Lernende Systeme

DIE PLATTFORM FÜR KÜNSTLICHE INTELLIGENZ

Edge AI: KI nahe am Endgerät

Technologie für mehr Datenschutz, Energieeffizienz und Anwendungen in Echtzeit

WHITEPAPER

Ecker, W., Houdeau, D. et al.
AG Technologische Wegbereiter und Data Science
AG IT-Sicherheit, Privacy, Recht und Ethik

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

acatech
DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

Inhalt

Zusammenfassung

Edge AI (Edge Artificial Intelligence) bezeichnet den Einsatz von Künstlicher Intelligenz lokal auf Servern, Fahrzeugen und Robotern oder Endgeräten wie Smartphones statt zentral auf umfangreichen Rechen- und Speicherinfrastrukturen (Cloud AI), wie dies derzeit bei vielen generativen KI-Technologien meist der Fall ist. Dies bietet Vorteile wie geringere Latenzzeiten, Anwendung in Echtzeit, reduzierten Energieverbrauch sowie erhöhte Privatsphäre und Sicherheit, da Daten direkt am Gerät, sozusagen „vor Ort", verarbeitet werden: So können Gesundheitszustände in Echtzeit überwacht werden oder Fahrassistenzsysteme können blitzschnell auf Hindernisse reagieren, was die Sicherheit im Straßenverkehr verbessert.

Zwar gehen mit dem Einsatz von KI „on the Edge" gewisse Herausforderungen einher – beschränkte Rechenleistung sowie Speicherkapazität – damit rücken aber (Energie-)Effizienz sowie auch Nachhaltigkeit ins Zentrum. Über letztgenannte Herausforderung des 21. Jahrhunderts hinaus ist Edge AI vielfältig einsetzbar und damit ein technologischer Baustein, um Herausforderungen wie den Klimawandel, die Digitale Souveränität oder die Energieversorgung zu bewältigen.

Expertinnen und Experten der Plattform Lernende Systeme unter Federführung der Arbeitsgruppen Technologische Wegbereiter und Data Science sowie IT-Sicherheit, Privacy, Recht und Ethik geben mit dem Whitepaper einen Überblick über Edge AI. Dabei beleuchten sie technische, gesellschaftliche und wirtschaftliche Potenziale von Edge-AI-Anwendungen, vergleichen die Edge-AI-Technologie hinsichtlich technischer Komponenten mit der ihr nahestehenden, aber konkurrierenden Cloud-AI-Technologie, zeigen Stärken der Edge-AI-Technologie für Deutschland und Europa auf und diskutieren bestehende wie anzugehende Herausforderungen. Anwendungsbeispiele aus der Automobilbranche, dem Maschinenbau und der Medizintechnik veranschaulichen diese Aspekte, wobei der Sicherheitsaspekt stets im Fokus bleibt. Abschließend werden Gestaltungsoptionen zur Nutzung und Heben des Potenzials von Edge AI formuliert: So sollten beispielsweise die Forschung und Entwicklung sowie die Anwendung von ressourcenbewusster Datenverarbeitung innerhalb der Limitationen von Edge-Geräten weiter vorangetrieben werden und Plattformen bzw. Module für Edge AI als Basisbausteine entwickelt werden, die je nach Bedürfnissen an die Branchen angepasst werden können, um den Transfer in die Anwendung zu erleichtern – was eine entsprechende Standardisierung voraussetzt.

Das Whitepaper basiert hauptsächlich auf den Ergebnissen eines in 2023 durchgeführten „Runden Tisches" zum gleichnamigen Thema sowie einem Workshop, der im Rahmen der Konferenz der Plattform Lernende Systeme im selbigen Jahr stattfand.

1 Einleitung

Die jüngsten Erfolge in der Entwicklung Künstlicher Intelligenz (KI), zum Beispiel KI-Software zur Text- und Bildgenerierung, basieren auf immer größeren, zentral verarbeiteten Datenmengen, immer größeren und skalierbareren neuronalen Netzen sowie auf immer höheren Rechenkapazitäten. Dies ist jedoch nur möglich, wenn auf physikalische Grenzen in Rechenzentren oder Geräten sowie auf sensible Daten keine Rücksicht genommen werden muss. Hinzu kommen vergleichsweise hohe Kosten und ein hoher Energieverbrauch. Parallel dazu findet eine Entwicklung in der Forschung und in der Industrie statt, die einen anderen Ansatz verfolgt – Edge AI, kurz für Edge Artificial Intelligence. Edge AI basiert auf der Idee des Edge Computing. Ziel ist es, Daten möglichst dezentral dort zu verarbeiten und zu analysieren, wo sie entstehen, beziehungsweise möglichst nahe am Ort ihrer Entstehung, also nahe beim Nutzenden. So kann der Einsatz komplexer Rechen-infrastruktur und aufwändiger Datenkommunikation reduziert werden.

Edge Computing kann als der Ort eines Kommunikationsnetzes (z. B. Internet, Funknetz) verstanden werden, an dem die Technik im Maximalfall direkt an das lokale und reale Geschehen bei den Nutzenden angrenzt. Sensoren in Fabrikrobotern erfassen reale Signale vor Ort, Smartphones erkennen den Standort und die Stimme des Nutzenden oder nehmen ein Foto einer Landschaft auf, Sensoren in Autos erkennen die Stärke des Regens und Objekte in der Umgebung oder körpernahe Wearables unterstützen Ärztinnen und Ärzte bei der Diagnose und Behandlung. Edge AI nutzt den Ansatz des Edge Computing und verbindet ihn mit Methoden der Künstlichen Intelligenz. Dabei sollen unter anderem das Training von Modellen, die KI-Inferenz (also zum Beispiel die Berechnung von Vorhersagen und Klassifikationen) und generell die Ausführung von Algorithmen für Anwendungen der Künstlichen Intelligenz möglichst „on the Edge“ stattfinden. Wie sinnvoll die Nähe von KI-Inferenz und KI-Training zu den Nutzenden ist, hängt allerdings von der konkreten Anwendung und deren Einsatzgebiet sowie den jeweiligen Anforderungen an Speicher- und Rechenbedarf ab. Entwickelnden steht ein ganzes Spektrum von Möglichkeiten zur Auswahl (Abbildung 1).

Eine Möglichkeit, Edge AI umzusetzen, ist das verteilte Lernen, das wiederum auf unterschiedliche Weise realisiert werden kann, wie Autorinnen und Autoren der Plattform Lernende Systeme in der Kurzpublikation KI Kompakt „Verteiltes Maschinelles Lernen“ dargestellt haben (siehe: Split Learning, Federated Learning, Swarm Learning, in KI-Kompakt: Verteiltes maschinelles Lernen, Plattform Lernende Systeme, 2022).

2 Edge AI – technische, gesellschaftliche und wirtschaftliche Potenziale

Edge AI ermöglicht in verschiedenen Industrien, Märkten und Anwendungsfeldern Durchbrüche bei der Umsetzungsgeschwindigkeit, beim Datenschutz oder bei möglichen Einsparungen (siehe hierzu auch Abschnitt 2.1 und 3). Dabei steht sie in direkter Konkurrenz zu Cloud AI, die bis dato häufiger verbreitet ist. Daher werden im Folgenden die Vor- und Nachteile der beiden Ansätze beleuchtet. Weiterhin wird auf Bereiche eingegangen, in denen Edge AI als Lösungsoption in Betracht gezogen werden kann. Zunächst wird jedoch der technische Hintergrund genauer betrachtet.

2.1 Technischer Hintergrund und Begrifflichkeiten

Edge AI bezieht unterschiedliche Teilgebiete der Informatik und Elektrotechnik ein. Aus algorithmischer Perspektive ist vor allem die Verarbeitung von Datenströmen möglichst in Echtzeit bei eingeschränkter Rechenleistung der Hardware ein wichtiger Aspekt oder, dass zumindest die maximale Zeit, die zur Berechnung einer Aufgabe auf einer Hardwareplattform benötigt wird, garantiert ist (worst case execution time). Edge AI bedingt dabei eine enge Verbindung von Software und Hardware bis hin zum Software-Hardware-Co-Design.

Dies ist ein Anlass für die Entwicklung vieler neuer KI-Verfahren. Diese Verfahren werden für kostengünstige Hardware, einen möglichst geringen Bedarf an Energie, Speicher und Kommunikation für unterschiedliche Rechnerarchitekturen der lokalen Geräte entwickelt. Klassischerweise basieren Computerprozessoren auf der Von-Neumann-Architektur, bei der ein gemeinsamer Speicher sowohl Programmbefehle als auch Daten enthält, oder auf dem Harvard-Prinzip, bei dem Daten und Programme in der Regel getrennt gespeichert sind. Im Kontext von ressourcenbeschränkten Anwendungsszenarien wie Edge AI wird dies jedoch nicht mehr als Norm vorausgesetzt. Es werden Multicore-Architekturen, die mehrere Prozessoren auf einem Chip realisieren, genutzt, wobei die verschiedenen Cores unter anderem auch nach dem von-Neumann- oder Harvard-Prinzip gestaltet sein können. Zudem kommen Field Programmable Gate Arrays (FPGA) zum Einsatz, das heißt, programmierbare digital integrierte Schaltkreise, in denen („vor Ort“) eine logische Schaltung geladen werden kann. In der Anwendung eignen sich FPGA oft nur für Kleinserien, da diese für die Massenware noch mit zu hohen Kosten verbunden sind. Schließlich sind auch nicht-volatile Speicherung und Prozessoren mit stark eingeschränkter Arithmetik zu nennen. Dies sind nur einige Beispiel aus der Vielfalt an technischen Lösungsoptionen.

Maschinelles Lernen ist im Rahmen von Edge AI nicht auf die derzeit prominenten neuronalen Netze beschränkt. Andere Methoden werden ebenfalls umgesetzt, wie beispielsweise Entscheidungsbäume und -wälder, probabilistische Markov-Modelle, Reinforcement Learning und Clustering (Morik & Marwedel, 2023). Zugleich entstehen für das verteilte maschinelle Lernen neue, sogenannte Machine Learning Sensors, bei denen die lokalen Daten und Modelle auch physisch von den Prozessoren der Anwendung, die eventuell mit einer Cloud kommunizieren, getrennt sind (Stewart et al., 2023).

Wie beim Management von Softwareupdates wird auch bei KI ein Lifecycle-Management benötigt. Bei Edge AI geht dieses aber wegen der engen Verzahnung von Software und Hardware über gewohnte Anforderungen hinaus. Dabei spielen insbesondere auch die (automatische) Zertifizierung und Evaluation der Modellgüte und des Ressourcenverbrauchs der Modelle eine bedeutende Rolle. Modellarchitekturen können sich

beispielsweise gänzlich verändern, sodass die Modellgüte sich zwar insgesamt verbessert, aber eventuell in einigen Teilbereichen verschlechtert. Sensoren können einem Concept Drift unterliegen, wenn sich die Betriebsumgebung verändert und das Modell diese nicht mehr adäquat abbildet (z. B. langsam über die Zeit oder abrupt beim Austausch eines Sensors).[1] Dies kann ebenfalls Einfluss auf die Modellgüte haben. Eine weitere mit Sensoren verbundene Herausforderung kann die unterschiedliche Güte und Varianz der Sensoren selbst sein sowie in manchen Fällen sogar deren Montage, die sich auf die Ergebnisse des KI-Systems auswirken kann.

KURZINFO

Begrifflichkeiten im Kontext von Edge AI

Distributed AI bezieht sich auf die Verarbeitung verteilter Daten, die nicht synchronisiert sind. Lokale Modelle der verteilten Geräte werden über unterschiedliche Verfahrensweisen zu einem globalen Modell zusammengeführt.

Edge AI ist eine Form der KI, die sich darauf konzentriert, KI-Modelle auf Geräten am Rande des Kommunikationsnetzwerkes (Edge) einzusetzen (z. B. Laptops, Mobiltelefone oder Sensoren), um Daten lokal zu verarbeiten. Edge AI kann verteilt (distributed) oder auf einem Rechnerknoten (undivided, as a whole) ausgeführt werden.

Tiny AI beinhaltet die Erstellung kompakter Modelle, die auf ressourcenbeschränkten Geräten ausgeführt werden können, also Geräten, mit denen Entwickelnde bei der Umsetzung von Edge AI häufig konfrontiert sind. Diese KI-Art wird fälschlicherweise oft als Edge AI bezeichnet, auch wenn die Geräte nicht mit dem Internet verbunden sind und sich damit eben nicht am Rand des Netzwerkes – also der Edge – befinden.

Federated Machine Learning ist ein Spezialfall von verteilter KI. Es ist ein KI-Ansatz, in dem lokale Daten und Modelle sich nicht über einen zentralen Knoten aneinander anpassen. Es ist eine Form des maschinellen Lernens, die auch für Edge AI eingesetzt werden kann.

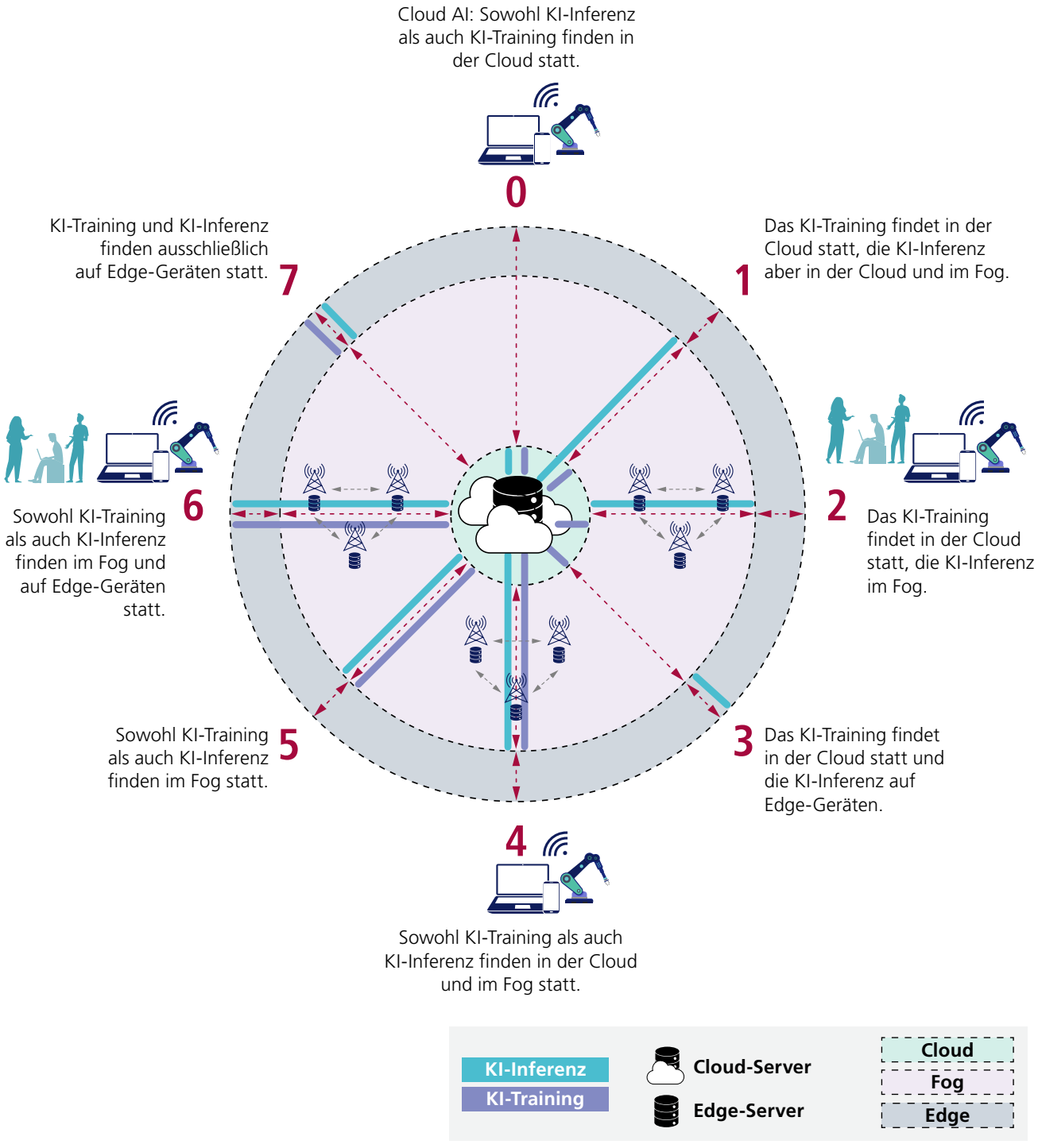
Fog AI bezeichnet gemäß der Idee des „Fog" eine Lösung, die zwischen Edge AI und Cloud AI angesiedelt ist. Fog AI ist in der Regel näher an der Edge angeordnet, zum Beispiel als Supervisor oder lokaler Rechencluster zum Training für mehrere Edge-AI-Knoten.

Ganz allgemein können unterschiedliche Lösungen für die technische Umsetzung von Edge AI gewählt werden (Abbildung 1, 1–7). Das KI-Training und KI-Inferenz können dabei an unterschiedlichen Stellen ausgeführt werden, wie beispielsweise via Cloud-Server oder Edge-Geräten oder zwischen diesen beiden Lösungen via Edge-Server (Fog). Allerdings kann zwischen KI-Inferenz und KI-Training nicht immer eine scharfe Trennlinie gezogen werden. Neuere Konzepte wie „active learning" können auch beide Phasen verbinden. Grundsätzlich ist in der gegenwärtigen Praxis jedoch zu betonen, dass das Training sehr viel mehr Speicherbedarf und Rechenressourcen benötig als die KI-Inferenz. Damit sind in vielen Fällen Lösungen praktikabler, bei

1 „Ändert sich das Einsatzgebiet, das heißt, passt das Modell nicht mehr zu den Daten, mit denen es trainiert wurde, wird die Abweichung zwischen den vom Modell prädizierten und den tatsächlichen (Ausgangs-)Daten so groß, dass das Modell nicht mehr sinnvoll eingesetzt werden kann." (Boll et al., 2022). Dieses Problem ist als Concept Drift bekannt.

denen das Training in der Cloud und die KI-Inferenz auf dem Edge-Gerät stattfindet (Abbildung 1, 2–3). Ein Beispiel macht dies deutlich: Wird ein neuronales Netzwerk etwa zur Geräuschlokalisierung trainiert, kann dieses aus Kostengründen auf 16 Kilobyte komprimiert werden. Die hierfür benötigten Trainingsdaten können jedoch 160 Megabyte umfassen, was für viele Anwendungen auf Edge-Geräten zu viel sein kann, wobei dies auch auf das Edge-Gerät ankommt (vgl. Apple iPhone mit M2-Prozessor versus Arduino).

Abbildung 1: **Unterschiedliche Verlagerung von Inferenz und Training für Edge AI**

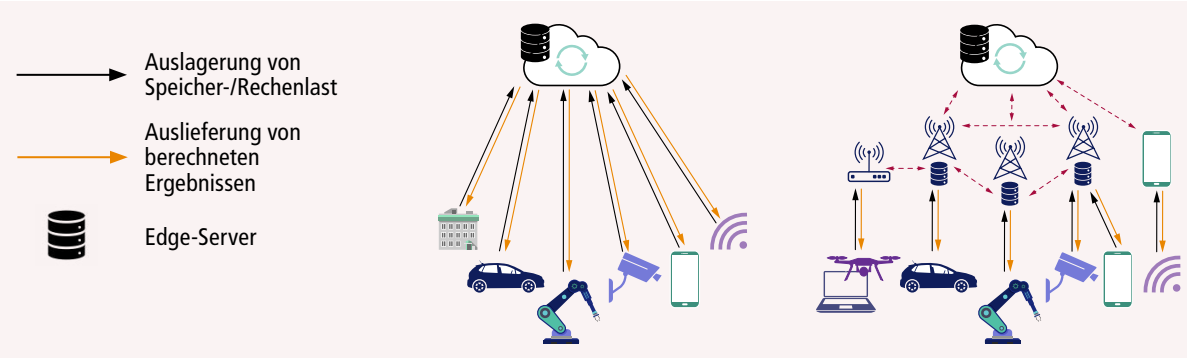


Quelle: Eigene Visualisierung basierend auf McEnroe et al. (2022). Als Edge-Server können auch Laptops, Mobiltelefone oder Bordcomputer und andere Geräte dienen (z. B. als Edge-Server für Sensoren).

2.2 Cloud AI und Edge AI – Vor- und Nachteile basierend auf technischen Unterschieden

Cloud AI und Edge AI weisen aufgrund ihrer jeweiligen technischen Basis unterschiedliche Vor- und Nachteile auf. Dies wird deutlich, wenn die beiden Technologien zu Kategorien wie Rechenleistung, Datenschutz, Sicherheit und Latenz miteinander verglichen werden (Tabelle 1). Aber auch bei Edge AI selbst sind die Vor- und Nachteile nicht immer eindeutig, wie Abbildung 2 zusammenfasst.

Tabelle 1: **Cloud AI und Edge AI: Vor- und Nachteile basierend auf technischen Unterschieden**

	Cloud AI	Edge AI
		
Siehe Abbildung 1:	Fall 0	Fall 1 bis 7
Rechenleistung, Speicherplatz	⊕ Viel Speicherkapazität, hohe Rechenleistung	⊖ Speicher-/Rechenleistung durch Limitationen von Geräten begrenzt
Sicherheit, Privatheit	⊖ Die Angriffsfläche ist geringer, aber ein gelungener Angriff kann erhebliche Folgen haben (vgl. Denial-of-Service-Attacke, Veröffentlichung umfangreicher Datenmengen etc.).	⊕ Daten werden verteilt verarbeitet, daher sind die Folgen eines erfolgreichen Angriffs geringer. ⊕ Daten werden über kürzere Distanzen versendet, sodass die Wahrscheinlichkeit, diese abzufangen, geringer ist.
Latenz	⊖ Längere Datenübertragungszeiten	⊕ Kurze Übertragungswege/-zeiten
Verlässlichkeit	⊖ Wenn der zentrale Server ausfällt oder nicht erreichbar ist, kann das den Anwendungsbetrieb stören.* ⊖ Die (i.d.R. variable) Datenübertragungszeit hat Auswirkung auf das Echtzeitverhalten.	⊕ Alternative, nahe liegende Edge-Server könnten einen Ausfall oder u. U. auch die Kapazitäten der Geräte selbst kompensieren. ⊕ Garantien für Echtzeitverhalten können gegeben werden, da nur die lokale Rechenzeit zu berücksichtigen ist.
Overhead der Kommunikation	⊖ Viele Daten müssen über große Distanzen hinweg übertragen werden.	⊕ Wenige Daten werden in kürzeren Zeitabständen übertragen.
Datenvollständigkeit	⊕ Eher vollständige Datenlage	⊖ Lokalen Modellen fehlen ggf. viele Daten, was ihre Qualität beeinträchtigen kann.
Anwendungen	z. B. Bild-, Film- und Musikgenerierung, Sprachroboter	z. B. Mensch/Maschine-Kommunikation, Geräte- und Maschinenüberwachung, Home Automation, effiziente Motorsteuerung, Abwehr virtueller Angriffe aus dem Netz, Regelung

Quelle: Eigene Zusammenstellung basierend auf McEnroe et al. (2022).
*In der Praxis werden kritische Anwendungen durch ganze Datenzentren gewährleistet, sodass der Ausfall eines Servers kompensiert werden kann.

Energieeffizienz und Nachhaltigkeit

Die Ressourcenbegrenzung in Edge-Geräten erfordert nicht nur energieeffiziente KI-Systeme, sondern auch kleine gelernte Modelle. Zumindest die Ausführung der Modelle muss unter starker Ressourcenbeschränkung möglich sein. Diese Einschränkungen sind jedoch nicht nur eine Herausforderung, sondern auch eine große Chance, da sie Treiber für ressourcenschonende KI-Innovationen sind. Sie haben zum Beispiel zu einem erhöhten Interesse an binarisierten neuronalen Netzen geführt (Buschjäger et al., 2021).

Insofern kann Edge AI als Vorbild für Cloud AI dienen, wie sie derzeit für große generative KI-Modelle eingesetzt wird. Das Training solcher KI-Modelle ist sehr energieintensiv und führt zu hohen CO_2-Emissionen, wenn Rechenzentren und Herstellungsprozesse für Hardware und Infrastruktur nicht mit erneuerbaren Energien betrieben werden. Neben den Prozessoren für die Rechner werden Strom für die Kühlung und das Energiemanagement benötigt, was beim Training eines Modells wie ChatGPT 3 zu einem Verbrauch von zehn Gigawattstunden führen kann, das heißt dem jährlichen Stromverbrauch von 1.000 US-Haushalten (McQuate, 2023). Ein Bildgenerator wie DALL-E oder Stable Diffusion benötigt beim Erstellen eines Bildes so viel Energie, wie für das Aufladen eines Handyakkus notwendig ist (Heikkilä, 2023). Daher wird nicht nur die Stellschraube der Energieerzeugung, sondern auch die Stellschraube der Energieeffizienz in der Forschung zu maschinellem Lernen beachtet (Piatkowski et al., 2016). Messungen zur Bestimmung des CO_2-Fußabdrucks von maschinellem Lernen werden seit einigen Jahren diskutiert (Henderson et al., 2020).

Zwei Beispiele machen den Punkt des Energiemanagement bei Edge AI vor allem deutlich:

Datenfilterung in der Edge: Datenfilter können exemplarisch am Beispiel der Anomalieerkennung bei Herzschrittmachern illustriert werden. Liegen Herzschläge im normalen Bereich, werden diese in der Edge weder aufgezeichnet noch gespeichert. Diese Daten liefern keinen Mehrwert, benötigen jedoch Speicherkapazität und Energie, wenn sie an ein Anzeigegerät übertragen werden sollen. Bei der Anomalieerkennung werden nur Daten analysiert, gesammelt und gespeichert, die jenseits des Normalbereichs liegen. Diese Daten geben Hinweise über den Gesundheitszustand der Tragenden beziehungsweise des Tragenden eines Herzschrittmachers.

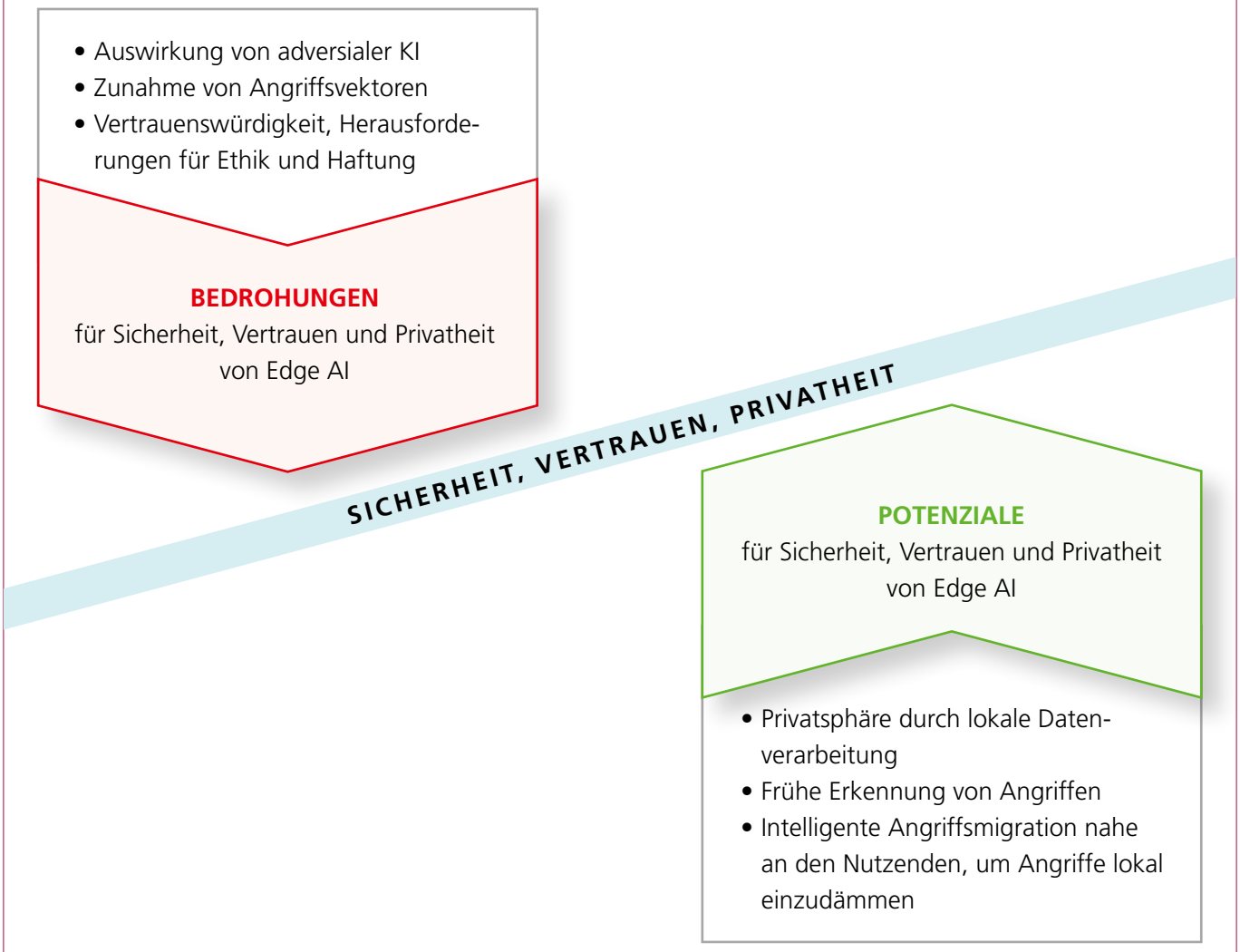
KI zur Steuerung der verwendeten Algorithmen in der Edge: Einige mobile Anwendungen verwenden mehrere KI-Algorithmen und -Anwendungen, wie etwa Autos oder Smartphones. Sie müssen aber mit der verfügbaren Energie (z. B. Batterie, Akkumulator) möglichst lange auskommen. So kann zum Beispiel eine Anwendung in einem teilautonomen E-Fahrzeug Müdigkeit erkennen und eine weitere den Wechsel einer Fahrspur anzeigen. Wenn das Fahrzeug an einer Ampel steht, können diese KI-Systeme und deren Sensoren ausgeschaltet werden, um die Reichweite eines E-Autos zu vergrößern. Im leistungsfähigen Smartphone werden heute rund zehn KI-Algorithmen eingesetzt. Würden alle KI-Systeme permanent in Betrieb gehalten werden, würde sich die Nutzungsdauer des Smartphones signifikant reduzieren.

Sicherheit und Privatsphäre

Diese positiven, technischen Eigenschaften von Edge AI im Bereich der Sicherheit und Privatsphäre, nämlich der geringen Latenzzeiten, der Zuverlässigkeit und des vergleichsweisen geringen Kommunikations-Overheads, bieten große Chancen im Hinblick auf das Angebot sicherer und verlässlicher Dienste und Datenanalysen in Echtzeit sowie Einsatzmöglichkeiten dort, wo es um besonders sensible Daten (z. B. Gesundheitsdaten) geht. Allerdings sind Datenschutz und Sicherheit nur so lange ein Vorteil, wie die Edge-Geräte selbst vor dem Zugriff von Unbefugten geschützt sind. Denn je mehr Geräte an KI-Inferenz und KI-Training beteiligt

sind, desto mehr Angriffsvektoren können entstehen. Damit können auch einzelne Geräte für Angreifende einfacher zugänglich werden, um Attacken (z. B. Seitenkanalattacken) durchzuführen, die es ermöglichen, Modellarchitekturen und -parameter zu ermitteln, geistiges Eigentum offenzulegen oder die Vertraulichkeit von Information und die Privatsphäre zu bedrohen.

Abbildung 2: **Bedrohung versus Potenzial für Sicherheit, Vertrauen und Privatheit von Edge AI**



Quelle: Visualisierung und Zusammenstellung basierend auf Ding et al. (2021) [Übersetzung d. Red.].

2.3 Gegenwärtige Herausforderungen mit Edge AI als Instrument angehen

Die positiven Eigenschaften von Edge AI können genutzt werden, um zur Überwindung verschiedener gesellschaftlicher und wirtschaftlicher Herausforderungen beizutragen – und damit auch zu bedeutenden Transformationsvorhaben des 21. Jahrhunderts. Die Energieversorgung, die Umwelt und der Klimawandel werden in Bevölkerungsumfragen der letzten Jahre häufig als drängende Probleme in Deutschland genannt, vor dem Hintergrund der Corona-Pandemie auch Gesundheit (Statista, 2023). Auch Unternehmen nennen Energieversorgung, Klimawandel und Pandemien als wichtige anzugehende Probleme, identifizieren aber auch Produktions- und Lieferengpässe sowie Cybervorfälle als große Herausforderungen, ebenso den Fachkräftemangel (Crunchtime, 2023).

Energieversorgung

In intelligenten Stromnetzen können Prognosen über Stromerzeugung und -verbrauch erstellt werden, die ein effizientes Lastmanagement im Stromnetz ermöglichen. Dies ist insbesondere für die Energiewende hin zu einem auf erneuerbaren Energien basierenden Energiesystem von Bedeutung, da die Energieerzeugung wetterbedingt schwanken kann und immer mehr Energieverbrauchende auch zu Energieerzeugenden werden (z. B. via Solarmodulen auf privaten Hausdächern). Edge AI kann hier auf der Ebene von intelligenten Stromzählern (Smart Meters) in den Häusern und einzelnen Windkraftanlagen oder Solarmodulen ansetzen, um das notwendige Monitoring und die nötigen Vorhersageleistungen zuverlässig und in Echtzeit zu liefern. Dies ermöglicht ein effektives Lastenmanagement, indem überschüssige Energie in Zeiten geringer Nachfrage in Energiespeicher eingespeist oder in Zeiten hoher Nachfrage effizient verteilt wird. In Kombination mit Cloud-basierten Lösungen kann diese Optimierung auch für ganze Flotten elektrischer Fahrzeuge ausgeweitet werden, um Lastspitzen zu reduzieren.

Umwelt und Nachhaltigkeit

In einer Kreislaufwirtschaft sollen Stoff- und Energiekreisläufe so optimiert werden, dass sie möglichst geschlossen und ökologisch sinnvoll sind (acatech, 2023). Edge AI kann auf Basis von Kamerasystemen und Sensoren in Recycling- oder Abfallbehandlungsanlagen eingesetzt werden, um verschiedene Wertstoffe im Abfall in Echtzeit zu erkennen und zu analysieren. In Verbindung mit Aktoren kann der Abfall dann automatisch sortiert und verschiedenen Weiterverarbeitungsprozessen zugeführt werden. Dies kann die Effizienz von Recyclingprozessen steigern und so einen wichtigen Beitrag zur Kreislaufwirtschaft und damit zur Nachhaltigkeit leisten.

Klima- und Katastrophenlagen

Durch den Klimawandel sind Extremwetterlagen wahrscheinlicher geworden. Hierdurch ergeben sich Herausforderungen für den Katastrophenschutz. KI-Technologie kann bei der Bewältigung solcher Herausforderungen unterstützen.

(1) Das europäische Projekt INSIGHT untersuchte in enger Zusammenarbeit mit der Stadt Dublin, wie bei Hochwasser die Früherkennung, die Kommunikation der Helfenden, die Verkehrslenkung und die örtliche Kommunikation durch maschinelles Lernen verbessert werden kann (Kinane et al., 2014). Heterogene Informationsquellen aus sozialen Medien, realzeitlichen Verkehrsdaten, Rundfunkanrufen, Pegelstandsmessungen wurden mit Glaubwürdigkeitstests zu einem Gesamtbild zusammengefasst. Die Ergebnisse wurden im Projekt VAVEL verallgemeinert und auf Warschau übertragen.

(2) Bereits bei der Hochwasserkatastrophe im Ahrtal (2021) wurden Drohnen eingesetzt. Die von den Drohnen erfassten Daten dienen der automatisierten Bilderkennung mithilfe von KI-Algorithmen und damit der Erstellung von Lagebildern (z. B. Überblick über Schäden, Veränderung der Hochwassersituation etc.). Solche Lagebilder können wiederum die Lageplanung und Koordination von Rettungskräften verbessern (DLR, 2022; McEnroe et al., 2022). Katastrophenlagen sind oft durch schwierige Bedingungen für die digitale Kommunikation gekennzeichnet, sodass die Konnektivität schwanken kann. Die Übertragung von Daten an zentrale Server wäre daher anfälliger für Störungen. Edge AI kann hier einerseits helfen, zuverlässige Echtzeit-Lagebilder zu erstellen, sodass im Einsatz KI-Modelle zur Lageerfassung lokal auf verschiedenen Drohnen ausgeführt und weiter trainiert werden können. Auch hier können die verschiedenen Teilmodelle wiederum zu einem effizienteren Gesamtmodell aggregiert werden, das dann wiederum lokal ausgeführt werden kann und so die Krisenkartierung noch weiter verbessert.

(3) Edge AI kann aber auch präventiv wirken, wie das Beispiel möglicher Felsstürze infolge des Klimawandels zeigt. So haben Forschende der ETH Zürich am Matterhorn ein drahtloses Sensornetzwerk installiert, das den Zustand von Felsgestein und Permafrost untersucht. Die einzelnen Sensoren können dabei auf Basis von KI selbst entscheiden, ob ein Ereignis relevant ist oder nicht (Rüegg, 2019).

Ökonomisches Potenzial

In Zeiten geringen Wirtschaftswachstums und hoher Energiekosten ist es besonders wichtig, sich damit zu beschäftigen, wie durch den Einsatz von Technologie Wettbewerbsfähigkeit hergestellt und wirtschaftliche Potenziale erschlossen werden können. Edge AI kann hier unterstützen, indem sie unter anderem die Automatisierung und Optimierung von Prozessen in Produktion und Logistik erhöht oder durch vorausschauende Wartung die Wahrscheinlichkeit von Produktionsausfällen aufgrund von Materialermüdung oder Ähnlichem reduziert (siehe hierzu auch Smart Factory/Industrie 4.0; Anwendungsfall 2: Maschinenbau – Föderiertes Lernen in der Robotik. In der Produktion bieten die verketteten Prozesse mit ihren sequenziellen Fertigungsstationen zudem die Möglichkeit, frühzeitig anhand der lokalen Auswertung an einer Station festzustellen, ob die gewünschte Qualität noch erreichbar ist: So wird gegebenenfalls der Prozess direkt verändert oder es wird nicht weiter in einen Prozess investiert (Deuse et al., 2022).

Darüber hinaus ergeben sich vielversprechende Anwendungsmöglichkeiten im Bereich des autonomen Fahrens (siehe Anwendungsfall 1: Automobilbranche – Vehicle-to-Vehicle-Kommunikation; Krieger et al., 2022) sowie, wie bereits dargestellt, im Energie- und Gesundheitssektor (siehe Anwendungsfall 3: Medizinsektor – Lokale Patienten-App). Weiterhin kann Edge AI für Unternehmen, die ihre sensiblen Rohdaten nicht teilen und extern verarbeiten lassen wollen oder allein nicht über genügend Daten verfügen, überhaupt erst die Voraussetzung schaffen, um von den derzeit rasanten Fortschritten im Bereich der Künstlichen Intelligenz profitieren zu können. Diese Potenziale spiegeln sich in Prognosen von Marktforschungsunternehmen wider. So erwartet das Unternehmen Gartner eine steigende Nachfrage nach Edge AI: Bis 2025 sollen mehr als 55 Prozent aller auf tiefen neuronalen Netzen basierenden Datenanalysen in einem Edge-System stattfinden, gegenüber weniger als zehn Prozent im Jahr 2021 (Gartner, 2023a).

Digitale Souveränität

Die Pandemie, aber auch der Krieg Russlands gegen die Ukraine haben gezeigt, wie fragil Lieferketten sein können und welche Herausforderungen einseitige Abhängigkeiten mit sich bringen. Hinsichtlich der IT-Infrastruktur ist in Deutschland ein deutlicher Trend zu erkennen: 75 Prozent der Unternehmen nutzen externe Rechenkapazitäten (IDC, 2022), um ihre digitalen Operationen effizienter zu gestalten. Die Herausforderung liegt jedoch in der zunehmenden Abhängigkeit von großen Cloud-Anbietern, die diese Kapazitäten bereitstellen, wobei viele Anbieter außerhalb Europas angesiedelt sind. Diese Abhängigkeit wirft Bedenken hinsichtlich der Kontrolle über sensible Daten, der Zuverlässigkeit des Betriebs kritischer Anwendungen (Tabelle 1) und somit der digitalen Souveränität auf. Hier setzt das Konzept der Edge AI an, das darauf abzielt, Datenströme näher an der Quelle zu verarbeiten und damit den Bedarf an großen Rechenzentren zu reduzieren. Dies führt auch insofern zu einer größeren digitalen Souveränität, als die eingesetzte Hardware nicht zwangsläufig auf modernster Halbleitertechnologie basieren muss. Dies eröffnet die Möglichkeit, Hardware vermehrt in Deutschland beziehungsweise in Europa zu produzieren und damit die Abhängigkeit von externen Zulieferern sowie auch von Störungen in der Chip-Lieferkette zu reduzieren. Auf diese Weise könnte auch Deutschland von diesem Wachstumsmarkt profitieren. Bei kleinen, spezialisierten KI-Beschleunigern für die Edge ist es zudem für deutsche und europäische Mitbewerber vergleichsweise einfacher, im Wettbewerb zu bestehen und den Vorsprung aufzuholen, als dies bei KI-spezifischen GPU-Prozessoren der Fall ist, also bei solchen, bei denen Marktanbieter mit Hauptsitz in den USA derzeit den Markt dominieren (Marktanteil von 80 Prozent, siehe Reuters, 2023).

Individuum

Nicht nur die Gesellschaft und Wirtschaft können von Edge AI profitieren. Vielmehr ermöglicht die Technologie auch die Personalisierung und Anpassung von Dienstleistungen an individuelle Bedürfnisse unter Wahrung von Datenschutz und Privatsphäre (Heppe et al., 2020). Nutzende müssen nicht mehr befürchten, dass ihre persönlichen und sensiblen Daten in entfernten Rechenzentren verarbeitet und zur Profilbildung genutzt werden. Stattdessen können Daten lokal verarbeitet und KI-Modelle beispielsweise auf dem privaten Tablet oder Mobiltelefon betrieben werden (Abbildung 1, Fall 7). Vom persönlichen KI-Assistenten bis hin zur Gesundheits-App kann dies die Dienste und Empfehlungen für den Einzelnen verbessern und gleichzeitig den Datenschutz von besonders kritischen personenbezogenen Informationen entsprechend der Datenschutzgrundverordnung (DSGVO) wahren.

Weitere Edge-AI-Projekte in verschiedenen Bereichen (Auswahl)

Titel	Bereich	Ziele (Auswahl)
Clever – Edge-Cloud Umgebung und Künstliche Intelligenz in elektronischen Systemen für Industrieanwendungen	Agrarsektor, Industrie und Produktion, Logistik, Handel	Latenzzeiten verkürzen, Energieeffizienz, technologische Souveränität
PLATON – Verteilte Rechenplattform für radarbasierte 3D-Umgebungserfassung im sicheren autonomen Fahren	Mobilität	Verlässlichkeit, Sicherheit, Energieeffizienz, Datenschutz
ImaB Edge – Intelligente, multimodale und autarke Bauwerksprüfung mittels Edge Computing	Bauen und Infrastruktur	Vorausschauende Wartung, Strukturmonitoring, Datenvorverarbeitung und -fusion
M/EDGE – Secure Low Power Medical Edge Computing	Gesundheit	Echtzeitfähigkeit, Verlässlichkeit, Datentransfer minimieren, Energieeffizienz
FloW – Föderales Lernen für bildgestützte optische Systeme im Wassermanagement	Klima-Umwelt-Nachhaltigkeit	Anomalieerkennung, Klassifikation von Verschmutzung
FAIRe – Ressourcenbewusste Edge AI ermöglicht KI-Anwendungen auf mobilen Geräten	Mensch-Roboter-Interaktion	Energieeffizienz, Privacy und Datenschutz, Verlässlichkeit, Datentransfer minimieren
NeuSPIN – Entwurf eines zuverlässigen neuromorphen Edge-Systems basierend auf Spintronik für Green AI	Hardwareentwicklung	Energieeffizienz, KI-Algorithmen On-Chip ausführen

Quellen: Zusammenstellung auf Basis der Projektbeschreibungen. In der Spalte „Ziele“ werden einige der in den Beschreibungen genannten Punkte ausgewiesen. Bei der Auswahl der Projekte wurde das Kriterium der Vielfalt berücksichtigt.

3 Stärken Deutschlands und Europas sowie Herausforderungen für den Transfer

Laut dem Marktforschungsunternehmen Gartner (2023b) befindet sich Edge AI derzeit im „Trough of Disillusionment", welches den Tiefpunkt des Hype Cycles darstellt. Es wird jedoch erwartet, dass das „Plateau of Productivity" der Technologie in weniger als zwei Jahren erreicht wird. Dies muss jedoch differenziert betrachtet werden, denn je nach Bereich hat die Technologie dieses Plateau bereits erreicht. In den Bereichen Maschinenbau, maschinelles Sehen sowie Sensorik ist Edge AI weit fortgeschritten und etabliert. Es zeigt sich also, dass sich eine realistische Erwartungshaltung hinsichtlich der Chancen sowie eine technologische Reife etabliert hat, die nahe an der Umsetzung von Edge-AI-Lösungen in der Praxis liegt. Die Verbindung von in Deutschland und Europa vorhandenem Wissen und Know-how sowie der vor Ort vorhandenen Industrien mit europäischen Werten kann als Stärke des Standortes genutzt werden. Hier sind jedoch auch Transferhemmnisse zu überwinden.

Vorhandenes Wissen und Know-how nutzen

Die Kombination aus vorhandenem technologischen Wissen, Domänenexpertise und Erfahrung in der Entwicklung physischer Produkte positioniert Deutschland ideal, um das Potenzial von Edge AI auszuschöpfen. Der Grundstein für den Erfolg von Edge AI ist bereits gelegt, da in Deutschland und Europa viel Expertise, Know-how und exzellente Forschung auf diesem Gebiet vorhanden sind. So arbeiteten in Deutschland im Sonderforschungsbereich „Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung" über zwölf Jahre hinweg Forschende aus maschinellem Lernen, eingebetteten Systemen, Kommunikationsnetzwerken und Datenbanken eng zusammen. Gemeinsam mit anderen Wissenschaften (Medizin und Physik) sowie mit realen Anwendungen in Medizin, Verkehr, Logistik und Produktion wurden Methoden für unterschiedliche Sensoren und Szenarien entwickelt.

Im Lamarr-Institut für Maschinelles Lernen und Künstliche Intelligenz wird der Bereich des ressourcenschonenden maschinellen Lernens weitergeführt und in Anwendungen in der Praxis eingesetzt (u.a. mit einem Pumpenhersteller). Projekte zum föderierten Lernen in der Robotik tragen die Edge-AI-Technologie in die industrielle Praxis (siehe Projekt Federated Learning for Robot Picking). Und das von der Europäischen Union geförderte Exzellenznetzwerk „dAIEDGE" unter Leitung des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) soll weiterhin die Vernetzung der Community vorantreiben, Projekte initiieren sowie Ideen, Tools, Services, Guidelines und Trends bereitstellen. Zu nennen ist auch das BMBF-geförderte Projekt Scale4Edge – Skalierbare Infrastruktur für Edge-Computing – in dem unter anderem drei sehr kleine, aber leistungsfähige KI-Beschleuniger mit unterschiedlichen PPA-Kennzahlen (Power, Performance, Area) entwickelt wurden.

Allerdings ist auf dem Gebiet des analogen Hardwaredesigns ein Mangel an verfügbarer Expertise durch Talente festzustellen (Stichwort: neuromorphes Computing). Dies bremst die Innovation bei Edge AI. Denn analogen Schaltungen wird ein großes Potenzial für energieeffiziente KI-Inferenz zugeschrieben. Allgemein sind sinkende Studierendenzahlen in der Elektrotechnik und fehlende Schwerpunkte bei analogen und digitalen Schaltungen in den Curricula zu verzeichnen. Waren im Wintersemester 2017/2018 noch 69.634 Studierende für das Fach eingeschrieben, waren es im Wintersemester 2022/2023 dagegen nur noch 62.875, damit unter dem Stand des Wintersemesters 2011/2012 mit 62.927 (Statistisches Bundesamt, 2023). Dies trägt zum genannten Fachkräftemangel bei und stellt ein Hemmnis für Forschung und Entwicklung dar.

Darüber hinaus ist beispielsweise in den in Deutschland führenden Industrien wie der Automobilindustrie, der Medizintechnik und der Elektrotechnik/Elektronik viel Domänenexpertise vorhanden, die ihr spezifisches Wissen in die Entwicklung und Umsetzung von Edge-AI-Anwendungen einbringen kann. Dies ermöglicht nicht nur die Entwicklung branchenspezifischer Lösungen, sondern bietet auch die Chance einer schnelleren Anpassung an die Anforderungen unterschiedlicher Branchen.

Schließlich verfügt die deutsche Industrie über langjährige Erfahrungen bei der Einführung konkreter physischer Produkte. Dies spiegelt sich unter anderem im Export wider. Als wichtigste deutsche Exportgüter sind an erster Stelle Kraftwagen und Kraftwagenteile zu nennen und an zweiter Stelle Maschinen. Datenverarbeitungsgeräte befinden sich auf dem vierten Platz (Destatis, 2023). Diese Erfahrung ist eine wichtige Wissensressource für die erfolgreiche Umsetzung von Edge-AI-Lösungen, da die deutsche Industrie mit den Herausforderungen der Produktentwicklung, Herstellung und Implementierung bestens vertraut ist.

Vorhandene industrielle Strukturen nutzen

Ein Vorteil für den Einsatz von Edge AI liegt vor allem in der räumlichen Nähe zu führenden Industrien wie der Automobilindustrie, Medizintechnik und Elektrotechnik/Elektronik. Diese Nähe ermöglicht eine schnellere Anwendung der Edge-AI-Technologie und eine breitere Akzeptanz in relevanten Branchen. Unter Nutzung der verschiedenen benannten Wissensressourcen können sich Unternehmen in Deutschland als Problemlöser für komplexe technische Unternehmungen auf Basis von Edge AI positionieren.

Europäische Werte als Marke nutzen

Unternehmen können die Landschaft europäischer Werte und europäischen Rechts als einen Wettbewerbsvorteil im Sinne von „KI made in Europe“ nutzen. Es gilt die technischen Eigenschaften von Edge AI zu nutzen, um sichere, verantwortungsvolle und datenschutzfreundliche KI-Anwendungen anzubieten.

Herausforderungen für den Transfer

Forschung und Entwicklung sind jedoch mit erheblichen Einschränkungen konfrontiert, da die Rechenkapazität und die Energieressourcen kleiner Geräte begrenzt sind und auch die Kosten tragbar bleiben müssen. Dies erfordert unter anderem KI-Modelle, die in ihrer Größe diesen Anforderungen Rechnung tragen (vgl. Tiny ML). Haben Forschende eine sinnvolle, technische Lösung für Edge AI gefunden, ist es immer noch eine große Herausforderung, um diese Lösung herum ein System aufzubauen, das allen Anforderungen gerecht wird. Dies erfordert einen ganzheitlichen Blick: KI, Software und Hardware müssen gemeinsam entwickelt werden, das heißt, dass Synergien zwischen Soft- und Hardware genutzt und erzeugt werden, um spezifische Ziele auf der Ebene des Gesamtsystems zu erreichen. So müssen Lösungen für spezifische Geräte, zum Beispiel für Ultra-Low-Power-Geräte, gefunden werden (Piatkowski et al., 2016). Lernalgorithmen müssen spezifisch für bestimmte Hardware neu entwickelt werden (Luk et al., 2022). Die Vielfalt der Implementationen und ihre ständige Weiterentwicklung führen zu einer Fülle an Programmbibliotheken, die schwer zu verbinden sind und schnell durch Updates überholt werden. Auf diese Weise sind Lösungen oft nicht leicht zu reproduzieren. Dies ist auch ein Grund dafür, dass die Anwendung in der Praxis schwierig ist. Es muss für eine Anwendung jeweils eine stabile Basis etabliert werden, die Hardware und Software verbindet. Zudem fehlt es in vielerlei Hinsicht an anerkannten und verbreiteten Benchmarks, weshalb Forschungsergebnisse von Endanwendenden oft nicht ernst genommen werden. Auch dies erschwert den Transfer in die Anwendung.

Die Herausforderung der sektoralen Anpassung soll anhand von drei Beispielen näher erläutert werden: der Automobilindustrie, dem Maschinenbau und der Medizintechnik. Dabei handelt es sich um Branchen, die für den deutschen Export von besonderer Bedeutung sind. Anhand der Beispiele wird jeweils auf die Chancen von Edge AI für die Branche sowie auf die branchenspezifischen Rahmenbedingungen eingegangen, die bei der Umsetzung von Edge AI herausfordernd sein können. Abschließend wird jeweils ein besonderes Augenmerk auf den Sicherheitsaspekt gelegt, der für das Vertrauen in den Einsatz dieser Technologie von zentraler Bedeutung ist.

ANWENDUNGSFALL 1

Automobilbranche

Vehicle-to-Vehicle-Kommunikation

Vehicle-to-Vehicle-Kommunikation ist eine Lösungsoption für das autonome Fahren oder für geringere Autonomiestufen. Durch Sensordaten des Autos (z. B. Lidar, Kameras, Radar) und Verkehrsdaten, die über Kommunikationsnetze zwischen Fahrzeugen ausgetauscht werden, können lokale KI-Modelle zum Einsatz kommen, die zuverlässig und in Echtzeit diese eingehenden Daten verarbeiten und so Anomalien oder auch mögliche Gefahrensituationen erkennen (Sliwa et al., 2021). So können Warnungen kommuniziert werden oder gar Maßnahmen eingeleitet werden, um ein Unfallrisiko zur vermeiden, wie beispielsweise Abstände vergrößern oder ausweichen.
Weiteres Beispiel: Kontinuierliche Verbesserung von Bilderkennungs-Algorithmen beim autonomen Fahren (siehe KI-Kompakt „Verteiltes maschinelles Lernen“, Plattform Lernende Systeme, 2022).

Herausforderungen für die sektorale Anpassung von Edge-AI-Lösungen sowie Cybersicherheit

In der Automobilbranche gelten strenge gesetzliche Vorgaben und die Zulassung von Fahrzeugen ist sehr teuer. Eine weitere Herausforderung besteht darin, dass sich viele Firmen im Bereich des autonomen Fahrens nicht genügend abstimmen und zu wenig miteinander kommunizieren. Neben solchen sektorspezifischen Herausforderungen können auch Cyberangriffe erfolgen, denn je höher der Automatisierungsgrad im Straßenfahrzeug angestrebt wird, desto mehr Funktionen werden durch Sensoren und Steuergeräte mit KI-Unterstützung benötigt. Dies erhöht die Intelligenz des Fahrzeugs, birgt aber gleichzeitig auch eine erhöhte Gefahrenoberfläche für Cyberangriffe, intrinsisch zum Beispiel auf drei Komponenten, mit Sensoren (z. B. Kamera), Kommunikationskanal (z. B. CAN-Bus) und Steuergerät, aber auch extrinsisch bei Over-the-Air (OTA)-Kommunikationen. Mögliche Angriffswerkzeuge reichen von der Code-Injection-Technik bis zum klassischen Ransomware-Angriff (siehe hierzu den ENISA-Bericht, Dede et al., 2021).

ANWENDUNGSFALL 2

Maschinenbau

Föderiertes Lernen in der Industrierobotik

Edge AI kann auf Basis des föderierten Lernens so umgesetzt werden, dass Kommissionierroboter in der Lage sind, Objekte und Greifpunkte zu erkennen und geeignete Greifverfahren für die Objekte auszuwählen. Dieser Ansatz ermöglicht es den Robotern, voneinander zu lernen, auch unbekannte Objekte zuverlässig zu greifen und dadurch schneller neue Aufgaben zu übernehmen. Die entsprechende kritische Masse an Daten für das Training des KI-Modells stammt unter anderem von Kameras in den Kommissionierzellen der Roboter und kann lokal oder eingeschränkt über Standort- und Unternehmensgrenzen hinweg ortsnahe verarbeitet werden, sodass lediglich die Parameter der lokalen KI-Modelle zentral zu einem globalen Modell aggregiert werden. Auf diese Weise entsteht ein effizienteres Modell, ohne dass sensible Unternehmensdaten oder Betriebsgeheimnisse geteilt werden müssen, wodurch die Zusammenarbeit zwischen Unternehmen erleichtert wird und die einzelnen Unternehmen von leistungsfähigeren Robotern und verbesserter Automatisierung profitieren können. Zusätzlich werden Mitarbeiterinnen und Mitarbeiter bei repetitiven, schweren und ermüdenden Tätigkeiten unterstützt (siehe Projekt Flairop, Ciupek, 2023).

Herausforderungen für die sektorale Anpassung von Edge-AI-Lösungen und Sicherheitsaspekte

Im Bereich des Maschinenbaus müssen Systeme international kompatibel sein und bestimmte Standards erfüllen, um anerkannt zu werden. Es kann jedoch vorkommen, dass aufgrund der inhärenten Intransparenz sowie des möglichen nicht-deterministischen Verhaltens einiger vielbesprochener Algorithmen eine Zertifizierung dieser Algorithmen schwierig wird. Auch Sicherheitsaspekte (Security und Safety) können bei Maschinen, wie etwa den Robotern im obigen Beispiel, eine Rolle spielen. Kommissionierroboter sind Teil der Intralogistik in einer Smart Factory. Die Intralogistik aus wirtschaftlichen Erwägungen ständig zu optimieren bedeutet, dass viele Sensordaten im Roboter mit vielen Logistikdaten des Roboters zum Beispiel in einem zentralen Leitstand zusammengeführt werden müssen. Dadurch entstehen viele Vernetzungen und gleichzeitig auch viele Daten, die gesammelt, ausgewertet, zwischengespeichert und weitergeleitet werden müssen. Hier müssen sowohl Maßnahmen für die End-Point-Sicherheit als auch für die Netzwerksicherheit ergriffen werden, um Cyberangriffe wirksam zu vermeiden. Der Bedarf an Cybersicherheit steigt nochmals, wenn kollaborative Roboter, sogenannte Cobots, zum Einsatz kommen, da dann auch Safety-Aspekte eine Rolle spielen.

ANWENDUNGSFALL 3

Medizinsektor

Lokale Patienten-App

Edge AI kann verwendet werden, um medizinische Daten von tragbaren oder körpernahen Geräten und Sensoren zu analysieren, um Patientinnen und Patienten in Echtzeit zu überwachen (vgl. Singh & Gill, 2023, S. 82), Abweichungen zu erkennen und frühzeitig auf potenzielle Gesundheitsprobleme hinzuweisen. Der Service würde auch dann weiterhin funktionieren, wenn die Konnektivität in einer Region schwankt. Anfallende Daten können lokal analysiert werden, um personalisierte Empfehlungen und Einblicke für Patientinnen und Patienten bereitzustellen und ihnen dabei zu helfen, ihre Gesundheit effektiver zu verbessern. Durch die lokale Datenverarbeitung kann die Patientendatensicherheit und -privatsphäre erhöht werden. Dies kann zur Einhaltung von Gesundheitsdatenschutzbestimmungen wie der DSGVO beitragen. Solche Anwendungen können nicht nur im Medizinsektor von Vorteil sein, sondern auch in der Pflege, wenn es um das Remote Monitoring von Vitalfunktionen in Echtzeit oder andere Informationen über die Situation von Pflegeempfangenden geht.
Weiteres Beispiel: Identifikation von Krankheitsfällen (Leukämie, Tuberkulose, Covid-19) einzelne Kliniken als „Edge Nodes" (siehe KI-Kompakt „Verteiltes maschinelles Lernen", Plattform Lernende Systeme, 2022).

Herausforderungen für die sektorale Anpassung von Edge-AI-Lösungen und Sicherheitsaspekte

Im Medizinsektor gibt es teils kleine Firmen, die jedoch von der Entwicklung bis hin zur Prüfung von Produkten alles abdecken müssen. Dies ist schwer zu leisten, vor allem dann, wenn es sich um die Einführung neuer Technologien handelt. Zudem herrschen im Gesundheitssektor strenge gesetzliche Vorgaben und umfassende Zulassungsprozesse, die bei Produkteinführungen beachtet werden müssen. Schließlich fehlt es an vielen Stellen auch an der notwendigen Digitalisierung und damit an der entsprechenden Datengrundlage (siehe: elektronische Patientenakte, kurz ePA).

Da bestimmte technische Spezifika von medizinischen Edge-AI-Geräten bzw. -Systemen einen möglichen Wettbewerbsvorteil gegenüber der Konkurrenz darstellen, müssen sie entsprechend geschützt werden, um diese Vorteile zu erhalten. Häufig kommen mehrere KI-Algorithmen zum Einsatz, etwa um…

a) Nutzendendaten zu filtern, die zum Beispiel für die Diagnose benötigt werden,
b) Abweichungen vom „Normal-Profil" auszuwerten, zum Beispiel um mögliche Behandlungsmaßnahmen vorzuschlagen und
c) andere KI-Systeme abzuschalten, wenn diese nicht benötigt werden, um den Stromverbrauch zu reduzieren.

Diese drei Sektoren haben sehr unterschiedliche Voraussetzungen. Dies bedeutet, dass bei der Umsetzung von Edge AI in jedem Bereich entsprechende Anpassungen erforderlich sind. Es werden daher technologische Lösungen benötigt, die entsprechend den unterschiedlichen Voraussetzungen in den Anwendungsbereichen flexibel gestaltet und eingesetzt werden können.

4 Gestaltungsoptionen

Edge AI bietet vielfältige Einsatzmöglichkeiten bei der Bewältigung aktueller Herausforderungen und vieler Transformationsvorhaben – von der Energiewende bis hin zur technologischen Souveränität (Abschnitt 2.3). Zudem kann diese Technologie einen Beitrag für die Entwicklung einer privatsphärenfreundlichen, energieeffizienten KI im Echtzeitbetrieb leisten (Abschnitt 2.2). Unterschiedliche Gestaltungsoptionen können in Betracht gezogen werden, um Edge AI in Deutschland und Europa voranzutreiben und um das beschriebene Potenzial von Edge AI zu heben.

Politikerinnen und Politiker sowie staatliche Einrichtungen

- Das größte Potenzial von Edge AI liegt in der synergetischen und holistischen Betrachtung von Hardware, Software, Künstlicher Intelligenz und Anwendungsszenarien. Edge AI voranzutreiben ist zudem auch im KI-Aktionsplan des Bundesministeriums für Bildung und Forschung (BMBF) festgelegt (BMBF, 2023, S. 9). Eine Aufgabe sollte darin bestehen, die per se sehr unterschiedlichen Disziplinen zusammenzubringen, sei es durch entsprechende Förderprojekte, durch Unterstützung bei der Schaffung von multidisziplinären Studiengängen und Curricula oder durch die Schaffung von Beratungsstellen, die Firmen dabei unterstützen, das fehlende Puzzlestück zur Gesamtlösung zu finden.

- Außerdem sollte die Förderung von KI-Forschungsprojekten einen niedrigen Energieverbrauch als Nebenziel einbeziehen sowie eine Projektförderung hinsichtlich Speicher- und Rechensparsamkeit von KI-Netzwerken und KI-Modellen mit ins Auge fassen.

Unternehmerinnen und Unternehmer

- Eine Strategie für Unternehmen kann es sein, sich als Anbieter von Lösungen von komplexen Problemstellungen zu positionieren, denen mit Edge AI begegnet werden kann.

- Der Aufwand für sektorale Anpassung sollte in Kooperation mit Forschenden minimiert werden, indem eine Plattform beziehungsweise Module für Edge AI entwickelt werden, die als Basisbausteine beispielsweise für Systeme der Bilderkennung dienen und dann, je nach den Bedürfnissen, an die Branchen angepasst werden können. Es kann beispielsweise eine Basis für Edge AI geschaffen werde, ähnlich dem „Robot Operating System" (kurz ROS) in der Robotik. Dafür ist eine Standardisierung einschließlich einheitlicher Schnittstellen zwischen den Branchen erforderlich. Eine einheitliche Hardware- und Entwicklungsumgebung könnte ebenfalls zweckdienlich sein, um solche Anpassungsprozesse zu unterstützen.

- Ein weiterer wichtiger Punkt ist die Lösung der Abhängigkeit von Unternehmen mit Vormachtstellung auf dem KI-Markt, die sich auch auf Edge AI ausweiten könnte. Dazu können in Unternehmen Entkopplungsstrategien verfolgt oder ein sogenannter „zweiter Entwicklungspfad" eingeschlagen werden, um einen Vendor Lock-in (Effekt, der verhindern soll, dass Kunden den Anbieter wechseln) zu vermeiden.

Forschende

- Eine Grundlage für Edge AI sind effiziente und leistungsfähige Computerchips, die auch in kleinen Geräten anspruchsvolle Berechnungen für KI-Training und KI-Inferenz durchführen können. Neuromorphe Chips und spezielle Edge-AI-Chips werden in der Forschung bereits als mögliche Lösungen entwickelt. Diese Forschungsrichtungen sollten weiterverfolgt werden, um neue Grundlagen für ressourcenbewusste Datenverarbeitung im Rahmen der Grenzen von Edge-Geräten zu schaffen (Whitten, 2022). Im KI-Aktionsplan des BMBF werden hierzu Beiträge ausgewiesen (BMBF, 2023, S. 9). Dies ist ein Schritt in die richtige Richtung.

- Neue und spezifische Verarbeitungsmöglichkeiten bei Hardware und Compilern sind erforderlich, die die Größe künstlicher neuronaler Netze oder Entscheidungsbaum-Ensembles (Random Forests) reduzieren und optimierten Code erzeugen. Kompressions-, Reduktions- und Destillationsverfahren für große KI-Modelle führen dazu, dass sie „on the Edge" ausgeführt werden können (Freund, 2023). Ebenfalls weiterentwickelt werden sollten grundlegend anders konzipierte neuronale KI-Modelle, die mit viel/wesentlich kleineren Parameterzahlen Vergleichbares leisten oder sogar ganz neue Anwendungen ermöglichen (z. B. Modelle, die sich an neuronalen Netzen von Insekten orientieren).

- Für die Vielfalt an Lösungen für Kombinationen von Lernverfahren und Rechnerarchitekturen mit ihren jeweiligen Ressourcenbeschränkungen sind Benchmarks und Testverfahren nötig, die die Modellwahl erleichtern können.

- Infrastrukturen und Methoden sollten entwickelt werden, um Daten adäquat für Edge AI zu sammeln und ein reibungsloses Training der Modelle maschinellen Lernens zu ermöglichen.

- Die neuen Herausforderungen in Bezug auf Sicherheit und Privatsphäre, die durch die Diversifizierung möglicher Angriffsvektoren bei Edge AI entstehen, sollten angegangen werden, um die Vorteile dezentraler KI-Inferenz und dezentralen KI-Trainings optimal auszuschöpfen.

- Offene Fragen bestehen ebenfalls bei der Erklärbarkeit von Edge AI, wenn mit Black-Box-Modellen gearbeitet wird, wie bei solchen KI-Modellen allgemein. Hierbei geht es nicht nur um Erklärbarkeit als Werkzeug der Transparenz, sondern auch als Werkzeug der Fehlerbehebung. Daneben spielt auch der Aspekt der Beobachtbarkeit („observability") eine Rolle, vor allem bei der kritischen Qualitätssicherung, zum Beispiel bezüglich der Erkennung von concept drifts („concept drift detection").

- Weitere offene Fragen betreffen die Datenportabilität und die Handhabung von digitalen Identitäten „on the edge".

Lehrende

- Hochschulen und auch die Forschung sollten systemisches Denken in der Lehre fördern. Die Zusammenarbeit von Hardware- und Software-Expertinnen und -Experten erlaubt Studienprojekte, die den Studierenden die Zusammenhänge der Inhalte aus verschiedenen Studienbereichen (z. B. Datenbanken und Rechnerarchitekturen) zeigen.

- Hierfür ist es notwendig, in Studiengängen und Curricula die entsprechende Multidisziplinarität zu fördern. Dies gilt auch über die MINT-Grenzen hinweg, beispielsweise mit den Sozialwissenschaften, um über die Akzeptanz oder gesellschaftlichen Folgen von immer stärker verbreiteten (Edge-)AI-Systemen zu diskutieren und Lösungen zu erarbeiten. Mit Open-Access-Büchern wie „Machine Learning under Resource Constraints" (Morik & Marwedel, 2023; Morik, K., Rahnenführer, J. & Wietfeld, C. 2023) des Sonderforschungsbereich (SFB) 876 liegen erstmals Texte für Forschung und Lehre vor, an denen sich Lehrende orientieren können. Der erste Band „Fundamentals" behandelt die Algorithmen und Architekturen von maschinellem Lernen und eingebetteten Systemen in einem einheitlichen Rahmen. Er bietet die Grundlagen für Edge AI. Der dritte Band „Applications" behandelt Anwendungen aus Medizin, Industrie 4.0, Smart Cities und Kommunikationsnetzwerken.

Verbände und Vereine

- Standardisierung für Edge-AI-Problemstellungen in Kooperation zwischen Standardisierungsorganisationen wie dem Deutschen Institut für Normung (DIN), Unternehmerinnen und Unternehmern und Forschenden angehen, insbesondere für Schnittstellen.

- Entwicklung von allgemein anerkannten Benchmarks in Zusammenarbeit mit Unternehmen und Forschenden.

Öffentlichkeit

- Es sollten nicht nur die großen, spektakulären KI-Modelle, die zentral trainiert und ausgeführt werden, in der Öffentlichkeit diskutiert werden. Auch alternative Möglichkeiten, KI umzusetzen, wie eben Edge AI, sollten Teil der öffentlichen Debatte sein, damit eine informierte und differenzierte öffentliche Debatte über Chancen und Risiken von KI stattfinden kann.

- Durch den stetigen und transparenten Austausch mit der Bevölkerung sollte das Bewusstsein für den Einsatz dieser Systeme beispielsweise in Endkundengeräten und deren Vorteile in bestimmten Gebieten gegenüber konventionellen Methoden ausgeweitet werden.

Literatur

acatech (2023): Circular Economy Initiative Deutschland. Deutsche Akademie der Technikwissenschaften (acatech). Online unter: https://www.acatech.de/projekt/circular-economy-initiative-deutschland/

Boll, S. & Schnell, M. et al. (2022): Mit Künstlicher Intelligenz zu nachhaltigen Geschäftsmodellen – Nachhaltigkeit von, durch und mit KI. Whitepaper aus der Plattform Lernende Systeme, München. https://doi.org/10.48669/pls_2022-1

Bundesministerium für Bildung und Forschung (BMBF) (2023): BMBF-Aktionsplan „Künstliche Intelligenz". Online unter: https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/kuenstliche-intelligenz/ki-aktionsplan.html

Buschjäger, S. et al. (2021): Margin-Maximization in Binarized Neural Networks for Optimizing Bit Error Tolerance. DATE 2021, S. 674–678. https://doi.org/10.23919/DATE51398.2021.9473918

Ciupek, M. (2023): Föderales Lernen trainiert Roboter ohne Übergabe sensibler Daten. Online unter: https://www.vdi-nachrichten.com/technik/automation/foederales-lernen-trainiert-roboter-ohne-uebergabe-sensibler-daten/

Crunchtime (2023): Geschäftsberichtsstudie Crunchtime Risikomonitor 2023. Online unter: https://www.crunchtime-communications.com/crunchtime-risikomonitor-2023

Dede, G. et al. (2021): Cybersecurity Challenges in the Uptake of Artificial Intelligence in Autonomous Driving. Online unter: https://www.enisa.europa.eu/publications/enisa-jrc-cybersecurity-challenges-in-the-uptake-of-artificial-intelligence-in-autonomous-driving/@@download/fullReport

Destatis (2023): Wichtigstes deutsches Exportgut 2022: Kraftfahrzeuge. Online unter: https://www.destatis.de/DE/Themen/Wirtschaft/Aussenhandel/handelswaren-jahr.html

Deuse, J. et al. (2022): Quality Assurance in Interlinked Manufacturing Processes. In: Morik, K., Rahnenführer, J. & Wietfeld, C. (Hrsg.), Volume 3 Machine Learning under Resource Constraints – Applications, S. 114–135. De Gruyter. https://doi.org/10.1515/9783110785982-015

Deutsches Zentrum für Luft- und Raumfahrt (DLR) (2022): Drohnen sammeln Daten für schnelle Katastrophenhilfe. Deutsches Zentrum für Luft- und Raumfahrt (DLR). Online unter: https://www.dlr.de/de/aktuelles/nachrichten/2022/04/drohnen-sammeln-daten-fuer-schnelle-katastrophenhilfe

Ding, A. et al. (2021): Roadmap for Edge AI: A Dagstuhl Perspective. ACM SIGCOMM Communication Review, Vol. 52, S. 28–33. et al. https://doi.org/10.1145/3523230.3523235

Freund, K. (2023): How To Run Large AI Models On An Edge Device. Online unter: https://www.forbes.com/sites/karlfreund/2023/07/10/how-to-run-large-ai-models-on-an-edge-device/

Gartner (2023a): Gartner Identifies Top Trends Shaping the Future of Data Science and Machine Learning. Online unter: https://www.gartner.com/en/newsroom/press-releases/2023-08-01-gartner-identifies-top-trends-shaping-future-of-data-science-and-machine-learning

Gartner (2023b): What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle. Online unter: https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

Heikkilä, M. (2023): Making an image with generative AI uses as much energy as charging your phone. Online unter: https://www.technologyreview.com/2023/12/01/1084189/making-an-image-with-generative-ai-uses-as-much-energy-as-charging-your-phone/

Henderson, P. et al. (2020): Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. Journal of Machine Learning Research (JMLR), Vol. 21(248), S. 1-43. Online unter: http://jmlr.org/papers/v21/20-312.html

Heppe, L. et al. (2020): Resource-Constrained On-Device Learning by Dynamic Averaging. PKDD/ECML Workshops 2020. https://doi.org/10.48550/arXiv.2009.12098

IDC (2022): IDC-Studie: 60 Prozent der deutschen Unternehmen investieren in zukunftsfähige Data Center. Online unter: https://www.idc.com/getdoc.jsp?containerId=prEUR149762022 (Letzter Zugriff: 14.12.2023)

Kinane, D. et al. (2014): Intelligent Synthesis and Real-time Response using Massive Streaming of Heterogeneous Data (INSIGHT) and its anticipated effect on Intelligent Transport Systems (ITS) in Dublin City, Ireland. 10th ITS European Congress, Helsinki. Online unter: https://api.semanticscholar.org/CorpusID:44086093

Krieger, C., Sliwa, B. & Wietfeld, C. (2022): Vehicle to Vehicle Communications: Machine Learning Enabled Predictive Routing. In: Morik, K, Rahnenführer, J. & Wietfeld, C. (Hrsg.), Volume 3 Machine Learning under Resource Constraints – Applications, S. 272–284. De Gruyter. https://doi.org/10.1515/9783110785982-023

Luk, W. et al. (2022): Hardware-Aware Execution. In: Morik, K. & Marwedel, P. (Hrsg.), Volume 1 Machine Learning under Resource Constraints – Fundamentals, S. 249–304. De Gruyter. https://doi.org/10.1515/9783110785944-006

McEnroe, P., Wang, S. & Liyanage, M. (2022): A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges. IEEE Internet of Things Journal, Vol. 9(17), S. 15435–15459. https://doi.org/10.1109/JIOT.2022.3176400

McQuate, S. (2023): Q&A: UW researcher discusses just how much energy ChatGPT uses. Online unter: https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/

Morik, K., Marwedel, P. (2023): Machine Learning under Resource Constraints: Fundamentals, DeGruyter. Online unter: https://www.degruyter.com/document/doi/10.1515/9783110785944/html

Morik, K. & Rhode, W. (2023): Volume 2 Machine Learning under Resource Constraints - Discovery in Physics. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110785968

Morik, K., Rahnenführer, J. & Wietfeld, C. (2023): Volume 3 Machine Learning under Resource Constraints - Applications. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110785982

Piatkowski, N., Lee, S. & Morik, K. (2016): Integer undirected graphical models for resource-constrained systems. Neurocomputing, Vol, 173(1), S. 9–23. https://doi.org/10.1016/j.neucom.2015.01.091

Plattform Lernende Systeme (2022): KI Kompakt: Verteiltes maschinelles Lernen. Besserer Datenschutz für KI-Anwendungen? (Publikationsreihe). Online unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/KI_Kompakt/PLS_KI_Kompakt_ML.pdf

Reuters (2023): EU examines Nvidia-dominated AI chip market's alleged abuses, Bloomberg reports. Online unter: https://www.reuters.com/technology/eu-starts-early-stage-probe-into-nvidia-dominated-ai-chip-market-abuses-2023-09-29/

Rüegg, P. (2019): Datenschatz vom Matterhorn. Online unter: https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2019/08/10-jahre-matterhorn-permafrost.html

Singh, R. & Gill, S. (2023): Edge AI: A survey. Internet of Things and Cyber-Physical Systems, Vol. 3, S. 71–92. https://doi.org/10.1016/j.iotcps.2023.02.004

Sliwa, B. et al. (2020): PARRoT: Predictive Ad-hoc Routing Fueled by Reinforcement Learning and Trajectory Knowledge. 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), S.1–7. https://doi.org/10.48550/arXiv.2012.05490

Statista (2023): Welches sind Ihrer Meinung nach die wichtigsten Probleme, denen Deutschland derzeit gegenübersteht? Online unter: https://de.statista.com/statistik/daten/studie/2739/umfrage/ansicht-zu-den-wichtigsten-problemen-deutschlands/

Statistisches Bundesamt (2023): Anzahl der Elektrotechnik- und Elektronikstudierende in Deutschland in den Wintersemestern 2010/11 bis 2022/23. Statista. Online unter: https://de.statista.com/statistik/daten/studie/261137/umfrage/studierendenanzahl-im-bereich-elektrotechnik-elektronik-in-deutschland/

Steward, M. et al. (2023): Machine Learning Sensors – A Design Paradigm for the Future of Intelligent Systems. Communications of the ACM, vol. 66(11), S. 25–29. https://doi.org/10.1145/3586991

Whitten, A. (2022): New Chip Expands the Possibilities for AI. Online unter: https://www.quantamagazine.org/a-brain-inspired-chip-can-run-ai-with-far-less-energy-20221110/

Über dieses Whitepaper

Das Whitepaper entstand arbeitsgruppenübergreifend unter Federführung der zwei Arbeitsgruppen Technologische Wegbereiter und Data Science sowie IT-Sicherheit, Privacy, Recht und Ethik.

Die Grundlagen zum Whitepaper wurden im Rahmen eines „Runden Tisches" mit 13 Teilnehmenden aus Wissenschaft und Wirtschaft im März 2023 in der Geschäftsstelle der acatech in München erarbeitet. Wir danken allen Workshopteilnehmenden an dieser Stelle ganz herzlich für die aktive Teilnahme und die inhaltlichen Beiträge.

Autorinnen und Autoren
Prof. Dr. Wolfgang Ecker, Technische Universität München / Infineon Technologies AG
Prof. Dr. Björn Eskofier, Friedrich-Alexander-Universität Erlangen-Nürnberg
Dr. Detlef Houdeau, Infineon Technologies AG
Prof. Dr. Katharina Morik, Technische Universität Dortmund / Lamarr-Institut für Maschinelles Lernen und Künstliche Intelligenz
Dr. Remo Lachmann, IAV GmbH

Autoren mit Gaststatus
Dr. Sebastian Buschjäger, Technische Universität Dortmund
Prof. Dr. Jan S. Rellermeyer, Gottfried Wilhelm Leibniz Universität Hannover

Redaktion
Dr. Maximilian Hösl, Geschäftsstelle der Plattform Lernende Systeme
Christine Wirth, Geschäftsstelle der Plattform Lernende Systeme

Impressum

Herausgeber
Lernende Systeme –
Die Plattform für Künstliche Intelligenz
Geschäftsstelle | c/o acatech
Karolinenplatz 4 | 80333 München
www.plattform-lernende-systeme.de

Gestaltung und Produktion
PRpetuum GmbH, München

Stand
Juli 2024

Bildnachweis
gorodenkoff/iStock/Titel

Empfohlene Zitierweise
Ecker, W., Houdeau, D. et al. (2024): Edge AI: KI nahe am Endgerät. Technologie für mehr Datenschutz, Energieeffizienz und Anwendungen in Echtzeit. Whitepaper aus der Plattform Lernende Systeme, München.
DOI: https://doi.org/10.48669/pls_2024-4

Bei Fragen oder Anmerkungen zu dieser Publikation kontaktieren Sie bitte Dr. Thomas Schmidt (Leiter der Geschäftsstelle):
kontakt@plattform-lernende-systeme.de

Über die Plattform Lernende Systeme

Die Plattform Lernende Systeme ist ein Netzwerk von Expertinnen und Experten zum Thema Künstliche Intelligenz (KI). Sie bündelt vorhandenes Fachwissen und fördert als unabhängiger Makler den interdisziplinären Austausch und gesellschaftlichen Dialog. Die knapp 200 Mitglieder aus Wissenschaft, Wirtschaft und Gesellschaft entwickeln in Arbeitsgruppen Positionen zu Chancen und Herausforderungen von KI und benennen Handlungsoptionen für ihre verantwortliche Gestaltung. Damit unterstützen sie den Weg Deutschlands zu einem führenden Anbieter von vertrauenswürdiger KI sowie den Einsatz der Schlüsseltechnologie in Wirtschaft und Gesellschaft. Die Plattform Lernende Systeme wurde 2017 vom Bundesministerium für Bildung und Forschung (BMBF) auf Anregung des Hightech-Forums und acatech – Deutsche Akademie der Technikwissenschaften gegründet und wird von einem Lenkungskreis gesteuert.