangs aus Vertraulichkeit, Integrität und Verfügbarkeit (vgl. Abbildung 2) Vor- und Nachteile, wobei diese oft im Zielkonflikt zueinanderstehen. So wird die Verfügbarkeit durch die Unabhängigkeit von funktionierenden Netzwerkverbindungen im Allgemeinen besser. Gleichzeitig steht und fällt der Einsatz der Edge-KI mit der Verfügbarkeit des jeweiligen Endgerätes. Ist dieses defekt, kann bei sehr ausgeprägten Arten der Edge-KI nicht mehr auf das KI-Modell zugegriffen werden – weil das Endgerät den Single Point of Failure darstellt und kein Back-up existiert.

Vertraulichkeit

Stellt sicher, dass Daten und Informationen vor unerlaubtem Zugriff gesichert sind



Integrität

Stellt sicher, dass Daten oder Informationen vor unerlaubten Veränderungen gesichert sind

Verfügbarkeit

Stellt sicher, dass bei Bedarf auf die benötigten Daten oder Informationen zugegriffen werden kann

Abbildung 2: Die Abbildung zeigt den in der Cybersicherheit bekannten Dreiklang von Vertraulichkeit, Integrität und Verfügbarkeit, wobei alle drei Elemente erfüllt sein müssen, um die Cybersicherheit eines Systems zu gewährleisten.

Vertraulichkeit und Integrität wiederum können gestärkt werden, weil Daten beim Einsatz von Edge-KI nicht mehr auf Cloud-Infrastrukturen verarbeitet werden müssen. Zum einen befinden sich aufgrund der verteilten Datenverarbeitung weniger Daten konzentriert an einem Ort. Erfolgreiche Angriffe wirken sich so tendenziell weniger drastisch auf Vertraulichkeit und Integrität insgesamt aus, da weniger Daten betroffen sind [1]. Auch die individuelle Vertraulichkeit kann durch Edge-KI und die Stärkung der digitalen Souveränität gefördert

werden. Immerhin wird die Abhängigkeit von Cloud-Diensten in einer Umfrage von 2025 als eine der Top-3-Schwächen der Schweiz beurteilt [8]. Diese Abhängigkeit zu reduzieren kann insbesondere für Edge-KI-Anwendungen, die Personendaten verarbeiten, vertrauensbildend wirken. Schliesslich fallen Datentransfers zwischen Edge und Cloud weg. Somit sinkt das Risiko von Integritätsverletzungen durch abgefangene Daten.

Gleichzeitig stehen die Betreiber des Edge in der Verantwortung, ihre Anwendungen und Daten lokal ausreichend zu schützen, um Vertraulichkeit und Integrität zu gewährleisten. Dieser Schutz wird aber gerade aufgrund der Nähe von Daten, KI-Modellen und Hardware (z. B. Sensoren) deutlich komplexer. Dadurch erhöhen sich nämlich lokal sowohl die Angriffsfläche (die Zahl der Einstiegspunkte) als auch die Angriffsvektoren – also die möglichen Methoden, mit welchen eine Edge-KI angegriffen werden kann [9]. Es kommt so zu einer Verlagerung der Cybersicherheitsverantwortung von der Cloud hin zum Edge. Dieser Herausforderung muss mit entsprechenden Massnahmen begegnet werden (vgl. Handlungsoptionen).



Box 2: Schweizer Akteure in den Bereichen Edge-KI und Cybersicherheit

In der Schweiz wird viel Forschung zu Edge-KI betrieben, sei es im Bereich Automatisierung (z. B. [CSEM](#)), physische KI (z. B. [ETH Zürich](#)) oder Nanoelektronik (z. B. [EPFL](#)). Und auch im Bereich Cybersicherheit weist die Schweiz sowohl eine lebendige Forschungslandschaft [11] als auch eine [aktive Start-up-Szene](#) auf.

Im Schnittpunkt von Edge-KI und Cybersicherheit sind die Akteure in der Schweiz naturgemäss weniger zahlreich – aber es gibt sie und sie sind international aufgestellt. Beispiele solcher Akteure im Bereich Edge-KI und Cybersicherheit sind (nicht abschliessend):

- Das europaweite [Konsortium EdgeAI-Trust](#), dem neben der ZHAW auch ein lokales Elektronikunternehmen und eine Firma im Bereich Internet of Things angehören, fokussiert auf die Weiterentwicklung von Edge-KI-Systemen, um deren Erklär- und Verstehbarkeit zu verbessern.
- Die von Horizon Europe finanzierte Initiative [CONVOLVE](#), an der neben ausländischen Unternehmen und Universitäten auch die ETH Zürich beteiligt ist, beschäftigt sich mit der Entwicklung und dem Design von Hardware für Edge-KI. Anspruch der Initiative ist es, dass Europa im Bereich Edge-KI-Hardware zur Weltspitze gehört.

Handlungsoptionen: Vorreiterrolle bei Cybersicherheit für Edge-KI übernehmen

Damit den Implikationen von Edge-KI auf die Cybersicherheit Rechnung getragen werden, zeigt die Themenplattform Cybersecurity der Schweizerischen Akademie der Technischen Wissenschaften SATW die nachfolgenden Handlungsoptionen auf.

So könnte die Wirtschaft für sichere Edge-KI-Anwendungen eine Vorreiterrolle spielen, indem sie proaktiv Industriestandards bzw. anerkannte Richtlinien implementiert. Denkbar sind folgende Massnahmen:

- **Implementierung der Zero-Trust-Architekturprinzipien (gemäss NIST SP 800-207):** Zero Trust fasst unter dem Grundsatz «Vertraue nie, überprüfe immer» eine Reihe von Prinzipien zusammen, die darauf abzielen, keinem Subjekt wie etwa einem Individuum oder einer Maschine einfach zu vertrauen. Entsprechend wird separat für jeden Zugriff auf jede Ressource – wie etwa auf eine Edge-KI-Applikation, einen Datensatz oder einen Service – ein separater Zugang erstellt, der auch jederzeit bei sich ändernder Sicherheitslage wieder entzogen werden kann. Damit kann den Cybersicherheitsrisiken von Edge-KI aufgrund der grösseren Angriffsfläche und den vielen Angriffsvektoren begegnet werden.
- **Implementierung internationaler Normen:** Wird Edge-KI im industriellen Kontext eingesetzt, bietet sich eine Implementierung der Normenreihe für Industrieautomatisierung (IEC-62443) an. Der Ansatz dieser Normenreihe, sicherheitsrelevante Zonen und Kommunikationskanäle zu definieren, erlaubt es, den verteilten Charakter von Edge-KI-Anwendungen abzubilden und diese in einer abgetrennten Zone zu betreiben.
- **Sicherheitsstandards für KI-Modelle:** Neue bzw. angepasste Standards, die spezifische Sicherheitsprobleme von KI-Modellen wie Datenvergiftung (manipulierte Trainingsdaten), Modellvergiftung (manipulierte Modellgewichte) oder feindliche Angriffe (manipulierte Eingabedaten) adressieren, sollten eingeführt werden. Denn: Edge-KI verstärkt diese Gefahren, weil die Systeme verteilt sind, Hard- und Software physisch nahe beieinander sind und die Angriffsfläche grösser sowie die Angriffsvektoren zahlreicher sind. Solche Sicherheitsprobleme könnten dank verbindlichen Red-Teaming-Tests, also gezielten Angriffssimulationen, vor dem effektiven Einsatz von Edge-KI-Systemen adressiert werden.

Wo freiwillige Initiativen für die Gewährleistung der Cybersicherheit von Edge-KI-Anwendungen nicht ausreichen, sollten Regulierungen die Anbieter zu einem verbindlichen Cybersicherheitsrahmen verpflichten. Hochrisiko-KI-Systeme, also Edge-KI-Systeme, die etwa im Rahmen kritischer Infrastruktur (bspw. im Verkehr oder Strom- und Wassernetz), bei der biometrischen Identifikation oder in der Bildungsevaluation zum Einsatz kommen, sollten zum Grundsatz der Security by Design verpflichtet werden. Dieser Grundsatz verlangt, dass die Genauigkeit, Robustheit und Cybersicherheit von Systemen nicht erst

nachträglich gewährleistet werden, sondern von Beginn weg im Architektur- und Entwicklungsprozess verankert sind. Eine solche Verpflichtung fordert auch Art. 15 im [EU-Gesetz über KI](#).

Da bei Edge-KI-Modellen lokal Daten bearbeitet und Code ausgeführt wird, wird auch der Schutz von Hardware-Komponenten während des Einsatzes immer wichtiger. Um diesen Schutz sicherzustellen, müssen Anbieter vermehrt auf vertrauenswürdige Ausführungsumgebungen (kurz: VAU) setzen. Bei diesen VAU handelt es sich um eine hardwareseitige Abtrennung des Bereichs, auf dem die Daten bearbeitet und der Code des Edge-KI-Modells ausgeführt werden [\[10\]](#).

Die erwähnten VAU sind ein gutes Beispiel dafür, dass im Bereich Edge-KI und Cybersicherheit nach wie vor viele Forschungsfragen offen sind – von den speziellen Anforderungen an Edge-KI-Chips bis hin zur Qualitätssicherung von Edge-KI-Modellen (für eine Übersicht, vgl. [\[2\]](#)). Obschon die Schweiz diesbezüglich bereits ein gewisses Forschungs- und Unternehmensökosystem aufweist (vgl. Box 2), sollte die Forschung in diesem Bereich weiter ausgebaut und der Transfer in die Praxis gezielt gefördert werden.

Fazit

Die rasante Entwicklung im KI-Bereich findet vermehrt auch im Lokalen statt. Die Schweiz hat als hochindustrielles Land eine gute Ausgangslage, im Bereich Edge-KI eine führende Rolle einzunehmen. Damit dies auf sichere und vertrauenswürdige Art und Weise geschieht, sollten die hier diskutierten Handlungsoptionen indes schnell an die Hand genommen werden.



► **Moderner Edge-KI-Computer in einer Industrieumgebung mit Tablet und Smartphone. Symbolbild für sichere, vernetzte und mobile KI-Anwendungen.**

Schweizerische Akademie der Technischen Wissenschaften SATW

Die SATW identifiziert im Auftrag des Bundes industriell relevante technologische Entwicklungen und informiert Politik und Gesellschaft über deren Bedeutung und Konsequenzen. Als Fachorganisation mit hoher Glaubwürdigkeit vermittelt sie unabhängige, objektive und gesamtheitliche Informationen über die Technik – als Grundlage für eine fundierte Meinungsbildung. Die SATW fördert auch das Technikinteresse und -verständnis in der Bevölkerung, insbesondere bei Jugendlichen. Sie ist politisch unabhängig und nicht kommerziell.



satw

Schweizerische Akademie der
Technischen Wissenschaften SATW

St. Annagasse 18
8001 Zürich

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

Impressum

Projektleitung: Tobias Schlegel

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

Redaktion: Esther Lombardini

Grafik: Brandlook GmbH

Bilder: iStockphoto (Quardia), OpenAI (ChatGPT)

Illustrationen: Brandlook GmbH

Übersetzung: Translingua AG

Druck: Egger AG

2026