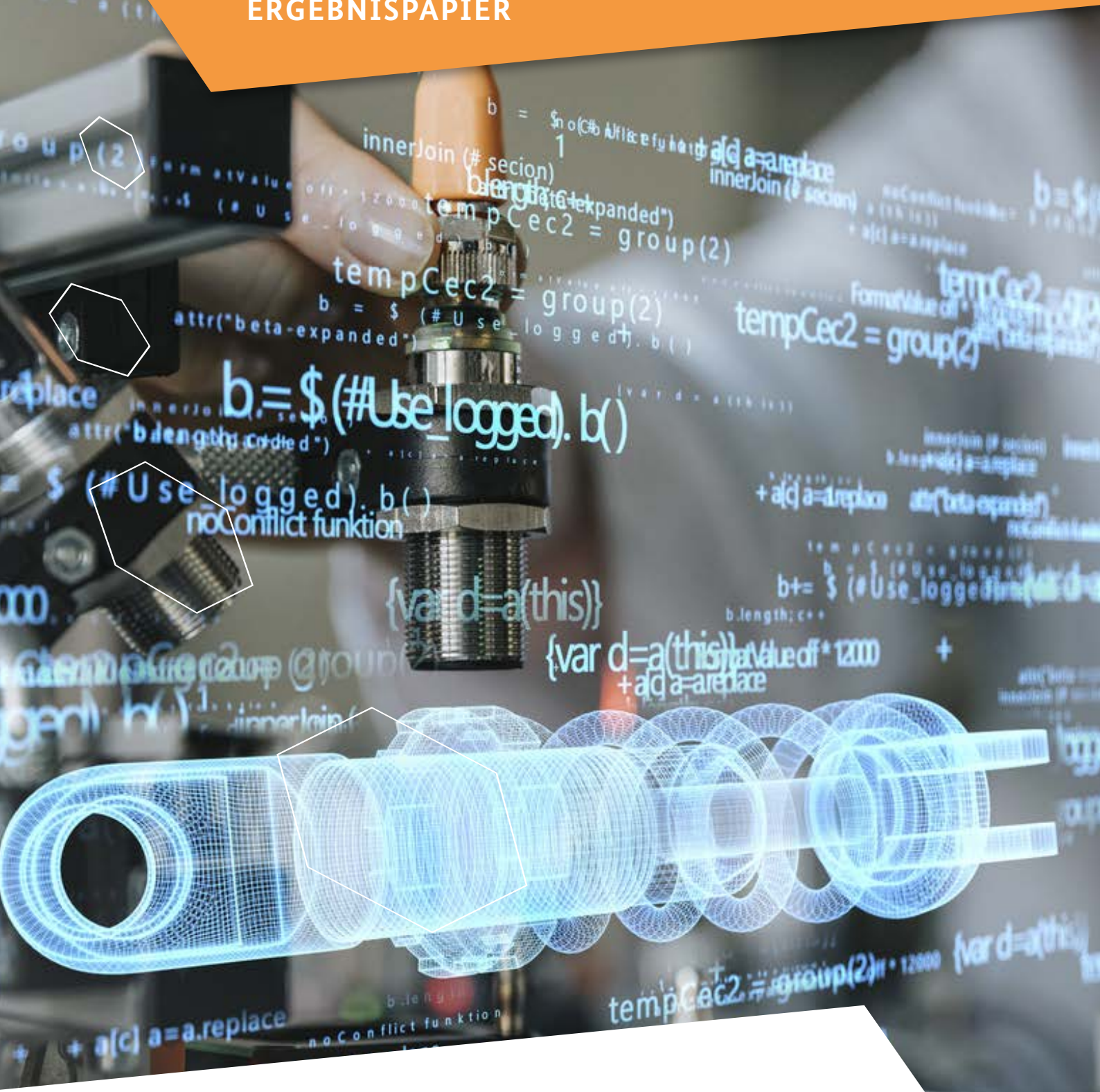
PLATTFORM INDUSTRIE 4.0

ERGEBNISPAPIER



Industrial Security und die Entwicklung von KI-Anwendungen in der Edge

Impressum

Herausgeber
Bundesministerium für Wirtschaft und Energie (BMWi)
Öffentlichkeitsarbeit
11019 Berlin
www.bmwi.de

Redaktionelle Verantwortung
Plattform Industrie 4.0
Bülowstraße 78
10783 Berlin

Stand
Juli 2021

Diese Broschüre wird ausschließlich als Download angeboten.

Gestaltung
PRpetuum GmbH, 80801 München

Bildnachweis
Infinite Lux/Westend61 / Adobe Stock / Titel, S. 10
xiaoliangge / Adobe Stock, ihor lishchyshyn / iStock / S. 5
D3Damon / iStock / S. 13
da-kuk / iStock / S. 16
Thomas Söllner / Adobe Stock / S. 23

Zentraler Bestellservice für Publikationen der Bundesregierung:
E-Mail: publikationen@bundesregierung.de
Telefon: 030 182722721
Bestellfax: 030 18102722721

Diese Publikation wird vom Bundesministerium für Wirtschaft und Energie im Rahmen der Öffentlichkeitsarbeit herausgegeben. Die Publikation wird kostenlos abgegeben und ist nicht zum Verkauf bestimmt. Sie darf nicht zur Wahlwerbung politischer Parteien oder Gruppen eingesetzt werden.

Inhaltsverzeichnis

1 Einleitung: Status und Trend der industriellen Anwendung von KI

Der Einsatz von KI in Produktions- und Verwaltungsanwendungen der Industrie wird vorwiegend durch das Streben nach Steigerung der Produktivität und Einführung neuer Leistungsmerkmale getrieben. KI-Systeme erweitern Dienstleistungen und Analysetätigkeiten an der Kundenschnittstelle, leisten Beiträge zu sensorgesteuerter Automatisierung der Produktion und bilden die Basis für neue Produktfähigkeiten und deren Umsetzung in neuen Geschäftsmodellen. Neue KI-basierte Verfahren reduzieren den Personaleinsatz in Routinetätigkeiten und substituieren kostenaufwendige Analogtechnik durch preiswertere, vergleichsweise einfache digitale Sensortechnologie. Das vorliegende Papier fokussiert auf Neuronale Netze. Weitere Ansätze, die KI unterstützen können (Clustering, Kombinatorik etc.), werden in diesem Papier nicht behandelt.

Mit dem flächendeckenden Vordringen der KI in ehemals durch Menschen und analoge Technik geprägte Tätigkeitsprofile entstehen neue Sicherheitsrisiken. Eine Ursache dafür liegt in der weiterhin schnell abnehmenden Erklärbarkeit der Ergebnisse fortgeschrittener Neuronaler Netze. Dieses Thema wird am Ende dieser Einleitung kurz beleuchtet.

Seit 2018 hat darüber hinaus eine technische Revolution im Bereich des Machine Learning (ML) begonnen, in der die Einsatzbereiche von Cloud und Edge neu definiert werden: Während in der frühen Phase des Einsatzes von ML sowohl für die Erstellung (Datenakquisition, Datenauswertung, Training) als auch den operativen Einsatz des fertigen Erkennungssystems (Inferenz) erhebliche Rechenressourcen (normalerweise auf einem umfangreichen Serverpool in einem Cloud-RZ) nötig waren, beschränkt sich dies heute im Wesentlichen auf das Training, dessen Effizienz durch skalierende Netzarchitekturen und Aspekte des AutoML™ [1] sowie Transfer- und Multitask-Learning dramatisch erhöht wird. Dagegen wird mit neuen Technologien wie Quantisierung [1], Flatbuffer-Format-Zugriffs-Technologie [2] und Edge-Beschleunigern (Graphics-Processing Unit (GPU), „Tensor-Processing-Unit“ (TPU), „Intelligence-Processing-Unit“ (IPU) u. ä.) in kostengünstigen, dezentral lokal betriebenen Edge-Devices Inferenz-Performance erreicht, die schon wegen heutiger und realistisch absehbarer Netzlatenz zentralen Instanzen verschlossen ist. Inferenz, der praktische Einsatz des fertig trainierten Systems, ist in der Cloud nur noch im Hinblick auf Vorteile bei Aktualisierung, Pflege und extremem applikationsspezifischem Bedarf an Rechen- und Speicherressourcen sinnvoll.

Der produktive Einsatz der KI findet voraussichtlich langfristig hauptsächlich in der Edge statt. So entstehen neben allen positiven Eigenschaften gewissermaßen als Trade-off-Effekt erhebliche Potenziale an Sicherheitsrisiken durch Fremdüberwachung und Spionage. Andererseits wird durch die Verlagerung komplexer Aufgaben in die Edge die Außenkommunikation neu strukturiert. Es entstehen neue Methoden und Möglichkeiten, Privatheit und Geheimhaltung zu stärken. Beispielsweise kann die permanente Kommunikation auf Maschinenebene durch regelmäßige Lieferungen von Berichten durch Edge-Devices in einem sicherheitsoptimierten Format abgelöst werden, in dem nur solche Informationen enthalten sind, die der jeweilige Dienstleister zu Erfüllung seiner Aufgaben tatsächlich benötigt. Etwaigem Spionageverdacht kann so entgegengewirkt werden.

Nach dieser Einleitung werden in **Kapitel 1** des vorliegenden Papiers zusammenfassend die heute typischen Anwendungen des ML in der Industrie sowie deren Vordringen und Verdrängung früherer Alternativen beschrieben. Dabei wird auch die aktuelle Risikolage skizziert, die in solchen Anwendungen durch missbräuchlichen Einsatz von KI für Angriffszwecke entsteht. Auf Grundlage der zuvor beschriebenen Entwicklungen werden in **Kapitel 2** die strukturellen Veränderungen des industriellen KI-Einsatzes rekapituliert. **Kapitel 3** gibt einen Überblick über ausgewählte Klassen industrieller Prozesse, in denen KI heute konkrete Unterstützung bieten kann, und stellt Auswahlkriterien für KI-Methoden vor. **Kapitel 4** leitet über auf besondere Sicherheitsaspekte von Industrie 4.0-Anwendungsfällen. **Kapitel 5** liefert eine Zusammenstellung der aktuellen Möglichkeiten, die für Industrie 4.0 wichtigen Sicherheitsthemen durch KI zu unterstützen. Dabei werden die Potenziale der KI-Unterstützung in den grundlegenden als auch in den aktuell neu entstandenen Bereichen erörtert. In dieser Diskussion wird auch auf voraussichtlich zu erwartende Chancen und Risiken durch die Verlagerung der produktiven KI-Anwendung an die Edge innerhalb der Unternehmensgrenzen hingewiesen. Grundlegende Handlungsempfehlungen liefert das **Kapitel 6** zusammen mit einer abschließenden **Fazit-Betrachtung**.

1 Kostenpflichtige Software zur Optimierung des Trainings von KI.

Mit dem vorliegenden Dokument sollen einerseits Domänenexperten in die Lage versetzt werden, die Eignung von KI-Technik für ihren jeweiligen Anwendungsbereich besser einzuschätzen. Andererseits richtet sich das vorliegende Dokument an industrielle Anwender von KI-basierten Systemen und Entwickler von KI-Algorithmen, die über Grundlagenwissen in den Methoden der Künstlichen Intelligenz verfügen sollten.

Weitere Arbeiten zur industriellen Anwendung dieser technologischen Entwicklung sind unter anderem im Zentralverband der Elektroindustrie (ZVEI) entstanden [3]. Im Laufe seines Entstehens ist das Dokument kontinuierlich sowohl in technischer Hinsicht als auch bezüglich wichtiger politischer Aussagen und Maßnahmen aktualisiert worden. Das Abschlussdatum ist der 1.7.2021.

Nicht-Erklärbarkeit von KI-Entscheidungen

Eine Schwierigkeit, die bei KI-behafteten Entscheidungen und Bewertungen auftritt, ist die Nicht-Erklärbarkeit [4] des Ergebnisses der KI-Anwendung. Es wird häufig nach Erklärungen gesucht, wenn die KI ein Ergebnis liefert, welches der menschlichen Erwartungshaltung nicht genügt. KI ist nicht in der Lage, Ursachen für Fehlentscheidungen zu erkennen oder gar die aufgetretenen Probleme zu generalisieren oder zu abstrahieren, wie der Mensch es kann.

Die Unterarbeitsgruppe „KI für Industrie-4.0-Security“ hat zu diesem Thema bereits zwei Berichte veröffentlicht [5] [6].

Es besteht derzeit ein klarer Widerspruch zwischen Präzision und Erklärbarkeit von KI-Ergebnissen, was auch in deren übermenschlichem Leistungsniveau in bestimmten Anwendungen begründet ist. Zwar existieren inzwischen diverse Bestrebungen, die Erklärbarkeit von KI-Entscheidungen zu steigern, befriedigende Ergebnisse sind aber noch kaum entstanden.

Der Anspruch des Menschen, eine für ihn geeignete Erklärung finden zu wollen, mag im Rahmen technischer Betrachtungen, wie sie z. B. in diesem Dokument beschrieben sind, noch auf die eine oder andere Art befriedigt werden können (siehe nächste Absätze). Mittlerweile kann eine KI jedoch auch Dinge erkennen, für die der Mensch keine Erklärungsmuster mehr hat, so dass dieser Anspruch zunehmend schwieriger zu erfüllen sein wird. In Referenz [7] wird beispielsweise über ein KI-basiertes Netzhaut-Bildauswertungsverfahren berichtet, welches es unter anderem erlaubt, Geschlecht, Krankheiten und Lebensgewohnheiten eines Menschen zu erkennen (sog. Opportunistisches Lernen). Die Erkennung einiger dieser Attribute ist (zumindest bislang) für den Menschen nicht erklärbar, es war zuvor wissenschaftlich nicht bekannt, dass bestimmte Daten überhaupt in sichtbaren Merkmalen der Netzhaut enthalten sind.

Bessere Erklärbarkeit von Ergebnissen erhält man bei Verwendung älterer KI-Methoden wie etwa mit Support-Vector-Machines (SVM) – eine Kernel-basierte Familie von Machine-Learning-Methoden. Allerdings erzielt man hiermit üblicherweise eine signifikant niedrigere Erkennungsrate im Vergleich zu Neuronalen Netzen.

Einige Bestrebungen, bessere Erklärbarkeit von auf Neuronalen Netzen basierender KI zu erzielen, gehen dahin, die Verschachtelungstiefe der automatisch erzeugten Features in modernen Neuronalen Netzen mit typisch mehr als 100 Schichten durch zusätzliche KI-Methoden mit geringerer Tiefe zu übersetzen.

Folgende „Best Practices“ können die Erklärbarkeit von Ergebnissen erhöhen:

- Intensives Testen (auch unter Zuhilfenahme einer Test-KI), mit entsprechender Fokussierung auf Entscheidungsgrenzfälle. Dies gilt auch im Allgemeinen, unabhängig von der Benutzung einer KI zur Entscheidungsfindung.

- Anlernen der KI mit den relevanten Daten, so dass kein Falschlernen erfolgt.

- Wenn möglich, Granularisierung der KI-basierten Bewertungen, so dass nachvollziehbar ist, welcher Teil eines KI-Regelwerkes welche Bewertung vorgenommen hat.

2 Strukturelle Veränderungen des industriellen KI-Einsatzes

Wie bereits einleitend skizziert, entwickelt sich KI in vielen speziellen Anwendungsbereichen in zwei Zielrichtungen mit unvermindert großer Geschwindigkeit weiter: immer höhere Leistung, bis in übermenschliche Fähigkeiten, und gleichzeitig immer mehr Verbreitung in der unmittelbar erlebten alltäglichen Realität. Beide Tendenzen beziehen sich auf den Aspekt des Machine Learning, der inzwischen oft als der eigentlich wirtschaftlich relevante Aspekt von KI gesehen wird. In Autos und in der Mehrzahl der technisch anspruchsvollen Geräte in Haushalt, Unterhaltung, Produktion, Diagnose, Militär, Nachrichtentechnik, Landwirtschaft – KI ist überall und Grenzen der Verbreitung sind kaum erkennbar, da immer neue Nutzenpotenziale erschlossen werden. Gleichzeitig werden immer mehr Bereiche von KI in vielfacher Hinsicht weitaus „intelligenter“ bearbeitet, als Menschen es können. Maschinen sind nicht nur wirtschaftlicher, viel schneller und zuverlässiger, sondern in ihrem speziellen Bereich auch viel intellektuell leistungsfähiger. So wie die Universalgelehrten in den letzten 500 Jahren von hochspezialisierten Extremtalenten abgelöst wurden, so setzt sich der Prozess nun mit ML in eine neue Epoche weiterer Spezialisierung fort. ALPHA-GO kann nur GO spielen, sonst nichts. Er hat keine „Bildung“ und keine menschlichen Qualitäten. Aber in dem komplexesten Strategiespiel, das die Menschheit hervorgebracht hat, brauchen weder einzelne Menschen noch Gruppen und auch keine von Menschen programmierten Maschinen mehr gegen ihn antreten.

Dieses Kapitel versucht, die aktuelle Entwicklung in diesen zwei Dimensionen aus Sicht der Industrie 4.0-Security zu erläutern.

2.1 KI in der Edge

Wie bereits einleitend erwähnt wurde, hat sich in vielen Fällen Leistungsfähigkeit in der Edge in den letzten drei Jahren signifikant vergrößert. Dadurch kann KI in weiten Bereichen des täglichen Lebens seine Wirkung entfalten. Dies hat schon der bekannte Stanford-Professor Andrew Ng propagiert, als er prophezeite „AI is the new electricity“. KI wird so selbstverständlich wie Elektrizität, sie ist überall, der Anteil technischer Geräte, die keinen KI-Bezug haben, schrumpft auf eine vernachlässigbare Größe.

Um die neue Situation zu verstehen, sollte man sich darüber bewusst sein, dass ein KI-System in einem ersten Schritt trainiert werden muss. Dabei werden bei einem Neuronalen Netz derzeit in der Regel zwischen 50 und 150 Millionen Parameter (Gewichte) bestimmt. Dieser Schritt ist aufwendig, benötigt viel Rechenkapazität, Hauptspeicher und Gleitkommaarithmetik. Sind diese Parameter bestimmt und das Netz somit „trainiert“, dann folgt die Anwendung, also die Inferenz. Das Netz wird dann normalerweise über einen langen Zeitraum nicht mehr verändert. Es macht dann nur noch Vorwärtsschritte beispielsweise zur Klassifizierung von Eingabewerten. Dies kann mit einer viel kompakteren Speicherstruktur und viel gröberer Arithmetik erfolgen. Das Netz wird dazu in das Flatbuffer-Format umgewandelt und mittels Quantisierung [1] wird die Arithmetik von Float32 auf Integer8 umgestellt. Das Ergebnis ist selbst bei aufwendigen Neuronalen Netzen ein Lademodul, das weniger als ein Gigabyte Speicher braucht und bei entsprechender Beschleunigung der Integer8-Arithmetik eine komplette Inferenzoperation in 5–15 Millisekun-

den erledigt. Das Gerät der Wahl für die Inferenz ist deshalb keineswegs die Serverfarm in der Cloud, sondern in der Regel ein sehr kleines Edge-Device oder ein TPU-beschleunigtes Smartphone (beispielsweise ein Google Pixel 3) – vor Ort und deshalb ohne Netzlatenz.

Durch Fortschritte in der Halbleitertechnik wächst auch die Leistung von Edge-Devices. Geräte im Format eines Raspberry Pi beispielsweise haben heute Speicher bis zu acht Gigabyte und eine Multicore-CPU. Als HW-Basis für die Inferenz haben Geräte mit solchen Ressourcen erhebliche Leistungsreserven selbst für anspruchsvolle Machine Learning (ML)-Systeme in den typischen Industrieanwendungen, wie beispielsweise Qualitätskontrolle auf der Basis von Bildanalyse.

Das Ziel, hochwertige KI in alle Bereiche der Wirtschaft und des Privatlebens zu bringen, rückt somit immer näher. Die führende Plattform für Machine Learning auf sehr kleinen Geräten ist heute TensorFlow Lite. Inzwischen existiert eine Kooperation zwischen TensorFlow und mehr als zehn Systemen aus dem Bereich der Microcontroller. Dies sind Systeme, die zu klein sind, um aufwendige Betriebssysteme zu tragen, sondern typischerweise eine einzige Anwendung als embedded System laden und ausführen können. Seit März 2019 existiert die tinyML Foundation [8], eine professionelle Non-Profit-Organisation. Sie verfolgt das Ziel, Plattformen für Machine Learning zu erschließen, die in dem Sinne klein sind, dass ihr Stromverbrauch im einstelligen Milliwatt-Bereich liegt. Damit öffnet sich die Perspektive für Machine Learning-basierte Anwendungen sogar in dem Bereich von Systemen, die unabhängig vom Stromnetz sind, weil sie über Jahre mit der Leistung einer einzigen kleinen Batterie auskommen oder mit kleinen Solarpanels und kleinem Akku betrieben werden können. Solche Systeme sind also, wie oben dargelegt, zum einen sehr kompakt und zum anderen auch über verhältnismäßig lange Zeiträume weitestgehend autark von externen Stromquellen. Die Summe dieser Eigenschaften führt dazu, dass die Devices in Zukunft also weitgehend aus dem Verborgenen heraus Umwelt, Verkehr, Personen und andere Phänomene als Bild, Ton oder anderen, aus Sensorik abgeleiteten Merkmalen erkennen, aufzeichnen und über geeignete Kanäle an interessierte zentrale Interessenten weitergeben. Die intelligente Edge des Internet-of-Things (IoT) muss dabei nicht permanent mit einer Verbindung zum Internet arbeiten. Ein mögliches Anwendungsbeispiel ist der Förster, der wöchentlich prüft, ob die Microcontroller Borkenkäfer oder andere Schädlinge gesehen, gehört oder gerochen haben. Den möglichen Anwendungsszenarien sind kaum Grenzen gesetzt.

Selbst auf kleinen Geräten bleibt heute Raum für mehrere ML-Systeme, beispielsweise als Ensemble zur weiteren Steigerung der Genauigkeit; aber auch bösartige ML-Systeme, die zur Spionage oder zur Herbeiführung von Störungen dienen können, die also gezielt Sicherheitsprobleme erzeugen, finden auf kleinsten Geräten Platz. Typische Microcontroller haben Speicher im ein- bis dreistelligen kB-Bereich. In den 70er Jahren war das die Kapazität von Großrechnern. Das reicht natürlich nicht, um mit Inception V4 oder EfficientNet B7 [8] die Welt zu erkennen – dazu braucht man „große“ Systeme wie einen Raspberry Pi für 30 Euro (s. o., oder ein ähnliches Gerät) – aber es reicht für die Erkennung einfacher Muster, und dies selbst dann, wenn nebenbei noch weitere Aufgaben erfüllt werden. Dazu sei noch angemerkt, dass die Analyse und Überwachung vermeintlich primitiver Geräte – selbst wenn diese nicht einmal über die technischen Voraussetzungen verfügen, ein Betriebssystem auszuführen – keineswegs trivial sind.

Das Ecosystem der IT-Industrie aus Hardware-, Software-, System- und Anwendungsherstellern hat begonnen, die Möglichkeiten, die sich eröffnen, wenn IT unterschiedlichste Aspekte der Umwelt wahrnehmen und erkennen kann, umfassend zu erschließen. Das gilt in diversen vertikalen Segmenten, im professionellen wie im Consumer-Bereich und über ein weites Spektrum von Komplexität und Wirtschaftlichkeitsaspekten. Techniken wie Sprachsteuerung, musterbasierte Zugangskontrolle oder Dateneingabe für Applikationen aller Arten – von industriellen Produkten wie Spurhalteassistenten im Auto bis zur stimmenbasierten Erkennung von Atemwegserkrankungen oder der industriellen Produktion wie KI in der optischen Qualitätskontrolle bis zu Schweiß-in-Prozess-Qualitätskontrolle – werden immer selbstverständlicher.

In der Anwendungsdomäne Industrie 4.0 [10] erfolgt die Migration von Anwendungsfunktionalität von „außen“ (in der Cloud) in die Operational Technology Edge, die normalerweise innerhalb der Unternehmens-IT steht. Daten müssen das Unternehmen nicht mehr verlassen, um durch einen extern bereitgestellten Dienst verarbeitet zu werden, da der Dienst bei dem Betreiber ohne Anbindung zum Herausgeber durchgeführt wird.

Der wohl wichtigste Nutzen ist neben Kostensenkung und Performancezuwachs die neue Möglichkeit, die Risiken der unternehmensübergreifenden Kommunikation zu reduzieren.

Spionage durch gezielte Auswertung und Interpretation von Nachrichtenströmen und Infiltration mit Schadsoftware kann durch den Übergang auf menschlich erklärbare Protokollebenen deutlich verringert werden.

Die Mustererkennung erfolgt innerhalb des Unternehmens in der Edge in einem Feature der Maschine. Dadurch, dass nur noch periodische Berichte als PDF erstellt werden, aus denen der nächste Wartungstermin hervorgeht, und eine kontinuierliche Datenverbindung nicht mehr notwendig ist, werden die Angriffsmöglichkeiten reduziert.

Da sich sowohl die KI als auch andere Regelwerke zur Prüfung von Auffälligkeiten oder Regelverstößen in diesem Edge-Device befinden können, kann an dieser Stelle schon eine Blockierung und Alarmierung ungültiger und ungewöhnlicher Zugriffe oder Zugriffsversuche erfolgen.

Andererseits entsteht durch den Einsatz entsprechend leistungsfähiger Edge-Devices ein Überfluss an Computerleistung in den zentralen Leistungsmerkmalen der Hardware. Ein einfaches Kleinstsystem wie der Raspberry Pi 4 hat beim Einsatz eines modernen High-End-Netzes wie dem EfficientNet B3 in Inferenzaufgaben eine Speicherauslastung im einstelligen Prozentbereich. Beim Anschluss eines Beschleunigers wie der Coral TPU [11] sinkt auch die beobachtbare CPU-Leistung (des 4-core Prozessors) auf unauffällige Werte.

Edge-Devices können einen großen Teil solcher sicherheitsrelevanter KI-Funktionen abbilden und lokal erkennen, ob ein verändertes Verhalten vorliegt, sowie ggf. entscheiden, wie mit externen Anfragen weiter zu verfahren ist. Allerdings kann das lokale Device nicht erkennen, ob in der Gesamtinfrastruktur ein endpunkt- und teilnehmerübergreifendes verändertes Verhalten vorliegt. Hierzu müsste eine KI an einer zentralen Stelle bereitgestellt werden, die ggf. Daten aus den Edge-Devices zur übergreifenden Auswertung erhält und zentral auswertet. Die Lokation einer solchen zentralen KI ist zurzeit noch nicht festlegbar, da die Architektur der Infrastruktur noch nicht (vollständig) definiert ist.

Zusammenfassend ist festzustellen, dass KI-Systeme durch Machine Learning mit hohem Ressourcenaufwand in Serverpools einer Cloud entstehen, aber in softwaretechnisch optimierter Form für dezentrale kleine Geräte in der Edge des Internets und sogar darüber hinaus in Teilen der Welt ohne Elektrizitäts- und Kommunikationsnetze zum Einsatz kommen. Die Reduktion auf Flatbuffer und I8-Quantisierung eignet sich nicht für Training, aber sehr wohl für Inferenz, wenn nötig auch mit zweistelligen Milliarden von Rechenoperationen pro Sekunde zum Preis eines T-Shirts. Dabei können energieautarke Sensornetze sich in zivilisationsferne Gebiete ausbreiten, sich selbst resilient konfigurieren und nur noch auf wenige Knoten angewiesen sein, die eine Verbindung zur Welt mit Elektrizität und Kommunikation aus einem festen Netzwerk brauchen. KI kann überall sein und ist bereits in vielen Bereichen heute selbstverständlich. Geräte aller Arten können sehen, hören, verstehen, Sprachen sprechen, Handzeichen geben und natürlich vielfältigste andere Sensorsignale wiedergeben. Jeder weiß, dass ein modernes Auto zumindest im Hinblick auf den Preisanteil der bestellten Features in erster Linie ein KI-basiertes System ist, das nicht über ein Telekomnetz gesteuert wird, sondern seine Funktion aus lokaler Intelligenz ableitet. Wer redet nicht mit seinem Navigationssystem, verlässt sich nicht auf wahrende Assistenten, genießt nicht den Komfort des elektronischen Fahrwerks und Antriebsstrangs ... das ist KI an der Edge und stellt nur ein unscharfes Abbild davon dar, was diese Technologie in Kernbereichen wie der Industrie 4.0 heute an der Edge bewirkt.

In diesem Papier werden wir erklären, dass diese Fähigkeiten nicht ohne Risiken kommen, denn schlaue Edge-Devices können auch Böses im Schild führen.

2.2 KI in der Cloud

Inzwischen sind Netzarchitekturen entstanden, die bewusst auf effiziente Skalierung zielen. Konkret bedeutet dies, dass die Lademodule bei gegebener Erkennungsqualität verschlankt werden und die Auflösung der Eingaben gleichzeitig erhöht wird. (Beispiel: Google EfficientNet B0-B8, compound scaling von 224X224 bis 672X672 mit nur 5,3 bis ca. 80 Millionen Parametern). Solche Architekturen sind auch wichtig, damit der Ressourcenaufwand für Training mit den qualitativen Zielsetzungen der Inferenz harmoni-

siert werden kann. Dieses Training muss weiterhin in einer umfangreich ausgestatteten Cloud erfolgen und stellt einen erheblichen Kostenaufwand dar.

Ein Cloud-Einsatz für Inferenz ist nur noch für extrem feingranulare Klassifizierung nötig und erfolgt dann in Kooperation mit der Edge für fünf- bis sechsstellige Klassenzahlen. Dies sind Netze wie die bekannte Google-Lens. In industrieller Mustererkennung kommen solche Anforderungen heute aber nicht vor. Sie zielen eher auf den Privatanwender, der mit seinem Smartphone Informationen über seine Lebensumgebung sucht und Pflanzen, Tiere, Autos oder Konsumgüter erkennen und vergleichen möchte.

An dieser Stelle muss auch darauf hingewiesen werden, dass ein Training für typische industrielle Anwendungen heute mit viel geringerem Aufwand erreicht werden kann als noch vor ca. drei Jahren. Die Technik des Transfer-Learning ist inzwischen so weit perfektioniert, dass auf Basis der berühmten Netzarchitekturen, die einst Weltrekorde im ILSVRC Wettbewerb (ImageNet Large Scale Visual Recognition Challenge) aufgestellt hatten, mit minimalem Aufwand hochleistungsfähige Netze erzeugt werden können. Praktisch alle Netze dieser Art sind heute als Open Source Software (OSS) unter typischen OSS-Lizenzen wie Apache 2.0 inklusive aller Gewichte, erzeugt mit ILSVRC2012-Daten, verfügbar. Definiert man, wie im Kontext industrieller Anwendungen üblich, ca. 20 neue Klassen und ersetzt damit die Klassifikationsschicht aus den 1000 ImageNet-Klassen, dann benötigt ein Training für diese Klassen in der Regel weniger als 24 Stunden, einschließlich eines Fine-Tunings, wenn man weniger als die letzten drei bis fünf Layer auftaut, den Rest der Gewichte aber fixiert hält. Solche Trainings für ein derartiges OSS-Netz erreichen Trefferraten von über 98 Prozent (nach Top-1-Kriterium). Der Kostenaufwand für ein Training dieser Art ist also marginal und kann sogar mit sehr überschaubarem Bestand an Trainingsdaten erfolgen, wenn moderne Augmentierungstechniken zum Einsatz kommen.

Für den Vorgang des Trainings großer neuer Architekturen außerhalb der Basissysteme für Transfer-Learning-Einsätze ist eine neue Klasse der realisierbaren Komplexität im Entstehen. Mit dem Einsatz von KI zur Bestimmung der optimalen Trainingsstrategie (in Systemen wie AutoML von Google) öffnen sich neue Horizonte für erreichbare Präzision. Heute sind Trainingsvorgänge vor allem deshalb so ressourcenintensiv, weil viele Kombinationen von Hyperparametern in heuristischer Vorgehensweise zum Training getestet werden müssen, um optimale Resultate in dem einen Lademodul zur Inferenz zu erreichen, das schließlich zum Einsatz kommt. KI-basierte Strategien werden es ermöglichen, kürzere, kostengünstigere Wege zum Ziel zu finden. Hier entsteht auch ein neuer Markt für proprietäre Software mit sehr hohem Wertversprechen.

Solche Systeme könnten auch den Einsatz von Reinforcement Learning (RL) und Generative Adversarial Networks (GAN) stark beflügeln, die heute nur extrem ressourcenintensiv und deshalb mit hohen Kosten verfolgt werden können. Dies sind die Teildisziplinen der KI, die über die erwiesenen übermenschlichen Fähigkeiten des Machine Learning in der Mustererkennung hinausgehende, kreative Fähigkeiten ermöglichen, die erst in Ansätzen erforscht sind. Die kreativen Handlungen des ALPHA-GO im Kampf um die Vorherrschaft auf dem Sektor der strategischen Spiele bieten ein Beispiel für überlegene maschinelle Handlungen, die biologische Gehirne nicht mehr verstehen können. GAN werden in der Erzeugung von Mustern aller Arten ähnliche Fähigkeiten zugetraut, die zu Fortschritt in technischer, medizinischer und allgemeiner wissenschaftlicher Forschung beitragen können.

Zusammenfassend lässt sich also feststellen, dass moderne KI in der Cloud Ergebnisse liefern wird, die zu immer mehr übermenschlicher maschineller Fähigkeit besonders in den Bereichen führen, die früher als differenzierend für das Gehirn galten. Kreativität und Strategiebildung galten bislang als menschliche Differenzierung, während die Leistungen von Sinnesorganen ja schon lange als technisch realisierbar galten. Beispielsweise können Hunde besser riechen als Menschen, Vögel besser sehen … es war schon lange klar, dass solche Fähigkeiten von Maschinen übernommen und weitaus besser ausgeführt werden können. Mit der Verbreitung von KI in kleinsten Edge-Devices, die überall auftreten und teils gar nicht mit KI in Verbindung gebracht werden, die aber die Lademodule aus solchen Netzen mit hohem Niveau an Fähigkeiten ausführen und dabei gleichzeitig Raum für Manipulation im selben Device lassen, entsteht ein neues „Schlachtfeld“ für IT-Security in der Industrie 4.0.

2.3 Globaler KI-Markt

Es ist allgemein bekannt, dass die finanzielle Förderung der Weiterentwicklung von KI heute mit sehr großem Abstand zu den Budgets im Rest der Welt in den USA und China erfolgt. Dreistellige Milliardenbeträge von staatlichen Organisationen, verbunden mit sehr großen Aufwänden privater Unternehmen im Internetsektor, sorgen dafür, dass heute fast alle weltweit anerkannten Publikationen, fortgeschrittene Produktentwicklungen und strategische Kursbestimmungen in Nordamerika und China entstehen. Wer heute nach aktuellen hochtechnischen wissenschaftlichen Publikationen zum Thema KI im Internet sucht, erhält in den gängigen Suchmaschinen anfangs fast ausschließlich amerikanische Fundstellen und geht dann allmählich in den Bereich von Suchergebnissen, die nur noch in Kanji-Schriftzeichen dokumentiert sind. Ergebnisse aus Europa, insbesondere aus Deutschland, existieren fast gar nicht. KI-Hochtechnologie findet hier zumindest in dieser Hinsicht nicht statt und jeder, der solche Suchen aufsetzt, kann dies deutlich sehen.

Die Führungsrolle in KI ist fokussiert auf wenige Unternehmen (Apple, Facebook, Google, Baidu, Amazon, Microsoft), Universitäten (Stanford, Berkeley, NYU, Montreal, Toronto, MIT, Oxford) und einige vorwiegend chinesische Regierungsorganisationen. Angesichts der Tatsache, dass die Erklärbarkeit von KI durch menschliche Gehirne weiterhin schnell abnimmt und der Perspektive, dass KI-basierte Erklärungssysteme ausschließlich in Nordamerika und China entstehen, gibt es Anlass zu Besorgnis. Erschwerend hinzu kommt der Umstand, dass sich die Lücke zwischen den wissenschaftlichen KI-bezogenen Fähigkeiten dieser Länder und Europa stetig vergrößert.

Die vor kurzem veröffentlichten Backdoors in IT- und OT(!)-Systemen [12], [13], [14], [15] gelten inzwischen als den Herstellern seit Langem bekannt. Umso erstaunlicher ist die Feststellung, dass die betroffenen Hersteller erst handeln, nachdem bei deren Kunden diese Angriffe ihre verheerenden wirtschaftlichen und geopolitischen Wirkungen gezeigt haben. Es gibt eine gefährliche „Lücke“ zwischen Erkennung und Behebung dieser Gefahren, welche nicht selten durch staatspolitische bzw. wirtschaftliche Bewertungsmaßstabe aufrechterhalten wird. Gemessen daran sind Trade-offs in KI-Lösungen in der Zukunft nicht anders zu bewerten.

3 Typische Anwendungen von KI in der Industrie

Nachdem in den ersten Kapiteln der Status, Trend und die durch KI ausgelösten strukturellen Veränderungen der Industrie betrachtet wurden, werden in diesem Kapitel die derzeit typischen Anwendungen von KI in der Industrie beleuchtet. Dazu werden Limitierungen und Kriterien für die richtige Wahl von KI-Lösungen vorgeschlagen.

Typische Anwendungen von KI in der Industrie finden sich unter anderem in der

- Automatisierung der Produktion,
- Steuerung von Industrieantrieben (Motion-Control),
- bedarfsgesteuerten Wartung des Maschinenparks,
- Prozessoptimierung,
- Qualitätssicherung durch zerstörungsfreie Prüfung.

Diese Anwendungen nutzen vor allem Sensortechnologien für die Gewinnung der nötigen Parameter. Neben den bildgebenden Sensoren (Bild, Farbe, Licht) wird im industriellen Bereich eine Vielzahl weiterer Sensoren zur Messung von Temperatur, Druck, Zug, Beschleunigung, Drehmoment, Berührung u. v. a. m. eingesetzt.

Die derzeit diskutierten KI-Methoden basieren auf maschinellem Sehen, Hören und Kommunizieren durch Nutzung der Signale von bildgebender und akustischer Sensorik und weniger auf Nutzung der anderen, erwähnten Sensorik. Sie werden eingeteilt in:

- Bild-/Video-Erkennung
- Spracherkennung
- Inhaltsangaben von Texten
- Stimmungsanalyse
- Beurteilung von Verträgen und Vertragspartnern sowie andere Verwaltungsanwendungen

Als besonders erfolgreich hat sich die Mustererkennung durch Verwendung von Deep-Learning mit Neuronalen Netzen erwiesen. Dies resultiert zum einen aus dem herausragenden Fortschritt bei der Entwicklung und Implementierung der Algorithmen für Neuronale Netze und zum anderen aus der Verfügbarkeit von Rechenkapazität und -geschwindigkeit in der Cloud, die letztlich die Anwendung von Neuronalen Netzen praktikabel macht. Ein weiterer Erfolgsfaktor ist der Umstand, dass die Algorithmen und Tools als Open-Source-Lösungen verfügbar sind.

Neuronale Netze werden mit großen Datenmengen trainiert. Für das Training stehen in der Cloud verschiedene Optionen auf speziellen Servern zur Verfügung. Beispiele für solche Technologien sind sogenannte TPU (Tensor Processing Unit)-Systeme von Google. TPUs sind speziell entwickelte Microchips, die die Parameter von Neuronalen Netzen während des Lernprozesses berechnen und kontinuierlich optimieren. Hierdurch wird das maschinelle Lernen stark beschleunigt. Andere ‚Lernbeschleuniger' sind IPU (Intelligence Processing Unit)-Systeme von Graph-

core, die sogenannte ‚In-Processor-Memories' realisiert haben. Dadurch erreichen sie eine sehr hohe Geschwindigkeit für das Training von Neuronalen Netzen. In der Cirrascale Cloud wird diese Technologie als Service zur Verfügung gestellt. Anbieter wie Baidu, Nvidia und andere bieten Open-Source Software für Deep-Learning an. So unterstützt beispielsweise PaddlePaddle von Baidu distributed computing zur effizienten Nutzung vieler Cloud-Server. Nvidias Cuda-X AI läuft am schnellsten auf Servern mit Nvidia-GPUs, die ebenfalls bei vielen Cloud-Anbietern zur Verfügung stehen.

Ein grundlegendes Problem beim Trainieren von Neuronalen Netzen ist die Verfügbarkeit von Trainingsdaten in hoher Anzahl und Qualität. Um eine ausreichend hohe Erkennungsrate zu erreichen, werden zigtausende von Trainingsdaten benötigt. In der Praxis stehen aber fast nie genügend Trainingsdaten für die eigene spezifische Anwendung zur Verfügung. Meist hat man nur wenige Datensätze zur Hand.

Einen Ausweg aus diesem Problem erreicht man mithilfe von Transfer-Learning. Für Transfer-Learning startet man mit einem vor-trainierten Netzwerk, das als Open-Source zur Verfügung steht. Dieses wird dann für die spezifische Anwendung mit einigen hundert Datensätzen nach-trainiert. Dadurch werden nur die letzten Lagen des Neuronalen Netzes auf das eigentliche Problem adaptiert. Dies funktioniert recht zuverlässig, wenn vor-trainierte Netze aus ähnlicher Anwendungsklasse, z. B. Bilderkennung, als Ausgangspunkt gewählt werden. Auch hier werden jedoch bisher die anwendungsspezifischen und proprietären Datensätze zum Nach-Trainieren in die Cloud hochgeladen. Dies stellt unter Umständen ein Sicherheitsproblem und ein Problem für den Schutz des eigenen Knowhows (IP) dar.

Transfer-Learning ist mittlerweile mit angemessenem Aufwand an Rechnerleistung auch auf eigenen Rechnern möglich, so dass das Hochladen der proprietären Daten in die Cloud entfallen kann.

Deep-Learning mit Neuronalen Netzen hat immer ein Ziel: das Erkennen von Datenmustern. Darin ist KI sehr gut und weit besser als der Mensch. Das Ergebnis hängt jedoch von den vorher trainierten Datensätzen ab. Sind in den Trainingsdaten versteckte Annahmen bzw. Verzerrungen (Bias) enthalten, werden diese automatisch auf die Ergebnisse übertragen. Auch ist KI nicht in der Lage, bei fundamental falschem Ergebnis die Ursache zu klären (siehe Erklärbarkeit von KI-Ergebnissen [4], Kapitel 1).

Limitierungen von KI-Lösungen

- Obwohl Trefferquoten von KI-Lösungen signifikant höher sind gegenüber bisherigen Standardverfahren, erreichen KI-Anwendungen keine hundertprozentige Abdeckung, da es nicht möglich ist, den gesamten Datenraum mit Beispielen zu trainieren.
- Ursachen von Fehlern sind für KI-Anwendungen schwer zu analysieren (limitierte Erklärbarkeit [4], Kapitel 1).
- Eine ungenügende Qualität der Trainingsdaten mit nicht erkanntem Bias repliziert diesen Bias und verzerrt somit das Ergebnis.
- Bei unerwarteten Problemen von KI-Anwendungen (Out-of-the-Box) kann (bisher) nur der Mensch helfen.
- KI-Lösungen enthalten per se keine Sicherheitsmaßnahmen gegen Angriffe. Daher müssen diese zusätzlich als Schutzschirm (Zwiebelschalenmodell) implementiert und mit entsprechenden Authentisierungsmethoden Zugriffsmodelle definiert werden.
- Implementierung von KI erfordert Expertenwissen.

Die Analyse von Sensornetzwerken mit Deep-Learning ist zum jetzigen Zeitpunkt noch unterrepräsentiert, da die verfügbaren Datensätze nicht immer den geeigneten Startpunkt für die eigene Anwendung liefern. Die erste Adresse für Datensätze findet man bei „Top Sources For Machine Learning Datasets" [16]. Auch stellt sich die Frage, inwieweit Deep-Learning hier generell der geeignete Weg ist, oder ob andere KI-Methoden wie z. B. SVM oder Gradient Boosted Decision Trees etwa mit Random Forest für die einige Anwendung geeigneter sind.

Kriterien für die Wahl von KI für industrielle Anwendungen

- Anwendung basiert auf maschinellem Sehen, Hören, Kommunizieren mit Text in natürlicher Sprache, Strategiebildung in komplexen Entscheidungssituationen, beispielsweise zur Steuerung autonomer Systeme
  - Deep Learning mit Neuronalen Netzen, vorzugsweise unter Einsatz von Transfer-Learning mit vortrainierten Netzen, soweit verfügbar

- Auswertung von Sensornetzwerken
  - einfache Anomalieerkennung mit Priorität auf Erklärbarkeit gegen maximal mögliche Trefferquote
    - ML Clustering mit entsprechenden statistischen Algorithmen
  - hochkomplexe Sensornetzwerke
    - Neuronale Netze + Transfer-Learning auf Servern vor Ort vermeidet das Laden proprietärer Daten in die Cloud und minimiert Sicherheitsrisiken wie den Verlust von IP.
      Voraussetzung: Es gibt ein vortrainiertes Neuronales Netz passend für diese Anwendung, um den Trainingsaufwand vor Ort zu minimieren. Ein solches Netz muss die gleiche Architektur mit einem geeigneten Input-Layer bereitstellen, auf den die verfügbaren Daten ohne Qualitätseinbußen angepasst werden können. Ebenfalls muss eine geeignete Schnittstelle zur Übergabe der Inferenzzwischenergebnisse aus den eingefrorenen Schichten zur neu geschriebenen Klassifizierungsschicht zur Verfügung stehen.
      Natürlich kann man ein Convolutional Neural Network (CNN) aus der Bilderkennung nicht verwenden, um diesem die Fähigkeiten eines Long **short-term** memory (LSTM) zur Spracherkennung zu vermitteln. Besonders für Bilderkennung erzielt Transfer-Learning aber erstaunliche Leistungen, da in den Faltungsschichten dieser Netze vor allem die Features zur Verarbeitung grafischer Primitive auftreten. Sie sind von den erkannten Bildinhalten weitgehend unabhängig. Deshalb ergibt Vortraining mit ImageNet-Daten (allgemeine Bilder von Tieren, Pflanzen, Gebäuden, Verkehrsmitteln …) oft eine sehr gute Basis für völlig andere Aufgaben grafischer Erkennung wie medizinische Diagnose oder zerstörungsfreie Prüfung in der produzierenden Industrie, u. a. allgemeine Voraussetzungen:
- Verfügbarkeit von geeignetem In-house-Expertenwissen oder Beraterfirmen.
- Ein Sicherheitskonzept muss in jedem Fall für die zugriffs- und integritätsgesicherte Speicherung und Verarbeitung extern zugekaufter und während der Nutzung entstehender KI-Trainings- und Anwendungsdaten über den gesamten Lebenszyklus durch maschinelle und menschliche Prozessbeteiligte entwickelt werden.
- Klärung vertragsrechtlicher Themen wie Rechte an den Daten oder daraus entstehender IP auch als Voraussetzung für die Anwendung eines geeigneten Sicherheitskonzepts.

Technologieausblick: TPU-ICs bestehen aus komplexen Rechner-Cores, die auf einem IC parallel in maximal möglicher Anzahl implementiert und verbunden werden. Die maximal mögliche Anzahl an Cores wird durch die gewählte Siliziumtechnologie und dann durch den Preis des ICs bestimmt (Preis–Leistung). Die Silizium-Technologie ermöglicht durch kontinuierliches Shrinken in neue Silizium-Prozessknoten, beispielsweise auf die halbe Strukturbreite bei gleicher Siliziumfläche, die Implementierung von viermal so viel Funktionen, in diesem Beispiel von viermal so viel TPU-Cores. Dies wird durch das Moore'sche Gesetz beschrieben, wonach sich die Rechnerleistung pro Jahr verdoppelt. Allerdings verliert das Moore'sche Gesetz beim Vordringen der Silizium-Technologie in den Bereich unter 20nm-Strukturbreite zunehmend an Gültigkeit. Dies kann unter Umständen durch bessere Architekturkonzepte für die Implementierung der Algorithmen zumindest teilweise kompensiert werden. Zum Beispiel werden die neuesten TPU-Cores zurzeit in 7nm-Technologie [17], [18] produziert und von der Rechnerarchitektur sind diese – abgesehen von einem nicht vorhersehbaren neuen Ansatz – bereits recht gut optimiert. Für die nächste Dekade sehen wir noch drei Shrink-Schritte bis zur 2nm-Technologie voraus. Die IPUs von Graphcore werden ebenfalls in 7nm-Technologie gefertigt. Da jeder dieser Shrinks die Cores bei etwa gleichem IC-Preis um das 2,5-fache schneller machen wird, würde sich daraus eine Beschleunigung der Rechnerleistung für IPUs und TPUs um den Faktor 15 in dieser Dekade ergeben, sofern das Problem der Energiedissipation in den ICs angemessen gelöst werden kann.

4 Sicherheitsaspekte für Industrie 4.0-Anwendungen

Im Kontext von Industrie 4.0 stellen Sicherheitsmerkmale Qualitätsmerkmale im weitesten Sinne dar. Da diese Eigenschaften keine inhärenten Bestandteile neuer Technologien, einschließlich KI, sind, widmet sich dieses Kapitel den grundlegenden Fragen der Sicherheit in KI-Anwendungen. Es bildet eine Überleitung von den vorangegangenen Kapiteln zum nachfolgenden Kapitel über die Anwendbarkeit von KI auf die verschiedenen sicherheitsrelevanten Aspekte von Industrie 4.0.

Ergänzend zur oben betrachteten beachtlichen Leistungsfähigkeit von KI im Allgemeinen und im Lichte bereits vorhandener KI-Anwendungen im industriellen Kontext ist allerdings zu beachten, dass die Einsatzszenarien im Bereich Industrie 4.0 in besonderem Maße auf Werten wie „Vertrauenswürdigkeit“ und „Sichere unternehmensübergreifende Kommunikation“ aufbauen. Wie sich zeigt, ist es für Menschen nicht immer einfach und anschaulich verständlich, woher sie die Evidenz beziehen sollen, um KI-Entscheidungen zu vertrauen, wenn die Ergebnisse nicht eingängig und verständlich sind. Schemata, die eine systematische Prüfbarkeit oder Zertifizierbarkeit im herkömmlichen Sinne ermöglichen, sind bisher nicht bekannt. Hierzu sind weitere Methoden zu entwickeln, um insbesondere den bisweilen unsichtbaren und auch sicherheitsrelevanten Einfluss von Verzerrungen (Bias) im Bewertungsprozess von KI zuverlässig zu erkennen.

Wie bereits in zwei vorhergehenden Publikationen [5] [6] erläutert wurde, existieren an dieser Stelle noch immer besondere Herausforderungen, die sich primär auf die Sicherheit von Komponenten, Maschinen und den Betrieb beziehen.

Die speziellen Industrie 4.0-Security-Anforderungen, die bisher durch weitestgehend etablierte kryptographische Methoden und Prozesse unterstützt werden, sind einerseits dahingehend zu untersuchen, wie durch gehärtete KI-Anwendungen im Bereich des Shopfloor, vornehmlich in Edge-Systemen, Verbesserungen der Security erreicht werden können. Inwiefern Edge-Systeme andererseits durch diese *gehärteten* KI-Anwendungen gegnerische Generative Adversarial Networks (GAN)-basierte KI-Angriffe abwehren können und dadurch im Vergleich zu konventionellen Intrusion Detection Protection (IDP)-Systemen eine höhere Resilienz ausweisen können, ist ebenfalls zu untersuchen. In jedem Fall werfen neue leistungsfähige Edge-Systeme auch neue Fragen in alle Richtungen hinsichtlich ihrer Einsatzmöglichkeiten, ebenso wie zu den möglichen „Trade-offs“ auf. Diesen Betrachtungen und Fragen mit Bezug auf „Industrial Security und der Entwicklung von KI-Anwendungen in der Edge“ wurde in der Industrie bislang kein besonderer Stellenwert beigemessen, zumal Edge-Architekturen erst im Aufkommen sind. Securitythemen wurden bislang erst spät betrachtet.

4.1 Relevanz von Sicherheitsanforderungen an die KI in der Industrie 4.0

Generell ist sicherzustellen, dass durch den KI-Einsatz in Industrie 4.0-Komponenten keine Verschlechterungen bezüglich der bestehenden Security-Lagen stattfinden bzw. sich auf den Status quo nicht negativ auswirken. Insofern geht es in der Diskussion beim Einsatz von KI-Anwendungen für Industrie 4.0 auch um die Frage, wie KI-Leistungen,

die bereits auf dem Markt sind, als auch kommende Entwicklungen Einfluss auf die geforderten Sicherheitseigenschaften von Komponenten, Maschinen und den Betrieb haben.

Dabei steht außer Frage, dass künftige Systementwickler nicht mehr unbedingt mit denselben Entwicklungstools der Vergangenheit oder Gegenwart werden arbeiten können, um Konstruktionsmerkmale und Sicherheitsanforderungen gegenüber KI-Anwendungen zu validieren. Diese KI-Anwendungen dürfen keine nachträglich „angeflanschten" oder aufgesetzten Technologien sein, sondern sie sind als inhärente Leistungsmerkmale des Designs von Komponenten und Maschinen bis hin zum Betrieb zu behandeln.

Vergleichbare Ansinnen bestehen bereits hinsichtlich der **Security-by-Design**-Forderung bei der Produktentwicklung gemäß ISO/IEC 62443. Analog hierzu lautet die Forderung für KI-Anwendungen im etablierten Maschinenbau in diesem Kontext u. a. die Beachtung von **Integrity-by-Design**. Damit würde das Ziel verfolgt, nicht ungehemmt der beliebigen Integration von KI-Anwendungen zu folgen, sondern ingenieurmäßige Integritätsmaßstäbe zu fordern, mit dem Ziel, die aus menschlicher Sicht mögliche Beherrschbarkeit von Systemen zu erreichen. Generell sollte es im Bereich Industrie 4.0 kein „blindes Vertrauen" für KI-Ergebnisse geben, sondern nachgewiesenes Verhalten und ein grundsätzliches Verständnis des Phänomens möglicher Verzerrungen. Zudem sollten Inferenz-Systeme in der Edge optimalerweise gegen Verletzbarkeiten und unerwünschte Seiteneffekte abgesichert werden.

4.2 Anwendung bekannter IT-Security-Vertrauensmodelle und -Präventionsmethoden auf Industrie 4.0-KI-Szenarien

Bei konsequenter Anwendung der vorangegangen Security-Überlegungen ist das aus dem Bereich digitaler Identitäten bekannte „Zero-Trust-Konzept" als Blaupause nutzbar. Dabei wird der Gedanke verfolgt, dass bei einer hohen Zahl agiler und ständig neuer Kommunikationsteilnehmender diese mit einem Zero-Trust-Anfangswert belegt werden und sich das Vertrauen der anderen Beteiligten im Rahmen eines Trust-Scoring erst „verdienen" müssen.

Übersetzt bedeutet das: Im Idealfall folgt jegliche Industrie 4.0-Kommunikation mit anderen Komponenten ausschließlich dem vorgenannten „Zero-Trust-Prinzip". Danach werden vor jedem Zugriff auf Anwendungen, Daten, Sensoren und Aktoren zuerst die beteiligten Identitäten und Berechtigungen geprüft. Erst nach deren Validierung und Gewährung werden Kommunikationsverbindungen freigeschaltet und Anwendungen sowie Datenräume für die menschlichen und maschinellen Nutzer sichtbar. Zusätzlich wird jeder Zugriff mit relevanten Metadaten protokolliert, um im Störungs- oder Angriffsfall forensische Analysen zu unterstützen. Dieses Prinzip kommt in verhaltensbasierten Intrusion Detection Systemen (IDS) und Intrusion Prevention Systemen (IPS) bereits zum Einsatz und kann in industriellen KI-Szenarien als Vorbild dienen. Inwieweit durch zusätzliche vorausschauende KI-Analytik die erforderliche Trustworthiness [19] von Entscheidungen der eigentlichen KI-basierten Systeme ermittelt, bewertet und erklärt werden kann, ist noch Forschungsgegenstand. Hierzu werden dynamische Veränderungen der beobachteten Systemmetriken im Hinblick auf ihre Plausibilität bewertet. Dies zielt auf KI-überwachte bzw. KI-erklärte Entscheidungen von KI-Systemen ab.

In realen Industrie 4.0-Implementierungen, u. a. auch im Bereich Kritischer Infrastrukturen, besteht die Herausforderung darin, in den nächsten Jahren eine vertrauenswürdige Kombination aus neuen, I4.0-nativen Anlagen mit bestehenden, bisweilen langfristig investierten Anlagen zu integrieren.

Im Vorfeld eines KI-basierten I4.0-Projektes sollte geklärt werden, ob ggf. ein Digitalisierungsprojekt erforderlich ist, um aus einem „Brownfield" eine homogene digitalisierte Prozessumgebung zu schaffen. Alternativ kann auch zusätzliche IIoT-Sensorik (z. B. Wärmebildkamera zur Temperaturüberwachung) verwendet werden, um Betriebsparameter zu bestimmen, die eben noch nicht digital zur Verfügung stehen. Um keine unbeabsichtigten Einfallstore für Sicherheitsvorfälle und Konfigurationsfehler zu schaffen, ist zu prüfen, ob und wie Kommunikationsschnittstellen und Protokolle durch industrietaugliche Hard- und Software auf einheitliche Standards übersetzt werden können. Edge-basierten KI-Systemen in diesem Kontext einer im Umbruch befindlichen Fertigungslandschaft einen ange-

messenen Platz im Rahmen vertrauenswürdiger standardisierter Systeme zuweisen zu können, stellt die eigentliche systemische Herausforderung dar. Kurz formuliert: Welcher KI-Entscheidung wird wann, von wem begründbar vertraut?

4.3 Interoperabilität auf organisationstechnischer Ebene

Im Zuge der Euphorie, die Potenziale neuer leistungsfähiger autonomer KI-basierter Systeme zu heben, stellen sich zugleich Interoperabilitätsfragen:

Folgen die neuen KI-basierten Edge-, Edge-Cloud- bzw. Cloud-Topologien in der industriellen Fertigung etablierten Sicherheitsstandards? Kann es hierbei zu Kompetenzgerangel auf den unterschiedlichen Ebenen im Netzwerk kommen? Wie wirken sich autonome Entscheidungen von Komponenten und Maschinen innerhalb einer historisch eher hierarchisch gewachsenen Prozesskontrollstruktur aus? Wie stark lassen sich KI-basierte autonome Maschine-zu-Maschine Entscheidungen vorantreiben und am Ende systemisch betrachtet trade-off-frei beherrschen?

Diese und andere wichtige Fragen bezüglich organisatorischen Interoperabilitätsanforderungen lassen sich im Rahmen dieses Dokuments nicht abschließend behandeln. Diese Fragen zeigen jedoch die Bedeutung und den Bedarf, bei der Einführung KI-basierter Systeme organisatorisch, konzeptionell und ganzheitlich zu planen. Die in den vorangegangenen Kapiteln aufgezeigten Stärken und möglichen Unwägbarkeiten von KI-Systemen haben Einfluss auf die funktionale Interoperabilität. Damit haben sie auch Auswirkungen auf die Stabilität von Industrie 4.0-Prozessen.

4.4 Berücksichtigung von Anforderungen an Industrie 4.0-Security-Maßnahmen

Wenn KI-Anwendungen in der Industrie 4.0 als „the new electricity“ Einzug halten sollen, um in diesem Bild aus dem Kapitel 1 zu bleiben, sind die Merkmale dieser Elektrizität absolut mission-critical. Schlimmstenfalls versagen alle Komponenten, wenn diese ‚Energie‘ zu gering ausfällt, oder es kommt zu Havarien, weil die ‚Energie‘ zu hoch, zu instabil, nicht berechenbar ist oder anderweitige Zustände unerklärbar sind. Diese Szenarien sind unter allen Umständen zu vermeiden.

Die Forderung nach Integrität und Stabilität, bestenfalls auch standardisierten Interfaces, erhält eine fundamentale Bedeutung. Um die Sicherheitsthemen der Industrie 4.0 mittels KI lösen zu können, sind einige Grundvoraussetzungen zu erfüllen. Exemplarisch werden nachfolgend einige Punkte zum besseren Verständnis genannt:

- Verständnis industrieller Sicherheitsanforderungen, speziell der Industrie 4.0
- Verständnis für die grundsätzliche Funktionsweise und Bedeutung von KI-Anwendungen und Einschätzung der Ergebnisse
- Schaffung geeigneter Modelle für das Zusammenwirken von KI und Sicherheit
- Schaffung von sicheren Modellen zur Validierung der Integrität
- Entwicklung geeigneter Prüfungsschemata und Methoden zur Zertifizierung von KI-Modellen unter Berücksichtigung der Prozesse zur Datenbeschaffung, Modellbildung, Maintenance, Überwachung im Feld etc. und des Betriebs

Im nachfolgenden Kapitel 5 wird näher untersucht, ob und wie sich KI-Anwendungen in den einzelnen sicherheitsrelevanten Themen von Industrie 4.0 mit welchem Mehrwert abbilden lassen. Dazu werden ausgewählte Aspekte aus der Arbeit der gesamten Arbeitsgruppe „Sicherheit vernetzter Systeme“ der Plattform Industrie 4.0 betrachtet.

5 KI im Kontext von Sicherheitsthemen der Industrie 4.0

In diesem Kapitel wird die Anwendbarkeit von KI auf die verschiedenen sicherheitsrelevanten Aspekte von Industrie 4.0 diskutiert und untersucht. Das Kapitel enthält hierzu entsprechende Industrie 4.0-spezifische Begriffe aus verschiedenen Diskussionspapieren der Plattform Industrie 4.0 zum Thema Sicherheit. Auch wird die Verwendbarkeit von KI bezüglich der verschiedenen Aspekte beispielhaft dargelegt, d.h. es werden Antworten auf folgende Frage gegeben:

Wie kann KI **aus Sicht von Sicherheitsaspekten** in die verschiedenen Themengebiete der Industrie 4.0 einfließen und helfen, ein erhöhtes Sicherheitsniveau zu erreichen?

5.1 Diskussion der Industrie 4.0-Sicherheitsthemen

Im Folgenden werden die verschiedenen Themengebiete der AG „Sicherheit vernetzter Systeme“ im Einzelnen betrachtet.

5.1.1 Die I40-Wertschöpfungskette

Beim Aufbau einer neuen Wertschöpfungskette[2] kann KI einen wichtigen Beitrag leisten, KI kann Anbieter in kurzer Zeit bezüglich ihrer Eignung zur Teilnahme in einer Wertschöpfungskette beurteilen. In die Beurteilung eines Teilnehmenden einer Wertschöpfungskette können verschiedenste Attribute einfließen, wie z.B. Lieferzuverlässigkeit, Zahlungsmoral, Gerüchte oder Berichte zum Zustand des Unternehmens, Qualitätsinformationen zu den Produkten, frühere Erfahrungen, Preispolitik, Umweltschutz, Marktbeherrschung, die Gesamtwirtschaftslage, lokale oder globale Pandemien in Regionen oder Ländern, oder der Umgang damit. Durch die mittlerweile massiv erhöhte Menge an verfügbaren Daten über alle diese Aspekte wird es dem Menschen schwer gemacht, die Beurteilung hinreichend gut durchzuführen, da sich die relevante Datenmenge und deren Filterung zu einem „Big-Data“-Problem aggregiert.

Eine KI kann unter Verwendung dieser Attribute eine Bewertung des Teilnehmenden auf Basis von Scores in Form einer Empfehlung abgeben. Dadurch können die Kandidaten in eine Wunschreihenfolge überführt werden und entsprechend der Reihenfolge zur Teilnahme eingeladen werden.

Das gleiche Auswahlverfahren kann beim Austausch von Teilnehmenden, auch unter Zeitnot, durchgeführt werden. Eine Zwischenbewertung der Teilnehmenden kann auch während des Produktivbetriebes der Wertschöpfungskette stattfinden, beispielsweise können politische Veränderungen, **Naturkatastrophen oder die Ausbreitung von Krankheiten schnelles Handeln und den Austausch eines Teilnehmenden nötig machen.**

2 Eine Beschreibung der Eigenschaften der I40-Wertschöpfungskette befindet sich in der Studie „IT-Sicherheit für Industrie 4.0“ [25].

Dabei kann die KI ggf. kontinuierlich mittels gezielter Update-Trainings weiterentwickelt werden, da durch die finale Entscheidung durch den Menschen letztendlich ein „Labeling“ der Auswahl stattfindet, welches bei Folgeentscheidungen berücksichtigt werden kann.

Häufig ist die Anzahl der in Frage kommenden Teilnehmenden an einer Wertschöpfungskette auf dem Markt begrenzt, da sich nur Konkurrenz den Markt aufteilt. Unter diesen Voraussetzungen könnte man die KI dazu verwenden, weitere potenzielle Teilnehmende vorzuschlagen, wenn sich diese am Markt etablieren, um grundlegende Abhängigkeiten durch eine marktbeherrschende Stellung eines Teilnehmenden zu verhindern, bzw. dieser Situation entgegenzuwirken.

Ein weiteres Beispiel der Verwendung von KI innerhalb der Wertschöpfungskette könnte die Erkennung von ungewöhnlichen Ereignissen in der produktiv laufenden I40-Wertschöpfungskette und eine entsprechende (Früh-) Warnung sein. Beispielsweise könnte eine KI eine ungewöhnlich hohe Datenanfrage mit ggf. ungewöhnlich hoher Ablehnungsrate bei einer Kommunikation zwischen zwei Teilnehmenden erkennen. Oder sie könnte ungewöhnliche Datenanfragen erkennen, deren Zugriff autorisiert wurde, die bisher nicht oder nur selten benötigt wurden.

KI kann ggf. vor dem Start des Produktivbetriebs dazu eingesetzt werden, intelligente Vorschläge zur hinreichenden Bereitstellung von Produktionsdaten für die Teilnehmenden innerhalb der Wertschöpfungskette zu erstellen. Die Frage, welche Teilnehmerin oder welcher Teilnehmer wann welche Daten von welchem anderen Teilnehmenden benötigt und dann einen genehmigten Zugriff haben sollte, ist bei einer größeren Zahl von Teilnehmenden einer Wertschöpfungskette nicht mehr trivial zu beantworten und kann vom Menschen ohne technische Unterstützung kaum noch konsistent eingerichtet werden.

Zusätzlich könnte eine KI Vorschläge zur Diversifikation von Wertschöpfungsketten machen, da es sich ggf. um unabhängige Produktionsschritte handelt, bei denen andere Wertschöpfungskettenteilnehmende nicht eingebunden sein müssen. Durch strikte Trennung (z. B. der Rohmaterialzulieferung von der Herstellung des Endproduktes) kann eine erhöhte Sicherheit erreicht werden, da sich die Abhängigkeiten in der Zulieferkette reduzieren und das Wissen der Teilnehmenden einer Wertschöpfungskette über die Teilnehmenden einer anderen Wertschöpfungskette massiv reduziert werden kann.

5.1.2 Sichere digitale Identitäten

Durch neue dezentrale Identitätsarchitekturen wird die Vergabe und das Management von digitalen Identitäten[3] komplexer. Dieses Konzept kam durch die Vorschläge des World Wide Web Consortium (W3C) in die weltweite Diskussion. Neben vielen privaten, kommerziellen und staatlichen Diskussionsgruppen beschäftigt sich auch die EU mit dem Thema im Rahmen des eIDAS-Ecosystems [20]. Dadurch wurden in den letzten zwei Jahren Lösungen weg von einer einzigen, zentral verwalteten Identität, hin zu einer „Self Sovereign Identity“ (SSI) [21] gesucht. Im globalen Lösungsraum können sich Identitäten durch verschiedene, verwendungsabhängige Identifizierungsmerkmale auszeichnen, ebenso wie auch deren Gültigkeit und Echtheit nur von verwendungsspezifischen Prüfstellen geprüft und bestätigt werden. Im Falle der Verwendung einer SSI verwaltet eine I40-Entität [22] ihre verschiedenen, verwendungsspezifischen Identitätsmerkmale selbst. Diese Verwaltung erfolgt nicht mehr zentral, sondern durch unterschiedliche dezentralisierte Instanzen unter Nutzung sog. „Decentralized Identifiers“ (DIDs). Die Konzepte sind zurzeit bei W3C-Arbeitsgremien in Entwicklung [23]. Es besteht darüber hinaus der Plan, die Interpretation von Identitäten verwendungsspezifisch für Verwendungszwecke im Industrie 4.0-Umfeld und auch bei GAIA-X (s. Kapitel 1) zu definieren.

Die zu fordernden Mindestsicherheitsattribute einer Identität können in Abhängigkeit von deren Verwendung, des Umfeldes und unter der Berücksichtigung von entsprechenden Kosten für deren Bereitstellung variieren. So werden sicherlich im Bereich kritischer Infrastrukturen höhere Sicherheitsanforderungen an die I40-Identitäten gestellt als in der Massenherstellung von Papieratemschutzmasken. Die Massenherstellung von Papieratemschutzmasken für einen Grundschutz gegen ansteckende Krankheiten bedarf zwar einer sehr guten Qualitätskontrolle, jedoch ist es aus

3 Eine Beschreibung der Eigenschaften einer sicheren Identität befindet sich im Ergebnispapier „Technischer Überblick: Sichere Identitäten“ [22]. Diese Lösung trägt ein gewisses Potenzial zur KI-Assistenz in sich und wird weiter unter erläutert.

Kostengründen möglicherweise schwierig, jeder einzelnen Schutzmaske einen eigenen, unfälschbaren Identitätsnachweis über die gesamte Wertschöpfungskette abzuverlangen.

KI kann bei der Aufstellung von Mindestsicherheitsanforderungen an Identitäten einen wichtigen Beitrag leisten. Es kann sich bei den Anforderungen durchaus um eine Liste mit abgestuften Sicherheitsanforderungen handeln, auch wenn die Identitäten sich in einer gemeinsamen Wertschöpfungskette wiederfinden. In die Erstellung der Anforderungen können verschiedenste Kriterien mit einfließen, wie Gesamtkosten der Herstellung, Kostenanforderungen an die Herstellung, Typ der Identität (Mensch, Maschine, Teilprodukt, Endprodukt), Einsatzmöglichkeiten des Endproduktes, Abnehmer des Endproduktes, Entfernung der Identität (z. B. in der Lieferkette) vom eigentlichen Endprodukt, Einfluss von Einzelteilen auf die Funktionsweise, Isolationsmöglichkeit einer Identität innerhalb einer Wertschöpfungskette, Herstellungsgeheimnisse und deren Schutz (IP-Schutz) usw.

Die KI kann dann eine Bewertung einzelner Sicherheitsattribute der in der Wertschöpfungskette verwendeten Identität pro Identität durchführen, z. B. bezüglich:

- Art und Weise der Absicherung der Identitäten hinsichtlich deren Unverfälschbarkeit (per Hardware, Software, Kombinationen, Personalausweis, Fingerabdruck und weiteren biometrischen Merkmalen)
- Anforderungen an die Identität als Kommunikationsteilnehmende (Attribute sicherer Kommunikation, wie Verschlüsselung, Level der Verschlüsselung, Verwendung von Kommunikationsprotokollen, Signaturen, die mindestens benötigt werden)
- Anforderungen an die Auditierbarkeit und Nachvollziehbarkeit (Feststellung der Notwendigkeit der Auditierbarkeit sowie Nachvollziehbarkeit und ggf. deren Detaillierungsgrad)

KI kann außerdem Beiträge leisten, den (erfolgreichen oder auch nicht erfolgreichen) Versuch der Fälschung oder des Austausches einer Sicheren Identität zu erkennen, z. B. durch Verhaltensmustererkennung und Abweichungserkennung von ebendiesem gelernten Verhaltensmuster. Beispiele für abweichende Verhaltensmuster sind:

- Veränderung des Kommunikationsverhaltens, wie Häufigkeit, ungewöhnliche Anfragen, Kontaktaufnahme zu anderen, bisher nicht kontaktierten Identitäten innerhalb der Wertschöpfungskette
- Verwendung anderer Sicherheitsattribute bei der Kommunikation (oder beim Kommunikationsversuch)[4]
- Veränderung der Geschwindigkeit der Kommunikation (z. B. Reaktion auf Anfragen, Zahl von Antwortblöcken)
- Veränderung von IP-Adressen, URLs, Mac-Adressen etc.
- Veränderung der Art der Auditierbarkeits-Daten (z. B. Log-Daten, deren Menge, die protokollierten Ereignisse etc.)
- Veränderung (oder Versuch der Veränderung) von Anmeldeprozessen an anderen Identitäten oder zentralen Komponenten

Da die Sichere Identität im Produktivbetrieb einer Wertschöpfungskette einem Teilnehmenden einer Wertschöpfungskette zugeordnet werden kann, gelten ebenfalls die bereits oben genannten KI-Beiträge zur Absicherung der Wertschöpfungskette.

5.1.3 Sichere Kommunikation

Sichere Kommunikation zeichnet sich vor allem darin aus, dass die Kommunikationsteilnehmenden mit den entsprechenden Sicherheitsattributen ausgestattet sind, wie entsprechende Kommunikationsprotokolle, die Sicherheit gewährleisten, entsprechende Schlüssel zur Verschlüsselung bzw. Entschlüsselung von Daten, vorlegbare Zertifikate zur Prüfung der Echtheit eines Kommunikationsteilnehmenden usw. Sie kann jedoch noch erheblich sicherer gestaltet werden, wenn bei der Einrichtung von Kommunikationswegen nur die Verbindungen entsprechend konfiguriert sind, die auch tatsächlich im Kommunikationsverbund der Wertschöpfungskette benötigt werden. Durch ständige

4 Eine Beschreibung der Eigenschaften einer Sicheren Kommunikation befindet sich im Diskussionspapier „Sichere Kommunikation für Industrie 4.0“ [26] als auch im Diskussionspapier „Sichere unternehmensübergreifende Kommunikation mit OPC UA“ [27].

Änderungen der Teilnehmenden und der damit verbundenen Kommunikationswege ist es leicht möglich, dass sich schnell eine für den Menschen unübersichtliche Kommunikationskonfiguration ergibt, die nur noch mit maschineller – ggf. mit KI-Unterstützung – gemanagt werden kann.

KI kann hilfreich bei der Erkennung der Durchführung ungewöhnlicher Kommunikationsversuche sein. Es gelten hier die gleichen Beispiele wie für die Wertschöpfungskette und die Sichere Identität, die im Zusammenhang mit der Veränderung der Kommunikation und der Kommunikationsattribute für eine Sichere Kommunikation verwendet werden: der Geschwindigkeit der Kommunikation, der Identität etc. KI kann zudem Beiträge dazu leisten, zu erkennen, ob Kommunikationswege fehlkonfiguriert oder unnötig konfiguriert sind. Kommunikationsteilnehmende, die aus der Wertschöpfungskette ausgeschieden sind, kommunizieren möglicherweise immer noch (erfolglos) mit ihren ehemaligen Kommunikationsteilnehmenden, die der Wertschöpfungskette angehören, fragen ggf. nach unnötigen Daten über eine unnötige Verbindung. Genauso ungewöhnlich erscheint es, wenn neue Teilnehmende in der Wertschöpfungskette gar nicht oder nur „wenig“ kommunizieren. In allen Fällen findet eine ungewöhnliche Kommunikation statt, die mit entsprechenden Meta-Daten möglicherweise auch ohne KI erkennbar[5] wäre, doch bei fehlenden Meta-Daten nur mithilfe einer KI erkennbar ist, da auf normalem Wege eine Plausibilität oder Nicht-Plausibilität nicht erkennbar ist.

5.1.4 Vertrauensinfrastruktur/Vertrauensprofile

Zurzeit ist das Thema, wie eine weltweite Vertrauensinfrastruktur aussehen könnte und geschaffen werden könnte, noch in intensiver Diskussion in den entsprechenden Arbeitsgruppen der Plattform Industrie 4.0, so dass in diesem Absatz noch keine konkreten KI-Hilfen zur Absicherung einzelner Entitäten, die die Vertrauensinfrastruktur am Ende definieren, genannt werden können.

Grundsätzlich ist die Vertrauensinfrastruktur jedoch Voraussetzung zum sicheren Betrieb von weltweit agierenden Wertschöpfungsketten und sie ist erforderlich für die Bereitstellung der notwendigen Sicherheitsmerkmale zum Aufbau von sicheren Identitäten und sicherer Kommunikation innerhalb von Wertschöpfungsketten. Daher ist jeder Cyber-Angriff auf eine Wertschöpfungskette oder der Versuch der Manipulation einer sicheren digitalen Identität auch ein Angriff auf die Vertrauensinfrastruktur, in der sich diese befinden. Bei der Diskussion, wie KI bei Sicherheitsaspekten hilfreich sein kann, gelten folglich auch die Beispiele aus den obigen Absätzen zur Wertschöpfungskette und zur Sicheren Identität.

Umgekehrt sind Angriffe auf die Komponenten einer weltweit existierenden Vertrauensinfrastruktur (z. B. auf deren Public Key Infrastrukturen (PKI) oder deren Certificate Authorities bzw. Certification Authorities (CA)) auch Angriffe auf die Sicheren Identitäten, sichere Kommunikation und den sicheren Betrieb von Wertschöpfungsketten. KI kann hier Beiträge dazu leisten, beispielsweise ungewöhnliche Verhaltensweisen bei der Kommunikation von PKIs, CAs und deren Netzwerk-Komponenten zu erkennen und daraus direkte oder indirekte Hinweise auf eine Manipulation solcher Einrichtungen zu liefern.

Insbesondere bei der Herstellung eines Erstvertrauens einer Identität (d. h. Ausstattung der Identität mit erforderlichen Schlüsseln und Zertifikaten) kann die KI möglicherweise ungewöhnliche Aktivitäten erkennen, wie z. B.:

- Verwendung ungewöhnlicher Kommunikationswege bei der Erst- oder Folge-Anfrage von Schlüsseln und Zertifikaten bei einer CA
- Ungewöhnliche Zusammensetzung der Zertifikate
- Ungewöhnliches Verhalten von Kommunikationsteilnehmenden (ggf. Sicheren Identitäten) direkt nach Erhalt und Erstverwendung ihrer neuen Sicherheitsmerkmale

Hinweis: Die Erkennung solcher Manipulationen wird mithilfe zusätzlicher Meta-Daten einfacher. Wenn z. B. bekannt ist, dass eine Sichere Identität in Wertschöpfungskette 1 und 2 involviert ist, kann (teilweise auch ohne KI) erkannt werden, dass ein ungewöhnliches Kommunikationsverhalten vorliegt, wenn die Identität versucht, sich mit Komponenten aus Wertschöpfungskette 3 zu verbinden.

5 Eine grobe Beschreibung der Eigenschaften einer Vertrauensinfrastruktur befindet sich im Dokument „Eberbacher Gespräch zu ‚Sicherheit in der Industrie 4.0‘“ (2013, Fraunhofer SIT, Darmstadt [28]). Eine aktuelle Publikation der Plattform Industrie 4.0 „Vertrauensinfrastrukturen im Kontext von Industrie 4.0“ ist unter Referenz [29] verzeichnet.

5.1.5 Attributbasierte Zugriffssteuerung

Durch die attributbasierte Zugriffsteuerung[6] (Attribute Based Access Control/ABAC) ist man in der Lage, komplexe Einzelregeln als auch Regelwerke zu entwerfen, die – jeweils einzeln untersucht – Sinn machen, bei denen es aber für den Menschen schwierig ist, das sinnhafte Verhalten des Gesamtregelwerkes zu beurteilen und ggf. Inkonsistenzen zu finden. Beispielsweise könnte eine Regel existieren, die den Zugriff in einer Zeit zwischen 10:00 und 16:00 Uhr erlaubt, und später könnte eine zweite Regel eingeführt werden, die den Zugriff zwischen 15:00 und 22:00 Uhr verbietet und damit eine Inkonsistenz für den Zeitraum 15:00 bis 16:00 Uhr erzeugt. Eine Testphase vor der Inbetriebnahme, die solche Arten von Inkonsistenzen erkennbar macht, ist auf jeden Fall erforderlich.

Zusätzlich können komplexe Regeln, die verschiedenste Attribute beinhalten, ggf. gar nicht mehr mit normalen Regelbeschreibungssprachen formuliert werden, da die exakte Formulierung zu viele ausprogrammierte Sonderfälle als Kombinationen von Attributwerten beinhaltet. Beispielsweise könnten in eine oder mehrere Regeln, die bestimmen sollen, ob eine Industrie 4.0-Komponente vom Wartungsstatus in den Produktionsstatus geschaltet werden darf, Attribute wie Komponentendaten (Druck, Temperatur, Stromverbrauch, Motorendrehzahlen etc.), Wetterdaten (Luftdruck, Niederschlag, Art des Niederschlags, Bewölkungsgrad, Luftfeuchtigkeit, Windgeschwindigkeit, Windrichtung, Wetterentwicklung, Gewitterwahrscheinlichkeit abnehmend/zunehmend etc.), Kommunikationsverhalten der Komponente („gewöhnlich“/„ungewöhnlich“), Zustand der Wertschöpfungskette (Abhängigkeiten innerhalb der Wertschöpfungskette, so dass ein Umschalten die Komponente gefährdet) etc. berücksichtigen. Der Mensch benötigt hier massive Hilfestellung seitens der Technik, um den Überblick zu behalten/gewinnen, falls dies überhaupt möglich ist.

KI kann z. B. sehr hilfreich sein, komplexe Regelwerke zu testen und dabei Inkonsistenzen zu finden. Dies kann z. B. durch Clusteringansätze unterstützt werden. Eine KI könnte die verschiedenen Attribute eines Regelwerkes auf eine intelligente Art und Weise iterativ kombinieren, um möglichst schnell zu den Grenzwerten der Antwortveränderung des Regelwerkes zu gelangen, und dann vom Menschen final beurteilen lassen, ob unter diesen Grenzfällen der Zugriff noch erlaubt sein soll oder nicht oder ob ggf. Inkonsistenzen vorliegen, so dass die Entscheidung falsch ausfällt.

KI kann nicht nur zur Kontrolle von Regelsätzen und deren Konsistenz verwendet werden, sondern kann auch innerhalb von Regeln benutzt werden. Das Beispiel des Umschaltens der Industrie 4.0-Komponente kann durch solche KI-basierten Regelwerke unterstützt werden. KI kann hier u. a. auch zeitliche Verläufe analysieren und daraus Rückschlüsse auf normale/ungewöhnliche bzw. unkritische/kritische zeitliche Abläufe ziehen, die – unabhängig von konsistenten und aktuellen Einzelattributwerten – innerhalb der Regel zur Ablehnung des Zugriffs führen können.

Im Zusammenhang mit der in Kapitel 1 dargestellten Nicht-Erklärbarkeit von KI-Bewertungen könnte in obigem Beispiel die Bewertung des zeitlichen Verlaufs des Zustandes der Wertschöpfungskette (beschrieben durch eine Vielzahl einzelner, aber zusammenwirkender Parameter) durch eine KI für den Menschen unnachvollziehbar sein. Als mögliche Abhilfen kommen hier die ebenfalls oben genannten Hilfsmittel zur (Teil-)Erklärbarkeit zum Einsatz: Intensive Tests, Datenprüfung der Gültigkeit der Lerndaten, Granularisierung der KI-Bewertungen.

5.1.6 Collaborative Condition Monitoring

Das Collaborative Condition Monitoring[7] ist ein geeignetes Szenario zum Aufzeigen des Zusammenwirkens der verschiedenen Aspekte, die bereits in den vorangegangenen Absätzen diskutiert wurden. Innerhalb einer Wertschöpfungskette kommunizieren (mindestens) zwei Sichere Identitäten (beispielsweise der Monitoring-Beauftragte und die I40-Entität) über eine sichere Kommunikation miteinander (gesicherte, verschlüsselte Verbindung), die über eine Vertrauensinfrastruktur sichergestellt wird und bei der der Anfragende über ein Attribute Based Access Control

6 Eine Beschreibung der Eigenschaften der Zugriffssteuerung für Industrie 4.0 findet sich im Diskussionspapier „Zugriffssteuerung für Industrie 4.0-Komponenten zur Anwendung von Herstellern, Betreibern und Integratoren“ [30].

7 Eine Beschreibung des Begriffs „Collaborative Condition Monitoring“ und eine Beschreibung des Szenarios befindet sich in der Publikation „Kollaborative datenbasierte Geschäftsmodelle“ [31].

lesenden Zugriff auf die Monitoring-Daten erhält. Es gelten daher alle Beispiele für KI-Hilfen aus den bisher diskutierten Begriffen.

KI kann im Zusammenhang mit diesem Szenario zusätzlich hilfreich sein, um unberechtigte oder auch ungewöhnliche Zugriffe auf Monitoring-Daten zu erkennen. Häufig kann der Monitoring-Beauftragte alle Informationen einem aggregierten Bericht entnehmen, der es unnötig macht, weiteren Zugriff auf Detail-Attribute der Entität zu erhalten, oder er ist nur in seltenen Fällen notwendig. Das Gleiche gilt für den schreibenden Zugriff auf bestimmte Attribute. Aus einer Häufung solcher Zugriffe über verschiedene I40-Entitäten hinweg kann gefolgert werden, ob der Zugriff oder die Häufung ungewöhnlich erscheint. Werden Monitoring-Daten regelmäßig/dauerhaft bereitgestellt und ggf. gesammelt, ergibt sich zusätzlich eine Notwendigkeit der Überwachung der Datenflüsse dieser Daten. KI kann hier unterstützen, ungewöhnliche Datenflüsse zu erkennen und ggf. zu verhindern oder sie vor dem Versand entsprechend bestimmter Empfänger-Attribute zu filtern. Im Zusammenhang mit der zunehmenden Verwendung von KI in der Edge können im Falle des Szenarios des „Collaborative Condition Monitorings“ Regelverstöße und auffällige Verhaltensweisen bei Anfragen bereits im Edge-Device erkannt und die Anfrage schon an dieser Stelle beendet werden.

5.1.7 Die Verwaltungsschale

Die Verwaltungsschale[8] ist das Kernelement der I40-Entitäten. Die Verwaltungsschalen besitzen auch in Bezug auf Sicherheit zentrale Bedeutung, da sich dort u. a. Betriebsparameter als auch -geheimnisse des Herstellers sowie des Betreibers der Entität befinden. Diese könnten bei einem Cyber-Angriff ausspioniert, zerstört oder manipuliert werden. Da die Verwaltungsschale Teil der I40-Entität ist und die I40-Entität durch eine Sichere Identität identifiziert ist, die sich wiederum innerhalb einer Wertschöpfungskette befindet und mit sicherer Kommunikation mit anderen Wertschöpfungskettenteilnehmenden kommuniziert sowie Zugriffskontrollmechanismen besitzt, gelten die bisherigen Betrachtungen vollumfänglich für die Verwaltungsschale.

KI kann speziell bei der Überwachung der Datenspionage und Datenmanipulation mit Schadensabsicht hilfreich sein. Die Datenspionage lässt sich auf die bereits ausführlich diskutierten ungewöhnlichen Datenzugriffe oder Datenzugriffsversuche abbilden. Die Datenmanipulation mit Schadensabsicht muss unterscheidbar sein von einer normalen und durchaus häufigen Veränderung von Daten der Verwaltungsschale (z. B. im Zusammenhang mit einem Einsatz der I40-Entität unter Verwendung neuer Betriebsparameter). Wenn die Verwaltungsschale sehr viele verschiedenste Attribute enthält, ist es dem Menschen nicht mehr möglich, jegliche Art von Kombinationen der verschiedenen Attributwerte zu überwachen, und eingebaute Software-Eingabe-Prüfungen sind ebenfalls nicht in der Lage, beliebige Kombinationen von Attributen und Attributwerten regelbasiert auf Gültigkeit zu prüfen. KI kann jedoch sehr gut genau solche Plausibilitäts-Checks durchführen und entsprechend alarmieren, falls eine Kombination dem Whitespace-Raum zuzuordnen ist.

Wie beim Begriff „Access Control“ und am Beginn des Kapitels bereits erläutert, besteht bei KI die Schwierigkeit, dass das Ergebnis einer KI-basierten Untersuchung nicht von der gleichen KI selbst erklärt werden kann. Sie liefert lediglich einen Score, der anzeigt: mehr plausibel oder weniger plausibel. Für die Prüfung der Plausibilität der Verwaltungsschalendaten müssen daher Hilfsmittel bereitgestellt werden, die final durch den Menschen analysierbar werden, wenn die KI den Alarm erzeugt. Geeignet erscheint hierbei die Speicherung der Historie der letzten Änderungen der Verwaltungsschalendaten, so dass die Datenversion von vor dem KI-Alarm maschinell mit der Version verglichen werden kann, bei der der Alarm auftrat. Dadurch kann der Mensch (als Administrator der Verwaltungsschale) hinreichend einfach nach entsprechender Recherche die Grundursache des Alarms erkennen. Zusätzlich kann eine weitere KI als Hilfsmittel für die Erklärung der Plausibilität bzw. Nicht-Plausibilität herangezogen werden. Der Mensch sollte hierbei jedoch die letzte Entscheidung haben, ob final die alarmierte Version der Verwaltungsschalendaten noch plausibel ist oder nicht.

8 Eine Beschreibung der Verwaltungsschale befindet sich in der Spezifikation „Details of the Asset Administration Shell“ der Plattform Industrie 4.0 [32]. Eine Beschreibung von Security-Anforderungen an die Verwaltungsschale befindet sich im Diskussionspapier „Security der Verwaltungsschale“ [33].

Da sich zum Zeitpunkt der Verfassung dieses Dokumentes verschiedene Aspekte der Verwaltungsschale noch in Diskussion innerhalb der Plattform Industrie 4.0 befinden, kann in diesem Dokument noch nicht detaillierter auf diese Aspekte eingegangen werden.

5.1.8 GAIA-X/Cloud Services/Edge-Devices

GAIA-X[9] steht im Prinzip für einen hochgradig einheitlichen (standardisierten) Datenverkehr im Internet zwischen verschiedensten Cloud-Services, die alle miteinander interagieren und somit komplexe Szenarien abbilden, die über verschiedenste Cloud-Service-Provider und Cloud-Infrastruktur-Anbieter verteilt sind. GAIA-X verkörpert als erstes Gebot: „Security und Privacy by Design". Dieses Paradigma ist ein entscheidender Erfolgsfaktor für GAIA-X, da die Cyber-Angriffsfläche einer solchen Dienstleistungsinfrastruktur noch erheblich höher erscheint als in reinen Industrie 4.0-Szenarien. Es ergibt sich zwangsläufig die Frage, wie bei sehr verteilten Verantwortlichkeiten in verschiedensten Umgebungen (Hyperscaler, Edge Devices, Firmennetzwerke ...) und bezüglich verschiedener Security-relevanter Ebenen (Netzwerk, Applikationen, Betriebssysteme, Datenbanken, GAIA-X-Verwaltungsinfrastruktur ...) Security als Ganzes betrachtet werden kann. Es müssen daher sehr gute Schutzmaßnahmen etabliert sein, die auf allen Ebenen wirken (Prävention, Detektion, schnelle Reaktion etc.), um die gesamte Infrastruktur abzusichern. Eine solche Service-Infrastruktur würde eine europäische Vertrauensinfrastruktur voraussetzen. Da sich irgendwann Kommunikationsteilnehmende innerhalb dieser Infrastruktur befinden und kommunizieren, sind die bisherigen Beispiele für die Sichere Identität, Sichere Kommunikation, Vertrauensinfrastruktur, Access Control (transferiert vom Industrie 4.0-Fall auf den GAIA-X-Fall) ebenfalls vollumfänglich gültig. Da innerhalb der GAIA-X-Service-Infrastruktur auch Industrie 4.0-Szenarien enthalten sind, gelten für diese auch die Betrachtungen zu den anderen Begriffen.

Die Hilfe von KI zur Absicherung dieser Service-Infrastruktur auf allen Ebenen ist sinnvoll. Da sich das GAIA-X-Projekt zum Zeitpunkt der Verfassung dieses Dokumentes erst kürzlich gegründet hat und sich daher noch teilweise in einer Definitionsphase befindet, sind hier nur einige offensichtliche Beispiele der Hilfe von KI aufgeführt.

KI kann eine Überwachung/Prüfung des Datenaustausches zwischen den Microservices, Edge-Devices, generell zwischen den Kommunikationsteilnehmenden, auf Nicht-Plausibilität durchführen, wie z. B.:

- ungewöhnliche Messwerte
- ungewöhnlich viele Anfragen
- ungewöhnliche Verbindungsversuche (erfolgreich/erfolglos)
- ungewöhnliche Abfragen von Daten
- ungewöhnliche Nutzung von APIs
- ungewöhnliche Verbindungsaufnahme des Microservices/Devices zu diversen Empfängern

Da im Falle von GAIA-X eine Orchestrierung der Kommunikationswege schwieriger erscheint als in einem reinen Industrie 4.0-Szenario mit einer sich zwar ändernden, aber definierten Anzahl von Teilnehmenden an einer Wertschöpfungskette, kann die Kommunikation mit einem festen, nicht KI-basierten Regelwerk kaum geprüft werden. Außerdem kann KI feststellen, wenn sich das Kommunikationsverhalten von Teilnehmenden verändert. Das erlaubt Rückschlüsse, ob diese Teilnehmenden manipuliert wurden. Verändertes Kommunikationsverhalten einzelner Teilnehmender kann sich u. a. auszeichnen durch:

- Versand abweichender Daten
- Benutzung veränderter Anmelderoutinen
- veränderte Kommunikationshäufigkeit
- veränderte Empfänger der Nachrichten des Teilnehmenden
- Änderung der Attribute der Sicheren Identität

9 Eine Beschreibung zu GAIA-X befindet sich auf der Website von GAIA-X [34].

6 Zusammenfassung und Fazit

πάντα ῥεῖ – Alles fließt

Kaum sind die aktuellen KI-Forschungsergebnisse Bestandteile von Produktentwicklungen geworden in einer sonst eher langsam getakteten Welt der Industrieinnovationen, zeichnen sich bereits neue Trends im KI-Bereich ab. Gewaltige Investitionen von Billion-Dollar-Companies üben weltweit eine starke Sogwirkung aus und brechen in der Folge auch in der industriellen Welt aufgrund dieser Dynamik weiterhin unentwegt etablierte Produktionsstrategien auf.

Dieses Dokument richtet sich an industrielle Anwender von KI-Systemen ebenso wie an Entwickler von KI-Algorithmen, um zu verdeutlichen, welche Auswirkungen der rapide Wandlungsprozess der Künstlichen Intelligenz auf die Gesamtheit der industriellen Produktion weiterhin hat. Da aktuell keine Abflachung dieses schnelllebigen Trends erkennbar ist, ist es unumgänglich, die unterschiedlichen Auswirkungen auf die Industrie 4.0-Produktion zu untersuchen und darzustellen. Hierbei geht es zum einen um die hohen Leistungspotenziale, die mit diesen KI-Entwicklungen einhergehen. Zum anderen geht es auch um mögliche Trade-off-Situationen durch neue Risiken aus Verwerfungen (Bias) oder unzureichenden bzw. kompromittierten Data-Sets, deren sich jeder Entwickler bewusstwerden muss. Anwender „vertrauen“ KI-assistierten Systemen zunehmend komplexe Entscheidungen an, ohne noch über weitere Rückversicherungen nachdenken zu können. Allein dieser Umstand nimmt Produktverantwortliche und Entwickler in die Pflicht, jetzt wieder „den in Vergessenheit geratenen systemischen Grundsätzen“ [24] bei Konstruktion und Implementierung von KI-basierten Systemen zu folgen.

Optimierungen, wohin das Auge schaut

Selbstverständlich sind auch im Bereich der Künstlichen Intelligenz permanente technische Anpassungen und Optimierungen ein ständiges Entwicklungsziel. So wurden noch vor drei Jahren die fast unvorstellbaren Leistungen, die Computer mit Machine Learning erzeugen können, in der Vorstellung vieler Außenstehender als Ergebnisse aus Megarechenzentren angesehen. Neuronale Netze, also die fast ausschließliche Erzeugungstechnologie solcher Leistungen, werden zwar in Mega-Data-Centers mit weiterhin stark wachsendem Ressourceneinsatz trainiert, aber die Inferenz, also die Anwendung des fertigen Netzes, inklusive des KI-Systems, findet inzwischen lokal, an der Edge des Internets in dedizierten Geräten statt. Optimierungsziele fordern somit z. B. keine kostbare Antwortzeit des Benutzers als Latenzzeit zu verschwenden, indem möglichst optimierte kompakte Geräte in der Edge zum Einsatz kommen. Die komplexesten Neuronalen Netze, die immer neue Weltrekorde in Bilderkennung und ähnlichen Aufgaben aufstellten, laufen aktuell auf Kleinstrechnern mit einer Stellfläche im Scheckkartenformat zu Bagatellkosten und Antwortzeiten ohne Netzlatenz mit nahezu beliebig skalierbarer Verfügbarkeit. Der Optimierungsgrad erreicht neue Höhenflüge.

Auswirkungen auf die Industrie 4.0

Was bedeutet das für Anwender und Hersteller aus dem Bereich der Industrie 4.0? Die bisherige Industrie 4.0-Konzeption der sicheren unternehmensübergreifenden Kommunikation erhält z. B. im Bereich der Maschine-zu-Maschine (M2M)-Kommunikation durch KI-Edge-Komponenten zusätzliche Möglichkeiten der Endpunkt-Security. Bestimmte Anwendungsbereiche können vereinfacht und kostengünstig durch KI abgesichert werden.

Beispielsweise brauchen Dienstleister zur prädiktiven Wartung keinen direkten Zugriff auf originäre Daten und lokale Sensordaten. Die jeweilige Maschine kann mittels KI-basierter prädiktiver Wartungsfeatures alle relevanten Aussagen durch eigene Messreihen und Analysen vor Ort in der Edge ermitteln und bereits als Handlungsanweisungen aktiv an den Dienstleister übertragen. Geheimhaltungsrisiken über rohe Messdaten entfallen. Befürchtungen der Betreiber, dass Wartungsdienstleister Aktivitäten der Maschine aus sensorischer Beobachtung schließen könnten, werden auf diese Weise weitgehend gegenstandslos.

Gleichzeitig bedeutet die Verschiebung der Intelligenz von der Cloud in die Edge aber auch die Entstehung neuer Securityrisiken, die durch hohe Rechenleistungen in der Edge entstehen. Anschaulich dargestellt können vier, acht oder noch mehr Kerne der CPU eines Smartphones allenfalls dann ausgelastet werden, wenn Video-Streaming oder andere Online-Live-Services zum Einsatz kommen. Die KI-Leistungen dieser Edge-Geräte sind beachtlich und erlauben neben den vom Nutzer gewünschten Video- und Stimmanalysen zu komfortablen Identifikationszwecken weitreichende unbemerkte und ungewollte Spionage- und Analysemöglichkeiten einschließlich lokaler Zeitreihenanalysen, ohne dabei die Prozessorlast merklich zu erhöhen. In dieser Hinsicht befinden wir uns bei dieser Technologie seit längerem in einem sicherheitskritischen Zustand. Aktuell wird das Vertrauen in diese Technologie durch das Ökosystem des Herstellers und dessen Vertrauensversprechen suggeriert. IT-Security spielt hier keine herausragende Rolle. Hier gibt es bislang kein belastbares Know-Your-Vendor-Prinzip (KYV), in Anlehnung an das im Finanzbereich bekannte Know-Your-Customer (KYC)-Konzept.

Durchaus anders ist diese Situation im Industriekontext zu bewerten: IT-Security richtet sich insbesondere im Industrie 4.0-Bereich gegen Bedrohungen mit dem Ziel, Erpressungen aufgrund von Vorfällen durch Cybercrime, Sabotage durch Staaten und Dienste oder Spionage, wie z. B. Datendiebstahl (illegale Kopien) aufgrund von Wirtschaftsspionage, zu verhindern. Weitergehende Ziele von Angriffen sind sekundäre, indirekt erzeugte wirtschaftliche Effekte aus der Erlangung von Transparenz über Ziele, Politiken und andere geschäftsrelevante Erkenntnisse über Wettbewerber. KI-basierte Angriffe sind in diesem Kontext nach heutigem Stand der IT-Security nur dem Prinzip nach identifizierbar. Um hier besser detektieren zu können, sind zusätzliche Sensoren im Netzwerk erforderlich, um „ungewöhnliche“ Verhaltensweisen von Edge-Komponenten zu identifizieren. Die Trainingsphasen derartiger Systeme können zwischen sechs bis zwölf Monaten liegen, um die False-Positive-Rate (FPR) auf ein erträgliches Maß an Falschmeldungen zu reduzieren. Bei Überkompensation dieses Verhaltens der Detektoren kann die gefährliche False-Negative-Rate (FNR) ansteigen, was dann zur Nicht-Detektion von Angriffen führt.

Andere Lösungen, wie „Security-Gateways“[10] zwischen Netzsegmenten, erkennen diese Art der Angriffe nicht sicher, weil diese Netzübergänge (Conduits) üblicherweise nicht in der Lage sind, Verschlüsselungen auf semantischer Ebene zu erkennen. Menschlich verständliche Botschaften werden auf KI-Ebene nicht erkannt und blockiert. Im KI-Zeitalter können Konzepte aus der Vergangenheit, die Protokolle auf Verschlüsselung prüften, nur noch bedingt gegen KI-basierte (GAN, adversarial examples, [5] [6]) synthetische Schlüssel schützen, die in Bildern, Geräuschen oder anderen als harmlos designten Berichtsmerkmalen verborgen sind. In Extremfällen tauschen intensiv vortrainierte KI-Angreifer aus der Edge überhaupt keine Primärdaten mehr aus, sondern übertragen vollkommen harmlose Nachrichten, deren subversive Bedeutung von dem jeweiligen KI-Control Center oder anderen Edge KI-Systemen im gleichen Netzwerk dennoch zu hundert Prozent richtig erraten wird, allerdings von der überwachenden KI-Instanz nicht mehr verstanden wird. Dieser Zustand entspricht der kompletten Übernahme des potenziellen Opfers durch den Angreifer.

10 Eine international genormte Beschreibung des Begriffs ist bisher nicht verfügbar.

Empfehlungen und Ausblick

Wie die obigen Ausführungen gezeigt haben, handelt es sich bei KI-Anwendungen in der Edge um eine zukunftsweisende Technologie, die einen souveränen Umgang erfordert, damit sie ihren Nutzen bei vertretbarem Risiko entfalten kann.

Um diese Situationen allgemein besser erfassen bzw. beeinflussen zu können, sind gute bis sehr gute Kenntnisse der eigenen Prozesse, Produktionssysteme und Infrastrukturen erforderlich. Auf das aus der Finanzwelt bekannte KYC-Prinzip sollte hier ein Know-Your-Artificial-Intelligence (KYAI)-Prinzip in der Industrie 4.0 folgen. Externe Hilfe zur Evaluierung der eigenen Prüfergebnisse wird in jedem Fall empfohlen.

KI-Systemen in der Edge kommt unter Security-Aspekten eine besondere Bedeutung zu. Erkannte Schwachstellen sind ernst zu nehmen, denn KI-Systeme sind nicht selbstkritisch und leisten keinen Eigenschutz vor dem eigenen „blinden Fleck“, wie z. B. einem Bias. Beim Einsatz von Überwachungssystemen für KI-Edge-Anwendungen ist zu klären, wann und unter welchen Bedingungen Eskalationen bzw. wann aufgrund der Gefahrenschemata Deeskalationen ausgelöst werden müssen, z. B., um über einen bestimmten Zeitraum eine hohe FPR zu kompensieren. Unbekannte KI-Systeme, -Algorithmen und unbekannte Data-Sets tragen grundsätzlich ein Sicherheitsrisiko in sich. Hersteller von KI-basierten Produkten brauchen neue Kriterien zur Schaffung und zum Erhalt der Vertrauenswürdigkeit von KI-basierten Produkten, basierend auf einem KYAI-Gedanken, der an dieser Stelle ausdrücklich empfohlen wird.

Literaturverzeichnis

[1] J. Benoit, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam und D. Kalenichenko, „Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference“, 1712.05877.

[2] Google for Games, „Flatbuffers“, [Online]. Available: https://google.github.io/flatbuffers/. [Zugriff am 21.03.2021].

[3] Zentralverband der Elektroindustrie (ZVEI), „AI to Industrial Automation White Paper“, Draft 2.2021. [Online].

[4] C. Molnar, „Interpretable Machine Learning A Guide for Making Black Box Models Explainable“, 08.03.2021. [Online]. Available: https://christophm.github.io/interpretable-ml-book/index.html. [Zugriff am 11.03.2021].

[5] Plattform Industrie 4.0 AG Security UAG KI, „Umgang mit Sicherheitsrisiken industrieller Anwendungen durch mangelnde Erklärbarkeit von KI-Ergebnissen“, 28.10.2019. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Umgang-mit-Sicherheitsrisiken.pdf?__blob=publicationFile&v=12. [Zugriff am 16.02.2021].

[6] Plattform Industrie 4.0 AG Security UAG KI, „Künstliche Intelligenz in Sicherheitsaspekten der Industrie 4.0“, 01.04.2019. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/KI-in-sicherheitsaspekten.pdf?__blob=publicationFile&v=8. [Zugriff am 16.02.2021].

[7] S. Waldstein, „The Lancet Digital Health“, 1.6.2020. [Online]. Available: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30080-7/fulltext. [Zugriff am 18.03.2021].

[8] „Tiny machine learning“, [Online]. Available: https://www.tinyml.org/. [Zugriff am 01.07.2021].

[9] Q. V. L. Mingxing Tan, „EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks“, [Online]. Available: https://arxiv.org/abs/1905.11946v5. [Zugriff am 21.03.2021].

[10] P. I. 4.0, „IT-Security in der Industrie 4.0 – Handlungsfelder für Betreiber“, [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/leitfaden-it-security-i40.html. [Zugriff am 26.03.2021].

[11] „Coral“, [Online]. Available: https://coral.ai/products/#production-products. [Zugriff am 01.07.2021].

[12] P. D. H. Pohl, „Der Patch ist der Angriff – Der Patch ist der Angriff“, 21.04.2021. [Online]. Available: https://www.it-daily.net/it-sicherheit/cybercrime/26735-der-patch-ist-der-angriff?start=1. [Zugriff am 11.03.2021].

[13] P. H. O'Neill, „How China's attack on Microsoft escalated into a "reckless" hacking spree“, 10.03.2021. [Online]. Available: https://www.technologyreview.com/2021/03/10/1020596/how-chinas-attack-on-microsoft-escalated-into-a-reckless-hacking-spree/. [Zugriff am 11.03.2021].

[14] Sophos, „HAFNIUM: Advice about the new nation-state attack“, 05.03.2021. [Online]. Available: https://news.sophos.com/en-us/2021/03/05/hafnium-advice-about-the-new-nation-state-attack/. [Zugriff am 11.03.2021].

[15] Bundesamt für Sicherheit in der Informationstechnik, BSI, „BSI warnt: Kritische Schwachstellen in Exchange-Servern“, 05.03.2021. [Online]. Available: https://www.bsi.bund.de/DE/Service-Navi/Presse/Pressemitteilungen/Presse2021/210305_Exchange-Schwachstelle.html. [Zugriff am 11.03.2021].

[16] W. Badr, „Top Sources For Machine Learning Datasets“, 13.01.2019. [Online]. Available: https://towardsdatascience.com/top-sources-for-machine-learning-datasets-bb6d0dc3378b. [Zugriff am 11.03.2021].

[17] M. Khairy, „TPU vs GPU vs Cerebras vs Graphcore: A Fair Comparison between ML Hardware“, 2020.

[18] N. P. J. et al, „Ten Lessons From Three Generations Shaped Google’s TPUv4i“, ACM/IEEE 48th Annual International Symposium on Computer Architecture, 2021.

[19] Plattform Industrie 4.0, Dr. Jürgen Neises, George Moldovan, Thomas Walloschke, „Trustworthiness as facilitator of Policy and Access Management in Supply Chains,“ 03 02 2021. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/2021_IT-Security-Conference_PPT_Neises.html. [Zugriff am 02.07.2021].

[20] EIDAS, „ec.europa.eu“, 1.5.2019. [Online]. Available: https://ec.europa.eu/futurium/en/system/files/ged/eidas_supported_ssi_may_2019_0.pdf. [Zugriff am 18.03.2021].

[21] „Wikipedia,“ 8.3.2021. [Online]. Available: https://de.wikipedia.org/wiki/Selbstbestimmte_Identit%C3%A4t. [Zugriff am 18.03.2021].

[22] Plattform Industrie 4.0, „Technischer Überblick: Sichere Identitäten“, 2016. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/sichere-identitaeten.html. [Zugriff am 11.03.2021].

[23] D. Reed, M. Sporny und D. Longley, „W3C“, 9.3.2021. [Online]. Available: https://www.w3.org/TR/did-core/. [Zugriff am 18.03.2021].

[24] G. Ropohl, „Allgemeine Systemtheorie, Einführung in transdisziplinäres Denken“, Berlin: edition sigma, 2012.

[25] Bundesministerium für Wirtschaft und Energie, BMWi, „IT-Sicherheit für Industrie 4.0“, 04.01.2016. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/bmwi-studie-it-sicherheit.pdf?__blob=publicationFile&v=5). [Zugriff am 11.03.2021].

[26] Plattform Industrie 4.0, „Sichere Kommunikation für Industrie 4.0“, 2017. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/sichere-kommunikation-i40.pdf?__blob=publicationFile&v=5. [Zugriff am 11.03.2021].

[27] Plattform Industrie 4.0, „Sichere unternehmensübergreifende Kommunikation mit OPC UA“, 2019. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/sichere-kommunikation-opc-ua.pdf?__blob=publicationFile&v=13. [Zugriff am 11.03.2021].

[28] Fraunhofer SIT, Darmstadt, „Eberbacher Gespräch zu ‚Sicherheit in der Industrie 4.0‘“, 10.2013. [Online]. Available: https://www.sit.fraunhofer.de/fileadmin/dokumente/studien_und_technical_reports/Eberbach-Industrie4.0_FraunhoferSIT.pdf?_=1420719894. [Zugriff am 11.03.2021].

[29] Plattform Industrie 4.0, „Vertrauensinfrastrukturen im Kontext von Industrie 4.0“, 2021. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Vertrauensinfrastrukturen.html. [Zugriff am 09.08.2021].

[30] Plattform Industrie 4.0, „Zugriffssteuerung für Industrie 4.0-Komponenten zur Anwendung von Herstellern, Betreibern und Integratoren“, 11.2018. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/zugriffssteuerung-industrie40-komponenten.pdf?__blob=publicationFile&v=8. [Zugriff am 11.03.2021].

[31] Plattform Industrie 4.0, „Kollaborative datenbasierte Geschäftsmodelle“, 07.2020. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/kollaborative-datenbasierte-geschaeftsmodelle.pdf?__blob=publicationFile&v=5. [Zugriff am 11.03.2021].

[32] Plattform Industrie 4.0, „Details of the Asset Administration Shell“, 2018/2020. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Details_of_the_Asset_Administration_Shell_Part1_V3.pdf?__blob=publicationFile&v=5. [Zugriff am 21.03.2021].

[33] Plattform Industrie 4.0, „Security der Verwaltungsschale“, 2017. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/security-der-verwaltungsschale.pdf?__blob=publicationFile&v=6. [Zugriff am 11.03.2021].

[34] Federal Ministry for Economic Affairs and Energy, „GAIA-X: A Federated Data Infrastructure for Europe“, 06.2020. [Online]. Available: https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html. [Zugriff am 11.03.2021].

AUTOREN

Dr. Bernd Kosch, Industrie-KI GmbH | Björn A. Flubacher, Bundesamt für Sicherheit in der Informationstechnik | Dr. Dipl.-Ing. Detlef Houdeau, Infineon Technologies AG | Dr. Michael Schmitt, SAP SE | Olaf Dressel, Bundesdruckerei GmbH | Peter Rost, secunet Security Networks AG | Thomas Walloschke, secon trust consult | Dr. Thomas Wille, NXP Semiconductors

Diese Publikation ist ein Ergebnis der Unterarbeitsgruppe „KI für I40-Security“ der Arbeitsgruppe „Sicherheit vernetzter Systeme“ der Plattform Industrie 4.0.

www.plattform-i40.de