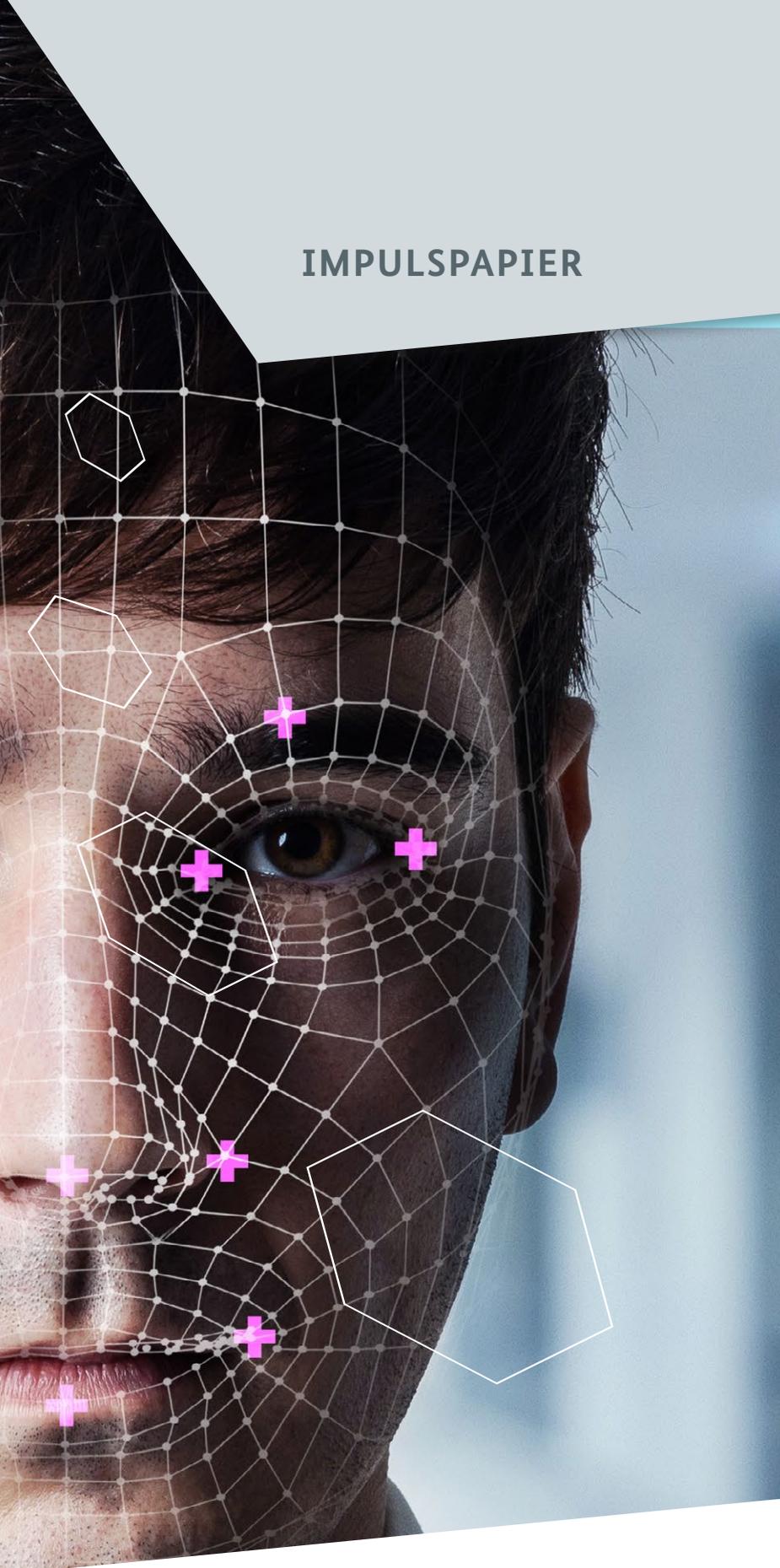


IMPULSPAPIER



- /Administration
- /Human Resources
- /Legal
- /Accounting
- /Finance
- /Marketing
- /Publicity
- /Promotion
- /Research
- /Business
- /Development
- /Engineering
- /Manufacturing
- /Planning



# Zero Trust in der I40

Perspektiven durch den Einsatz von ML

## Impressum

### **Herausgeber**

Bundesministerium für Wirtschaft und Klimaschutz (BMWK)  
Öffentlichkeitsarbeit  
11019 Berlin  
www.bmwk.de

### **Redaktionelle Verantwortung**

Plattform Industrie 4.0  
Bülowstraße 78  
10783 Berlin

### **Stand**

April 2024

Diese Publikation wird ausschließlich als Download angeboten.

### **Gestaltung und Produktion**

PRpetuum GmbH, 80801 München

### **Bildnachweis**

iStock  
gorodenkoff / S. 6  
ismagilov / S. 15  
metamorworks / Titel  
monsitj / S. 2, S. 14  
onurdongel / S. 8

### **Zentraler Bestellservice für Publikationen der Bundesregierung:**

E-Mail: [publikationen@bundesregierung.de](mailto:publikationen@bundesregierung.de)

Telefon: 030 182722721

Bestellfax: 030 18102722721

Diese Publikation wird vom Bundesministerium für Wirtschaft und Klimaschutz 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.



# Inhalt

<b>1. Einführung</b>	<b>1</b>
<b>2. Begriffsbestimmung Zero Trust heute</b>	<b>2</b>
2.1 Frühe und aktuelle Ansätze im Vergleich	4
<b>3. Bezug von Zero-Trust-Architecture zu Industrie 4.0</b>	<b>6</b>
<b>4. Verwendung von künstlicher Intelligenz, um Zero-Trust-Ansätze zu verwirklichen</b>	<b>8</b>
4.1 Stand der Technik: BeyondCorp Enterprise (BCE)	10
4.2 Umsetzung von ZTA mit allgemeinen Large Language Models	11
4.3 Umsetzung von ZTA mit fine-tuned Large Language Models	12
4.4 Einschätzung des Standes der Technik zur Anwendung von LLM zur Umsetzung von ZTA	13
<b>5. Rolle und Möglichkeiten des Einsatzes sowie Grenzen von Machine Learning in Verbindung mit ZTA-Prozessen</b>	<b>14</b>
<b>6. Handlungsempfehlungen</b>	<b>15</b>
<b>7. Zusammenfassung und Ausblick</b>	<b>16</b>
<b>8. Anhang</b>	<b>17</b>
8.1 Einführung eines Zero-Trust-Maturity-Modells	17
8.2 Detailbetrachtung der Beispiele zum Industrie 4.0-Kontext von Zero-Trust-Ansätzen aus Kapitel 5	17
8.2.1 Beispiel der Teilnahme an einer Industrie 4.0-Wertschöpfungskette	18
<b>9. Literaturverzeichnis</b>	<b>21</b>

# 1. Einführung

**Das heute übliche Sicherheitsmodell der industriellen Unternehmens-IT ist konzeptionell veraltet und bedarf deshalb einer grundlegenden Neudefinition**

Die heutigen IT-Landschaften haben sich ausgehend von zentralistisch organisierten Hardware-, Software- und Betriebsprinzipien in den siebziger Jahren durch vollständige Diversifizierung aller Betriebsmittel in unterschiedlichste Größenklassen, Bereitstellungsformen, Zweckbestimmungen, Eigentumsverhältnisse und viele andere Nutzungsaspekte drastisch verändert. Datenhaltung und Rechenleistung werden heute aus global verteilten Cloud-Instanzen von heterogen organisierten Dienstleistern bezogen. Alle Ressourcen werden aus vielfältigsten, individuell organisierten Zugängen angesprochen. Daher ist die Unterscheidung zwischen einem auf Vertrauen basierenden „Inneren“ und einem nicht kontrollierbaren „Äußeren“ immer weniger möglich: Es reift die Erkenntnis: „Trust is a Vulnerability“.

Das vorliegende Papier der Unterarbeitsgruppe „Künstliche Intelligenz für I40-Security“ der Arbeitsgruppe „Sicherheit vernetzter Systeme“ der Plattform Industrie 4.0 stellt mit diesem Diskussionspapier einen aktuellen technischen Status und die wichtigsten Trends zu ZTA und KI sachlich dar.

Das Dokument soll insbesondere den Klein- und mittleren Unternehmen, die z. B. im Bereich der Maschinen- und Anlagen-Entwicklung und -Herstellung tätig sind, einen Überblick geben, welche Prozess-Veränderungen mit der Zero Trust Architecture (ZTA) im Industrial Internet of Things (IIoT) verbunden sein können und wie KI-Algorithmen eine Teilautomatisierung ermöglichen. Neben den neuen Chancen werden auch neue Grenzen und mögliche Risiken der KI-basierten Umsetzung von ZTA-Entscheidungen aufgezeigt.



## 2. Begriffsbestimmung Zero Trust heute

Am 5. November 2010 veröffentlichte der Forrester-Analyst John Kindervag den Aufsatz „Build Security into Your Network’s DNA: The Zero Trust Network Architecture“<sup>[4]</sup>. Das Papier stellte eine konsolidierte Zusammenfassung vorheriger Forrester-Berichte aus 2009 und 2010 dar, die durch Ratschläge und Vorhersagen für Security-Professionals erweitert wurde und heute als Beginn der Diskussion um das Thema „Zero Trust Architecture (ZTA)“ gilt. Kindervag positioniert sein Papier selbst als Bestandteil einer Folge von Berichten, die sich den konzeptionellen Grundkonzepten des neuen Ansatzes widmet, Architecture-Komponenten, Funktionalitäten und erforderliche Technologien skizzieren und mit Fallstudien, Best Practices und Vorteilen darstellen will. Es handelt sich also nicht um eine unmittelbar umsetzbare technische Anleitung für die betriebliche Implementierung oder die Beschreibung eines integrierten Produktes

aus Hardware, Software und konkreten Dienstleistungen. Zero Trust Architecture bezeichnet vielmehr den Übergang zu einem strikten, radikal andersartigen Sicherheitsprinzip: Statt darauf zu vertrauen, dass das Netzwerk in einen gesicherten (inneren) und einen ungesicherten (äußeren) Bereich aufgeteilt werden kann, gilt in einer ZTA, dass der gesamte Netzverkehr als unsicher betrachtet wird: „Thus, security professionals must verify and secure all resources, limit and strictly enforce access control, and inspect and log all network traffic.“

Es existiert keine Beschreibung der Zero Trust Architecture, in der diese als Technologie formuliert wird. Es besteht aber ein allgemeines Verständnis darüber, welche Prinzipien in einer ZTA besonders vorrangig umgesetzt werden müssen.

- **IAM (identity and access management):** Die Identitäten aller beteiligten HW-Komponenten, SW-Komponenten und Personen müssen vor jedem Zugriff auf sensitive Daten geprüft und verifiziert werden.
- **Least privileged access:** Jeder Zugang wird nur für geringstmöglichen Umfang (Scope) und Zeit gewährt, die die jeweilige Transaktion erfordert. Zugriffe auf benachbarte Sachgebiete und/oder in unnötigen Zeitfenstern werden so unterbunden oder zumindest erschwert.
- **Mikrosegmentierung:** Trennung von Netzwerkkomponenten in eine fein granulare Struktur, um dadurch potenzielle Angriffsflächen zu minimieren. Auf diese Weise werden Angriffsrisiken verringert und die Ausbreitung von Schäden eingrenzbar gemacht. Mikrosegmentierung erleichtert die Erkennung von Bedrohungen und ermöglicht automatisierte Reaktionen in Echtzeit.
- **MFA (Multifaktor-Authentifizierung):** ZTA betrachtet Vertrauen als Verletzungsrisiko und ist deshalb bestrebt, dieses durch mehrfache Authentifizierungsnachweise und deren kontinuierliche Prüfung zu verringern. MFA ist ein Kernprinzip der ZTA und senkt das Kompromittierungsrisiko von Identitäten.
- **Aufklärung und Security-Analytik:** Sicherheitsalgorithmen, statistische Analysen und fortgeschrittene Methoden des Machine Learning können Einsichten in das Verhalten von Benutzenden innerhalb des Perimeters liefern. Insbesondere können hochdimensionale Auffälligkeiten erkannt werden, die Zweifel an der Identität oder den Zielen eines Benutzers begründen und in verstärkte Aufmerksamkeit der Überwachung münden können.

In einer zunehmend vernetzten digitalen Welt wird „Zero Trust“ inzwischen weithin als ein wichtiges Konzept zur Erhöhung der Sicherheit von IT-Systemen angesehen.

Im Jahr 2003 wurden die Herausforderungen bei der Festlegung der Grenzen der IT-Systeme eines Unternehmens auf dem Jericho-Forum hervorgehoben, als der Trend zur „Entperimeterisierung“, wie es damals genannt wurde, diskutiert wurde. Der Gedanke des klassischen Perimeterschutzes spielt im Bereich der produzierenden Industrie auch 21 Jahre später noch eine große Rolle, insbesondere vor dem Hintergrund des Spannungsfeldes, das durch unternehmensübergreifende Vernetzung der Daten im Rahmen von Digitalisierung und Industrie 4.0 entsteht. Die notwendige Auflösung des Perimeterschutzes ist daher eine der Triebfedern für die Einführung der Zero Trust Architecture im Umfeld von Industrie 4.0.

Eine der Grundideen des OSSTMM (Open Source Security Testing Methodology Manual)-Rahmenwerks für die Durchführung von Sicherheitstests und -bewertungen von Informationssystemen ist, dass „Vertrauen als Schwachstelle“ betrachtet werden kann, da falsches Vertrauen in ein Informationssystem zu Sicherheitslücken führen kann. OSSTMM definierte 2007 hierzu verschiedene Vertrauensstufen, die auf die Zuverlässigkeit und Sicherheit von Informationssystemen angewendet werden können:

- **Kein Vertrauen (Zero Trust):** In dieser Vertrauensstufe wird dem Informationssystem oder seinen Komponenten kein Vertrauen entgegengebracht. Alle Elemente des Systems, einschließlich Benutzer, Anwendungen und Netzwerke, werden als potenziell unsicher betrachtet und müssen entsprechend streng kontrolliert und überwacht werden.
- **Begrenztes Vertrauen:** Bei diesem Vertrauensniveau wird bestimmten Elementen des Informationssystems begrenztes Vertrauen entgegengebracht, während anderen Elementen nicht vertraut wird. Besondere Kontrollen und Maßnahmen werden ergriffen, um die Teile des Systems zu schützen, denen vertraut wird, während andere Teile streng kontrolliert und überwacht werden.
- **Bedingtes Vertrauen:** Bei diesem Vertrauensniveau wird das Vertrauen auf der Grundlage bestimmter Bedingungen oder Kriterien gewährt. Es werden Kontrollen und Überwachungen durchgeführt, um sicherzustellen, dass diese Bedingungen erfüllt sind, bevor das Vertrauen ausgesprochen wird.

## 2.1 Frühe und aktuelle Ansätze im Vergleich

Der Begriff „Zero Trust“, wie er heute verstanden wird und erst in den letzten Jahren populär geworden ist, basiert im Kern auf älteren Prinzipien und Ideen. Viele ältere Dokumente oder Forschungsarbeiten, die sich mit ähnlichen Konzepten unter teilweise anderen Namen beschäftigen, können daher als Vorläufer des heutigen „Zero Trust“-Ansatzes betrachtet werden.

- **„The Protection of Information in Computer Systems“ von Jerome Saltzer und Michael D. Schroeder (1975):** Dieses Dokument, das oft als „Saltzer- und Schroeder-Papier“ bezeichnet wird, stellt acht Prinzipien für den Schutz von Informationen in Computersystemen vor. Diese Prinzipien betonen die Notwendigkeit, den Zugang zu Informationen zu beschränken und die Vertraulichkeit, Integrität und Verfügbarkeit von Informationen zu gewährleisten, ähnlich wie es im „Zero Trust“-Ansatz betont wird.
- **„BeyondCorp: A New Approach to Enterprise Security“ von Neal Stephenson<sup>[2]</sup>:** Dieses Dokument von Neal Stephenson, einem Mitarbeiter von Google, beschreibt Googles Sicherheitsmodell „BeyondCorp“, das als Vorläufer des „Zero Trust“-Ansatzes gilt und bereits 2009 bei Google implementiert wurde. Es wird die Dringlichkeit betont, dass Unternehmen sich nicht auf vertrauenswürdige interne Netzwerke verlassen sollten, sondern stattdessen den Zugang auf Benutzerebene kontrollieren und kontextbezogene Informationen auf der Grundlage von Gerät, Standort und anderen Faktoren nutzen sollten.

- **„The Forrester Wave: Zero Trust eXtended Ecosystem Platform Providers, Q4 2019“ von Chase Cunningham (2019):** Dieser Bericht von Forrester Research beschreibt das Konzept „Zero Trust“ als neuen Ansatz für IT-Sicherheit und bewertet verschiedene Anbieter von Zero-Trust-Plattformen.

Nachfolgend eine tabellarische Zusammenfassung der Unterschiede zwischen diesen frühen Ansätzen von 1975 bis 2010 und den heutigen Zero-Trust-Ansätzen in Tabelle 1 (S. 5).

Zusammenfassend lässt sich feststellen, dass aktuelle Ansätze des „Zero Trust“-Konzepts eine Weiterentwicklung der ersten Ansätze von 1975 bis 2010 darstellen und darauf abzielen, den Anforderungen einer durch die Digitalisierung zunehmend vernetzten, dezentralen IT- und OT-Landschaft gerecht zu werden. Im Mittelpunkt stehen dabei der einzelne Nutzer, die kontextsensitive Authentifizierung und Autorisierung, die kontinuierliche Überwachung aller kommunizierenden Entitäten (Mensch, Maschine, Software etc.) sowie die dynamische Anpassung von Zugriffsrechten anhand verschiedener Faktoren wie Nutzerrolle, Gerätetyp, Standort oder Zugriffszeit. Dadurch können Zugriffsrechte dynamisch angepasst und Risiken besser eingeschätzt werden.

Im Gegensatz zu früheren Ansätzen, die primär auf Netzwerkebene und internen Vertrauenspositionen basierten, misstrauen aktuelle Ansätze allen Geräten und Nutzenden, unabhängig von Standort oder Netzwerkbereich. Zudem werden Nutzerverhalten und Gerätezustand umfassender überwacht, Anomalien erkannt und Zugriffsrechte kontinuierlich überprüft. Für komplexe IT-Umgebungen und Cloud-Infrastrukturen sind aktuelle Ansätze besser geeignet. Der Fokus liegt auf der Erkennung und Abwehr von Bedrohungen.

Tabelle 1 – Vergleich früher und aktueller Zero-Trust-Ansätze

Aspekt	Frühe Ansätze (vor 2010)	Aktuelle Ansätze (nach 2010)
Fokus	Hauptsächlich auf einzelnen Systemen und Zugriffskontrollen, basierend auf Netzwerkgrenzen und internen Vertrauensstellungen.	Ganzheitlicher Ansatz mit Fokus auf Benutzerebene, kontextbasierte Authentifizierung und Autorisierung sowie kontinuierliche Überwachung und Validierung von Benutzenden und Geräten.
Vertrauensmodell	Vertrauen in interne Netzwerke und Systeme.	Misstrauen gegenüber allen Geräten und Benutzenden, unabhängig von ihrem Standort oder Netzwerkbereich.
Zugriffskontrolle	Hauptsächlich auf Netzwerkebene mit Firewall-Regeln und VPNs, oft auf Basis von IP-Adressen.	Benutzer- und gerätezentrierte Zugriffskontrolle mit starkem Fokus auf Authentifizierung, Autorisierung und Identitätsmanagement.
Überwachung	Begrenzte Überwachung von Netzwerkverkehr und Systemprotokollen.	Umfassende Überwachung von Benutzerverhalten, Gerätezustand, Anomalieerkennung und kontinuierliche Überprüfung von Zugriffsrechten.
Kontextuelle Informationen	Wenig bis keine Verwendung von kontextuellen Informationen wie Benutzerrolle, Gerätetyp, Standort oder Zeitpunkt des Zugriffs.	Starke Verwendung von kontextuellen Informationen zur dynamischen Anpassung von Zugriffsrechten und Risikobewertung basierend auf verschiedenen Faktoren.
Skalierbarkeit	Weniger Skalierbarkeit für komplexe IT-Umgebungen und hybride Cloud-Infrastrukturen.	Höhere Skalierbarkeit für komplexe, dynamische IT-Umgebungen, einschließlich Cloud-Infrastrukturen und mobilen Benutzenden.
Bedrohungserkennung	Wenig bis keine Fokussierung auf Echtzeit-Bedrohungserkennung oder Incident Response.	Aktive Bedrohungserkennung und Incident Response mit kontinuierlicher Überwachung und Analyse von Sicherheitsereignissen.

Quelle: Plattform Industrie 4.0

**Zusätzliche Information:** In der jüngeren Vergangenheit wurden aufbauend auf den Zero-Trust-Prinzipien Reifegradmodelle (engl. Maturity Model) aufgesetzt, die es ermöglichen, Zero-Trust-Ansätze und deren Status in einem Unternehmen oder einer Organisation bezüglich bestimmter Kategorien und Kriterien besser zu beurteilen, verständlicher zu machen und bei deren Umsetzung schrittweise vorzugehen. Eine genauere Information hierzu befindet sich im Anhang (vgl. 8.1).



### 3. Bezug von Zero Trust Architecture zu Industrie 4.0

In der bisherigen Industrie 3.0 hatten Produktionsnetze strikt abgetrennte Perimeter, bei denen ein Datenaustausch mit Entitäten außerhalb des jeweiligen Produktionsnetzes nicht vorgesehen war.

Die neuen Anwendungsfälle der Industrie 4.0 sehen jedoch eine Vernetzung mit einer hohen Dynamik vor: Unternehmen vernetzen sich horizontal (in Bezug auf die Automatisierungspyramide) über die Lieferkette hinweg, sie tauschen sensible Daten aus und verarbeiten externe Daten auf allen Ebenen ihrer Betriebe. Der Shop-Floor von heute ist in vielen Fällen ohne Datenaustausch nicht mehr konkurrenzfähig. Folglich sind die Produktionsnetze also keine strikt abgetrennten Perimeter mehr, sodass sich die mögliche Angriffsfläche für Angreifer vergrößert (vgl.<sup>[3]</sup>).

In der Industrie werden diese Risiken bereits häufig berücksichtigt und zum Teil werden in heutigen kommerziellen Produkten und besonders in konzeptionellen Lösungen schon vereinzelt IT-Securitymaßnahmen eingeplant.

#### Beispiele für solche Anwendungsfälle im Industrie 4.0-Kontext sind:

- Die Teilnahme an einer Industrie 4.0-Wertschöpfungskette: siehe Anhang 8.2.1
- Die Teilnahme an einem Daten-Ökosystem oder Data Space: siehe Anhang

Eine Zero Trust Architecture bedarf allerdings einer gesamtgesellschaftlichen Sicherheitsbetrachtung und des Zusammenspiels aller Securitymaßnahmen. Die meisten der existierenden Teillösungen für Securityanforderungen, wie die Securitykonzepte der Plattform Industrie 4.0, sind geeignet, um in einer ZTA aufgenommen zu werden. Beispielhaft sei in diesem Zusammenhang auf die veröffentlichten Konzepte in den Publikationen „Zugriffssteuerung für Industrie 4.0-Komponenten zur Anwendung von Herstellern, Betreibern und Integratoren“<sup>[4]</sup>, „Sichere Kommunikation für Industrie 4.0“<sup>[5]</sup>, Vertrauensinfrastrukturen im Kontext von Industrie 4.0<sup>[6]</sup>, „Technischer Überblick: Sichere Identitäten“<sup>[7]</sup> verwiesen.

Die Autorisierung für den Zugriff auf eine Industriekomponente beruht heutzutage hauptsächlich auf der Identität des anfragenden Subjekts, die mit einem Zertifikat nachgewiesen wird. Dabei ist zu beachten, dass das Vertrauen in Identitäten nicht statisch ist<sup>[6]</sup>, da das damit repräsentierte Vertrauen beispielsweise durch den Verkauf des Assets an einen nicht vertrauenswürdigen neuen Verwender erlöschen kann. In der Regel sind Zertifikate daher nur über einen definierten Zeitraum gültig und werden nach Ablauf der Gültigkeitsdauer nicht mehr anerkannt.

Die Nebenbedingungen für die Ausstellung der Zertifikate werden heute nach der Ausstellung der Zertifikate in der Regel nicht erneut geprüft. Entsprechend der Dynamik der Identitäten sollte jedoch auch die Einschätzung der Vertrauenswürdigkeit der Nebenbedingungen regelmäßig erfolgen. Dies ist wichtig, da sich die äußeren Umstände wie die geopolitische Lage, die Bedrohungslage, die Betriebssicherheitsanforderungen oder das Betriebsvermögen etc. ändern können.

Vor diesem Hintergrund ist Constant Verification ein Prinzip der ZTA. Dabei werden Sessions möglichst kurz gehalten. D.h. anstatt eine Entität nur einmal zu authentifizieren und ggf. zu autorisieren, sieht ZTA die Authentifizierung und Autorisierung bei jeder Zugriffsanfrage vor.

Mit Attribute Based Access Control (ABAC) können neben dem Identitätsnachweis weitere Prüfkriterien für die Gewährung des Datenzugriffs festgelegt und geprüft werden. Bedingungen können beispielsweise für unterschiedlich vertrauliche Datenkategorien oder aktuelle Randbedingungen etc.<sup>[4], [6]</sup> in die Zugriffsregeln aufgenommen werden. ABAC unterstützt also eine der Grundlagen der ZTA.

Im Sinne von Zero Trust müssen die Prüfungen bei der Datenkommunikation und beim Datenzugriff zukünftig noch häufiger als bisher unter sich ändernden Randbedingungen (z.B. aktuelle geopolitische Veränderungen etc.) erfolgen. Zusammen mit dem aktuellen Fachkräftemangel und dem Axiom, dass ständige Authentisierung, Authentifizierung und Autorisierung den Echtzeitanforderungen nicht zuwiderlaufen dürfen, machen diese Anforderungen einen hohen Automatisierungsgrad bei der Umsetzung der Zugriffssteuerung in der ZTA unabdingbar.

Statische, zum größten Teil manuell aufgestellte Prüfregeln, wie bei ABAC, erfüllen nur begrenzt diese Anforderungen, da die Prüfregeln veralten, neue hinzukommen oder andere wegfallen. Zudem können komplexe Prüfungen gegebenenfalls nicht mehr in ein festes Regelwerk übersetzt werden. Daher können zur Lösung dieser Herausforderungen KI-basierte Prüfungen als zusätzliches Regelwerk geeignet sein.



## 4. Verwendung von künstlicher Intelligenz, um Zero-Trust-Ansätze zu verwirklichen

Nachdem die bahnbrechende Publikation von Hinton et al.<sup>[8]</sup> den Backpropagation-Algorithmus als mathematisch korrekte Methode zur Implementation der seit 30 Jahren aus neurologischer Perspektive ersonnenen Phantasien über neuronale Netze etablierte, war die Business-Zukunft von KI im Sinne von Machine Learning (ML) offenkundig. In den 90er-Jahren entstanden konkrete Verfahren zur Mustererkennung in Bild-, Text- und Sprachdaten, die wirtschaftliche Revolutionen auslösten. Beispiele sind die Automatisierung der Postsortierung, Sprachsteuerung von Geräten oder maschinelle Sprachübersetzung. Mit „Deep Learning“ und globalen Initiativen zur Beschaffung von Trainingsdaten wurden ab 2012 kontinuierlich neue Rekorde in Klassifikation, strategischer Zielerreichung und kreativer Datenerzeugung aufgestellt. Die frühere Vorstellung, dass das menschliche Gehirn mit acht Neuronenschichten eine ideale Konfiguration hätte, wurde immer fragwürdiger, und im Zeitraum bis 2017 sank die Fehlerrate bei Bildklassifikation durch immer tiefere Neuronale Netze von 28 Prozent auf 2,3 Prozent. Schon ab 2015 waren die neuen Rekorde von ResNet, Inception und ähnlichen Netzen der geschätzten menschlichen Rate von 4,6 Prozent deutlich überlegen. Es war ein Zustand erreicht, in dem hoch qualifizierte Systeme sehr viele hoch spezialisierte Aufgaben nicht nur viel schneller, sondern auch qualitativ viel besser bewältigen konnten als menschliche Spezialisten. Neue Firmen fokussierten sich auf Anwendungsfelder, die nur mit „superhuman“ Performance erfolgreich erschlossen werden konnten.

Das Argument gegen die wachsende Überlegenheit maschinellen Lernens schrumpfte zu der Aussage, dass maschinelle Intelligenz nur sehr spezielle Aufgaben besser als Menschen ausführt – aber eben bloß ein einziges Thema perfekt beherrscht. Eine Perspektive zu „Allgemeiner Künstlicher Intelligenz“ (GAI) schien damals noch unvorstellbar und eventuell gefährlich, aber sicherlich noch Jahrzehnte entfernt. Noch während in 2012–2017 die Schlacht der DL Architecture die Bilderkennungsrekorde auf 2,3 Prozent Fehler reduzierte, entstand ab 2014 weitgehend unbeachtet eine völlig neue Idee auf dem Gebiet der Verarbeitung natürlicher Sprache (NLP). Diese Idee wurde perfektioniert und in dem richtungsweisenden Paper „Attention is all you need“<sup>[18]</sup> von Google Brain, Google Research und der Universität Toronto veröffentlicht. Kern dieses Papers ist die revolutionäre Architecture des „Transformers“ auf Basis der „Attention-Mechanismen“.

Transformers wandeln eine Eingabesequenz in eine Ausgabe-sequenz über eine Hidden-Layer-Struktur. Es gibt Möglichkeiten, den Prozess zu verkürzen und encoder- oder decoder-only Strukturen zu erzeugen. Transformers galten anfangs als Weiterentwicklung der Rekurrenten NN aus den 90er-Jahren (LSTM, GRU), bis klar wurde, dass der Attention-Mechanismus nicht nur eine Prozess-Parallelisierung, eine viel größere Erinnerungsreichweite und eine Verallgemeinerung auf diverse völlig andere Anwendungsfelder (hochauflösende Bilder, Videos, Grafiken, etc.) ermöglicht.

Gleichzeitig entstand die Idee für eine neue Form „unsupervised training“ zur Erzeugung sehr großer Sprachmodelle mit Word-embeddings in der Form sogenannter Tokens. Das sind Kombinationen oder Bestandteile von Wörtern, die unter einer hochdimensionalen Abstandsmetrik in Beziehung gesetzt werden – typisch sind über 1000 Dimensionen. Je kürzer der Abstand, desto enger ist die Beziehung der Tokens. Large Language Models (LLMs) sind generative Modelle, die auf einer (unsupervised) vortrainierten Menge von Tokens Aufgaben der Art „Füllen von (maskierten) Lücken in einem Text“ bearbeiten. Bekannt wurde die LLM-Technik vorwiegend durch Systeme, die dem Sprachmodell mit einem Chat-Frontend den Einsatz als intelligentes Dialogsystem ermöglichten. Ein frei formulierter Text wird dabei in einem sogenannten „Prompt“ als Frage an das LLM gerichtet, die das LLM beantwortet. Das LLM erzeugt dann zunächst das erste Wort (aus den wahrscheinlichsten Tokens, die es für diesen Fragetext gelernt hat) und schreibt es in das „Chatfenster“, wo auch die Frage abgelegt wurde. Aus dem aktuellen Inhalt dieses Fensters (Frage und begonnene Antwort) erzeugt das LLM iterativ weitere Tokens für die Antwort, bis es die se entweder als beantwortet betrachtet oder bis der verfügbare Platz im Chatfenster erschöpft ist. Solange kein neuer Chat explizit gestartet wird, nimmt der Chatbot weitere Fragen entgegen. Es entsteht ein Dialog, der einen wachsenden Kontext für die Unterhaltung zwischen Menschen und Maschine bildet. Die Kunst, eine Sequenz guter Prompts zu schreiben, besteht darin, für eine komplexe Fragestellung schrittweise möglichst viel relevanten Kontext im Dialog aufzubauen („Prompt engineering“).

Solche Systeme bilden eine neue Art von KI-Systemen für neue Aufgaben, vorwiegend im Umfeld von Sprache. Sie können Texte in vielen Sprachen übersetzen, zusammenfassen oder erweitern, andere Systeme (wie DBMS, Tabellenkalkulation oder andere Office-Funktionen, Bilderkennung u.Ä.) in Verbände einbeziehen und sie können nach natürlich-sprachlicher Anweisung in diversen Computersprachen programmieren. Aber sie liefern Ergebnisse, die zumindest heute noch dem Risiko der Halluzination unterliegen und insofern nicht gesichert technisch korrekt sind, also einer Kontrolle bedürfen.

Chat ist nur eine spezielle Anwendung für LLM. Sie hat eine gewaltige Öffentlichkeitswirkung erzeugt: Jeder konnte ohne Vorkenntnisse mit einer Maschine in natürlicher Sprache reden und dabei, basiert auf einem gewaltigen gelernten Literaturwissen, erstaunlich kluge Antworten bekommen. Der Bot antwortet immer in der Sprache der jeweiligen Frage. Für weitergehende Anwendungen haben LLM eine API-Schnittstelle und themenspezifische „pretrained“ Varianten, beispielsweise für medizinische, juristische, steuerliche

oder andere Themenkreise, zu denen auch IT-Security gehört.

Für die meisten heute existierenden LLMs sind wenig Kapazitätsdaten veröffentlicht. Die aktuelle (noch frei verfügbare) Version 3.5 des GPT (Generative Pretrained Transformer) von OpenAI mag eine Vorstellung über die Dimension solcher Modelle vermitteln. Alle bisherigen Versionen des GPT sind decoder-only Transformer. Für GPT-3.5 gilt:

- 175 Mrd. Gewichte in 96 hidden layers
- Trainingsset aus 500 Mrd. Tokens (im Wesentlichen aus common crawl, books1, books2, Wikipedia, WebText2) im Umfang von 45 TB
- Training auf einer Konfiguration mit 285.000 CPU-Kernen und 10.000 GPUs, die kumulative Verweilzeit auf diesen GPUs (2020 Technologie) betrug ca. 355 GPU-Jahre (im Parallelbetrieb ca. 8 Monate)

Dieses System ist also etwa 2000-mal größer als die bislang größten pretrained Open source CNNs. Die kostenpflichtige Nachfolgeversion GPT-4 befindet sich noch im Training und wird mit etwa 6-facher Kapazität erwartet. Testversionen sind bereits verfügbar. Der konkurrierende Chatbot „BARD“ von Google ist noch in der Entwicklung, aber bereits zum kostenlosen Testen freigegeben. Die Benutzererfahrung ist ähnlich. Im Gegensatz zu GPT-3.5 sind die Daten tagesaktuell und die Antwortzeiten sind oft kürzer. BARD gibt auch Hinweise und Tipps, nach denen gar nicht gefragt wurde. Alle LLM-Systeme sind anfällig für Halluzination, besonders dann, wenn Prompts missverständlich sein könnten.

Zusammenfassend kann man zum aktuellen Stand der künstlichen Intelligenz feststellen:

- Neuronale Netze haben für diverse technische Anwendungsfelder übermenschliche Leistungsmerkmale in Präzision, Geschwindigkeit und Kapazität erreicht und haben neue Datentypen, insbesondere Graphen, als Anwendungsgebiete erschlossen (Graph Neural Networks – GNN). Solche Systeme dienen oft der Erkennung von Mustern in Bildern, Tönen, mathematischen Konstrukten oder systematischem Verhalten mit Zielen wie Klassifikation, Vorhersage, Lokalisierung oder semantischer Segmentierung. Meist haben diese Verfahren ein einziges spezielles Aufgabengebiet und erreichen darin hohe Zuverlässigkeit und geringe Fehlerraten. Man bezeichnet sie oft als klassisches ML.

- Aus einer neuen Methode zur Verarbeitung von natürlicher Sprache hat sich ein Ansatz (Transformer mit Attention-Mechanismen) entwickelt, der auf der Basis von sehr groß angelegtem, nicht überwachtem Lernen aus Literatur, Texten und Programmierung intelligente Antworten auf unterschiedlichste Fragen geben kann. Dieser Ansatz hat bisher nicht die Zuverlässigkeit erreicht, die in den klassischen monothematisch spezialisierten Systemen heute gegeben ist. Diese Systeme werden oft als Basismodelle (Foundation Models) oder als große Sprachmodelle (Large Language Models – LLM) bezeichnet und werden meist mit einer Dialogschnittstelle als Chatbots verwendet.

Eine Zero Trust Architecture (ZTA) besteht in der Regel aus mehreren Komponenten, die von spezialisierten Beratern integriert werden und für Überwachung oder Vertrauensmonitoring ML-Technologie verwenden. Die jeweiligen Softwarekomponenten sind dabei nicht unbedingt vom selben Hersteller. Mehrere bekannte Hersteller von IT-Security Software haben inzwischen Komponenten oder ganze Systeme angekündigt, die von einschlägig spezialisierten Integratoren zumindest in Teilbereichen den Aufbau einer ZTA und damit die Abkehr vom überkommenen Modell des Perimeterschutzes ermöglichen. Genaue Beschreibungen solcher Systeme sind, wie bei Security-SW üblich, schwer zu ermitteln, und Vergleiche können kaum angestellt werden.

Es wird davon ausgegangen, dass andere Systeme ähnliche Merkmale aufweisen wie die im Folgenden beschriebenen, die wir aus öffentlichen Quellen und aus Befragungen des Google Chatbots BARD für das Google-Cloud-System „BeyondCorp“ ableiten konnten. Sie soll als Beispiel für den Stand der heutigen Technik hier in ihren Kernkonzepten dargestellt werden.

#### 4.1 Stand der Technik: BeyondCorp Enterprise (BCE)

Die Softwaresuite BeyondCorp Enterprise (BCE) stellt eine ZTA dar, die im eigenen Konzern implementiert wurde und die inzwischen als Produkt im Markt angeboten wird. Es existiert ein frei verfügbares Papier (Rory Ward, Betsy Beyer: „BeyondCorp: A New Approach to Enterprise Security“, Usenix Mag. 39(6) (2014)), das die Konzepte und die Implementationssystematik darstellt. Mehrere LLMs berichten über die verwendeten Technologien, beispielsweise Graph Neural Networks (GNN) und andere ML-Methoden, die in BeyondCorp Enterprise (BCE) zum Einsatz kommen. GNNs gelten bereits seit dem bekannten Artikel von Keyvan Ramezanpour, Jithin Jagganath: „Intelligent Zero Trust

Architecture for 5G/6G Tactical Networks: Principles, Challenges and the Role of Machine Learning“<sup>[9]</sup> als die geeignete ML-Technik für die Implementation von Verfahren zur Erkennung von Anomalien in den Graphstrukturen von Netzwerken (siehe auch den Überblicksartikel „Automation and Orchestration of Zero Trust Architecture: Potential Solutions and Challenges“<sup>[10]</sup> und den speziellen Artikel „Anomalous distributed traffic: Detecting Cyber Security Attacks amongst microservices using Graph convolutional networks, ScienceDirect“<sup>[11]</sup>). Eine genaue Beschreibung der in BCE verwendeten und implementierten Technologien oder deren Konfigurationen ist natürlich nicht offengelegt.

Plattformen für BCE sind hybride Cloud- und On-Premise Umgebungen. Es werden Anwender wie Microsoft, Cisco und Amazon genannt, aber auch hier werden Konfigurationsdetails und Einsatzszenarien selbstverständlicherweise verschwiegen.

Googles Chatbot Bard beschreibt BCE auf Anfrage so: *„BeyondCorp ist ein Zero-Trust-Sicherheitsmodell, das von Google entwickelt wurde. Es basiert auf der Idee, dass kein Gerät oder Benutzer innerhalb eines Netzwerks automatisch vertrauenswürdig ist. Stattdessen wird der Zugriff auf Ressourcen auf der Grundlage des Kontexts gewährt, einschließlich der Identität des Benutzers, des Geräts, von dem er aus zugreift, und der Art der Ressource, auf die er zugreifen möchte. BeyondCorp basiert auf den Prinzipien:*

- *Vertrauen Sie niemandem: Kein Gerät oder Benutzer wird innerhalb eines Netzwerks automatisch als vertrauenswürdig angesehen.*
- *Gewähren Sie Zugriff auf Grundlage des Bedarfs: Der Zugriff auf Ressourcen wird auf der Grundlage des Kontexts gewährt, einschließlich der Identität des Benutzers, des Geräts, von dem er aus zugreift, und der Art der Ressource, auf die er zugreifen möchte.*
- *Untersuchen Sie verdächtige Aktivitäten: Anomalien und potenzielle Angriffe werden überwacht und untersucht.*

*BeyondCorp kann in einer Reihe von Umgebungen implementiert werden, einschließlich On-Premises-, Cloud- und hybriden Umgebungen. Es kann verwendet werden, um den Zugriff auf Unternehmensressourcen von überall aus zu schützen, unabhängig davon, wo sich die Benutzer oder Geräte befinden. BeyondCorp ist ein umfassendes Sicherheitsmodell, das Unternehmen dabei helfen kann, ihre Daten und Anwendungen vor unbefugtem Zugriff zu schützen.“*

Natürlich kennen auch andere Chatbots BeyondCorp und liefern ähnliche Beschreibungen.

BeyondCorp betreibt ein umfangreiches Partnernetzwerk unter dem Namen BeyondCorp Alliance, dem diverse namhafte große Hersteller aus dem Bereich Security-Software und -Lösungen angehören. Unter anderem werden Palo Alto Networks, Symantec, VMware, Citrix und CrowdStrike auf den öffentlichen Internetseiten<sup>1</sup> genannt.

Von der Autorenschaft des vorliegenden Diskussionspapiers konnten keine öffentlich verfügbaren, konkreten Leistungsdaten über ZTA-Implementationen im Hinblick auf Fehlalarme gefunden werden. Egal jedoch, mit welcher statistischen Methodik die eingesetzte KI arbeitet, die willkürliche Kalibrierung zur Steuerung der Anomalie-Erkennung für das Auftreten von false negative und false positive Alarmen bleibt ein intrinsisches Einstellungsproblem jeder speziellen Implementation.

## 4.2 Umsetzung von ZTA mit allgemeinen Large Language Models

Es existieren auch Vorschläge, wie man die ZTA-Prinzipien mit Unterstützung durch ein LLM umsetzen kann. Im Gegensatz zu dem oben beschriebenen Paradigmenwechsel, der zumindest bei Technologieführern offenbar angegangen ist, konnte die Autorenschaft in 2023 noch keine belastbaren Informationen über Nutzungsszenarien von LLMs in einer mit der Implementation von ZTA verbundenen Funktion recherchieren.

Auch Literatur-Recherchen zeigen noch keine konkreten Aussagen über Testinstallationen, Proofs of Concept (POC) oder erste Erfahrungsberichte und Testergebnisse. Offenbar gibt es aber in den Tiefen des Internets Quellen, aus denen LLMs gelernt haben, zu Aussagen über ihre Eignung zur Unterstützung von ZTA zu kommen. GPT-4 ist dabei eher zurückhaltend und bezeichnet LLMs nicht als zentrale Technologie für die Implementierung einer ZTA. GPT-3 (ChatGPT-3.5) und PaLM2 (BARD) präsentieren eine andere Einschätzung.

BARD und GPT-3.5 wurden von den Autoren nach solchen Vorschlägen befragt, und beide haben geantwortet, dass LLMs im Allgemeinen und ihre Systeme ChatGPT und BARD im Konkreten solche Unterstützung leisten können. ChatGPT verwechselte dabei BARD mit einem anderen System aus dem ZTA-Umfeld. Es werden zwar ausschweifende Antwort-

ten gegeben. Sie bleiben aber meist im Ungefähren, und man kann kaum beurteilen, welche Aussagen verlässlich sind, weil der BOT selbst keine Vorstellung von der „wahren“ Bedeutung seiner Aussagen haben kann. Es ist stets unklar, wo die Grenze zur Halluzination liegt: Hier ein Antwortbeispiel von BARD auf die von der Autorenschaft gestellte Frage nach dem Nutzen von LLM für ZTA:

*„Google Bard kann die Implementation einer Zero Trust Architecture unterstützen. Zero Trust basiert auf dem Prinzip, dass jeder Benutzer oder jedes Gerät, das versucht, auf eine geschützte Ressource zuzugreifen, als nicht vertrauenswürdig angesehen wird, bis seine Identität und Absichten verifiziert sind. Bard kann dabei helfen, dieses Prinzip umzusetzen, indem es folgende Aufgaben erfüllt:*

- **Authentifizierung und Autorisierung:** Bard kann verwendet werden, um Benutzer und Geräte zu authentifizieren und autorisieren. Dies kann durch die Verwendung von künstlicher Intelligenz und maschinellem Lernen erfolgen, um die Identität eines Benutzers oder Geräts zu überprüfen.
- **Netzwerksegmentierung:** Bard kann verwendet werden, um Netzwerke in Segmente zu unterteilen, die jeweils mit unterschiedlichen Sicherheitsrichtlinien versehen sind. Dies kann dazu beitragen, den Schaden zu begrenzen, wenn ein Angriff erfolgreich war.
- **Endpunktschutz:** Bard kann verwendet werden, um Endpunkte vor Angriffen zu schützen. Dies kann durch die Verwendung von künstlicher Intelligenz und maschinellem Lernen erfolgen, um bösartige Aktivitäten zu erkennen.
- **Log-Analyse:** Bard kann verwendet werden, um Sicherheitsprotokolle zu analysieren, um Anomalien und potenzielle Angriffe zu erkennen.

*Im Folgenden sind einige konkrete Beispiele dafür, wie Bard eine Zero Trust Architecture unterstützen kann:*

- Bard kann verwendet werden, um Benutzer anhand ihrer Verhaltensmuster zu authentifizieren. Dies kann dazu beitragen, bösartige Benutzer zu erkennen, die versuchen, sich mit gestohlenen Anmeldeinformationen anzumelden.
- Bard kann verwendet werden, um Geräte anhand ihrer Softwarekonfiguration zu authentifizieren. Dies kann dazu beitragen, Geräte zu erkennen, die mit Malware infiziert sind.
- Bard kann verwendet werden, um den Zugriff auf Ressourcen

<sup>1</sup> <https://cloud.google.com/blog/products/identity-security/google-cloud-announces-new-partners-in-its-beyondcorp-alliance>

cen auf der Grundlage des Benutzers, des Geräts und der Art der Ressource zu autorisieren. Dies kann dazu beitragen, den Schaden zu begrenzen, wenn ein Angriff erfolgreich ist.

- Bard kann verwendet werden, um Endpunkte auf Anomalien zu überwachen, die auf Angriffe hindeuten könnten. Dies kann dazu beitragen, Angriffe frühzeitig zu erkennen und darauf zu reagieren.“

Auch ChatGPT-3.5 behauptet, ähnliche Beiträge zu ZTA-Implementationen liefern zu können. Angesichts der Tatsache, dass LLM-basierte Chatbots bislang kaum echte seriöse Argumentationsfähigkeiten besitzen, die über den rein sprachbasierten Aspekt hinausgehen, ist in solchen Fällen eine sorgfältige Verifikation erforderlich.

In einem anderen Chat wurde GPT-3.5 nach Eigenschaften und Fähigkeiten von BARD befragt. Er verwechselte aber das Chat-Frontend mit einem ganz anderen Google-System, das speziell für die Unterstützung von ZTA positioniert ist, und kam deshalb zu offenkundigen Halluzinationen, die für beide Google-Produkte zweifelhaft waren. Es entstand der Eindruck, dass der GPT durch seine offensichtliche Verwirrung so sehr vom Kurs abkam, dass er über ein noch ganz anderes drittes Produkt sprach.

### 4.3 Umsetzung von ZTA mit fine-tuned Large Language Models

Neben der Integration von allgemeiner LLM-Technologie in eine individuell gestaltete ZTA existieren inzwischen vorgefertigte fine-tuned LLM-Varianten, die sich speziell auf den Security-Bereich richten und Schwerpunkte der ZTA-Prinzipien adressieren. SecPALM2 von Google ist ein Beispiel, das kurz vor der allgemeinen Produktfreigabe steht. Es wird nach MedPALM2 wahrscheinlich das zweite umfassende fine-tuned LLM auf der Basis von Googles PALM2 LLM sein. Aktuell sind noch keine konkreten Beschreibungen der Leistungsmerkmale und keine Positionierung der Einsatzschwerpunkte verfügbar. Insbesondere ist auch noch unbekannt, in welcher Hinsicht SecPALM2 über die allgemeinen Fähigkeiten von PALM2 und BARD im Hinblick auf Security und darin speziell auf ZTA hinausgeht.

Vor diesem Hintergrund wurde Bard dann von der Autorenschaft gefragt, ob BeyondCorp PALM2 nutzt und gab folgende Antwort:

„Ja, BeyondCorp nutzt PALM2, ein KI-Framework von Google, das für die Erkennung und Abwehr von Bedrohungen in Zero-Trust-Umgebungen entwickelt wurde. PALM2 verwendet maschinelles Lernen, um Anomalien in Benutzerverhalten und Geräteaktivitäten zu erkennen. Diese Anomalien können auf Angriffe hindeuten, wie z. B. Phishing, Malware oder Denial-of-Service-Angriffe. PALM2 wird in einer Reihe von Bereichen in BeyondCorp verwendet, darunter:

- **Authentifizierung:** PALM2 kann verwendet werden, um Benutzer anhand ihrer Verhaltensmuster zu authentifizieren. Dies kann dazu beitragen, bösartige Benutzer zu erkennen, die versuchen, sich mit gestohlenen Anmeldeinformationen anzumelden.
- **Autorisierung:** PALM2 kann verwendet werden, um den Zugriff auf Ressourcen auf der Grundlage des Kontexts zu autorisieren. Dies kann dazu beitragen, dass nur die Benutzer Zugriff auf Ressourcen erhalten, die sie für ihre Arbeit benötigen.
- **Verwaltung:** PALM2 kann verwendet werden, um die Sicherheit von BeyondCorp zu verwalten und zu verbessern. Dies kann dazu beitragen, dass das System auf dem neuesten Stand bleibt und Bedrohungen frühzeitig erkannt werden.

Im Folgenden sind einige konkrete Beispiele dafür, wie PALM2 in BeyondCorp verwendet wird:

- **Authentifizierung:** PALM2 kann verwendet werden, um die Tastaturgewohnheiten eines Benutzers zu analysieren, um festzustellen, ob sie sich von den gewohnten Gewohnheiten des Benutzers unterscheiden. Wenn PALM2 eine Anomalie erkennt, kann es einen Benutzer daran hindern, sich anzumelden.
- **Autorisierung:** PALM2 kann verwendet werden, um die Rolle eines Benutzers zu bestimmen, um festzustellen, ob er Zugriff auf eine bestimmte Ressource hat. Wenn PALM2 feststellt, dass ein Benutzer nicht autorisiert ist, kann es den Zugriff auf die Ressource verweigern.
- **Verwaltung:** PALM2 kann verwendet werden, um Sicherheitsprotokolle zu analysieren, um Anomalien und potenzielle Angriffe zu erkennen. Wenn PALM2 eine Bedrohung erkennt, kann es einen Sicherheitsadministrator benachrichtigen. PALM2 ist ein wichtiges Element von BeyondCorp, da es dazu beiträgt, die Sicherheit des Systems zu verbessern und Bedrohungen frühzeitig zu erkennen.“

#### 4.4 Einschätzung des Standes der Technik zur Anwendung von LLM zur Umsetzung von ZTA

LLM-basierte Chatbots haben bemerkenswerte Fähigkeiten, wenn die Aufgabe ausschließlich auf rein textbasierte Aspekte fokussiert, die weitgehend frei von semantischen Inhalten sind (übersetzen, zusammenfassen, themenbasiert entwerfen, programmieren nach Anweisung etc.), können aber in manchen Fällen versagen, in denen semantische Argumentation gefragt ist<sup>2</sup>. Im Prompt Engineering müssen Sondermaßnahmen ergriffen werden, um Halluzination zu verhindern – hier liegt heute noch eine große Schwäche der LLMs, deren Überwindung bislang noch nicht absehbar ist.

Microsoft hat die Testversion des GPT-4 im Edge Browser zum Test kostenlos bereitgestellt und an OpenAI einen zweistelligen Milliardenollarbetrag übertragen. Google spricht schon seit der globalen Entwicklerkonferenz 2023 über den PALM2-Nachfolger GEMINI, der GPT-4 übertreffen soll. Auch Microsoft hat bekannt gegeben, in das KI-Hardwaregeschäft mit eigenen Prozessoren einzusteigen. Das Rennen um die Vormachtstellung in KI-Performance nimmt immer mehr Fahrt auf, weil die Geschäftserwartungen förmlich explodieren.

Ob das neue Mega-Thema LLM die vollständige Automatisierung von ZTA beflügeln oder komplett umsetzen kann, bleibt eine offene Frage. Klar ist aber, dass IT-Security in diesem Fall ein Thema sein wird, das zu nennenswerten Teilen in den Oligopolbereich sehr weniger dominierender Cloud-Provider oder derer Partnernetzwerke überführt werden könnte.

2 Gemeint sind Aufgabenstellungen wie das Ablegen einer Prüfung nach dem Erlernen eines universitären Unterrichtsstoffs aus Vorlesungen, beispielsweise in Medizin, Jura, Betriebswirtschaftslehre oder anderen Fächern, in denen weitgehend Textwissen aufgenommen wird. Selbst an den namhaften amerikanischen Hochschulen bestehen LLMs Prüfungen zwar nicht unbedingt mit einem A-degree, aber mit guten B-degrees und sind somit qualifiziert/approbiert als Arzt, Anwalt, Buchhalter, Wirtschaftsprüfer o.Ä.

## 5. Rolle und Möglichkeiten des Einsatzes sowie Grenzen von Machine Learning in Verbindung mit ZTA-Prozessen

Die Verwendung von künstlicher Intelligenz im Allgemeinen sowie zur Verwirklichung von Zero-Trust-Netzwerk Architectures ist auf dem Vormarsch. Generelle Restriktionen beziehungsweise Grenzen der Umsetzung mit Hilfe von KI wurden ausführlich im Papier „Künstliche Intelligenz in Sicherheitsaspekten der Industrie 4.0“<sup>[12]</sup> diskutiert. Solche generellen Restriktionen beim Einsatz von KI sind beispielsweise:

- **Teilweise oder ungeeignete/unvollständige Lerndaten:** Wenn die Lerndaten das adressierte Problem nicht vollständig beschreiben, kann die KI auch nur ungeeignete Ergebnisse liefern.
- **Aspekte des Schutzes der KI gegen Manipulationen beziehungsweise Cyber-Angriffe:** Wie jede andere Security-Maßnahme ist das KI-gestützte Bewertungssystem gegen jegliche Manipulation zu schützen, angefangen beim Lernprozess, beispielsweise durch verfälschte/manipulierte Lerndaten, bis zum produktiven Einsatz.

Beispielsweise Manipulation des fertig gelernten neuronalen Netzwerks oder Angriffe mithilfe von KI, bei denen die manipulierte Datenzugriffsanfrage von der legitimen Anfrage kaum mehr zu unterscheiden ist.

Grundsätzlich ist eine eigenständige Aktion oder Reaktion einer KI – genau wie andere automatische Aktionen außerhalb einer KI-Steuerung – an unternehmensspezifische Policies gebunden und muss entsprechend dieser Policies trainiert sein. Im Rahmen der Erarbeitung solcher Policies ist eine Risiko-Betrachtung bzgl. der Eigenständigkeit einer eingesetzten KI erforderlich. Grundsätzlich ist zu bedenken, dass eine KI ein Werkzeug ist und daher bei Fehlern nicht zur Verantwortung gezogen werden kann. Im Industrie 4.0-Kontext befinden sich die Verantwortlichkeiten im Bereich des Betreibers oder des Herstellers. Rechtliche Aspekte des Einsatzes von KI werden in diesem Dokument jedoch nicht näher behandelt, sondern in anderen Industrie 4.0-Arbeitsgruppen.

Attribute Based Access Control (ABAC) kann eine Basis zur Umsetzung von ZTA sein: Nach NIST 800-162 werden für ABAC mehrere Instanzen beschrieben, die in ZTAs zur Anwendung kommen sollten. Dazu gehören:

- Anfrage an Policy Administration (PA)
- ML-basierte Prüfung der Policy Engine (PE)
- Verbindungsherstellung über Policy Enforcement Point (PEP)

Diese unterschiedlichen Instanzen sind für eine ZTA-Einführung in die industrielle Produktion zu definieren, festzulegen und einzuführen. Sie könnten Teile eines ZTA-Management-Systems sein.

Für Anlagen, Geräte und Maschinen können beispielsweise folgende „Credentials“ definiert werden und zur Anwendung kommen: MAC-Adresse, IP-Adresse, Port, Ort, Zeitpunkt, Zweck, aber auch Stromverbrauch etc. Für Personen in der Produktion und in der Verwaltung kommt übliche Multifaktor-Authentifizierung (z. B. Name, Rolle, PIN, Biometrie etc.) zum Einsatz.



## 6. Handlungsempfehlungen

Die AG „Sicherheit vernetzter Systeme“ hat bereits in der Publikation „Künstliche Intelligenz (KI) in Sicherheitsaspekten der Industrie 4.0“<sup>[13]</sup> darauf hingewiesen, dass KI sowohl zur Verbesserung des Schutzes eingesetzt werden als auch durch Angreifer als Angriffswerkzeug genutzt werden kann.

Die seinerzeit beschriebenen Maßnahmen und Sicherheitsregeln werden mit weitergehender Automatisierung und KI-gestützter Weiterentwicklung der Gefahrenabwehr immer wichtiger: Eine ZTA sollte sich nie auf Standard-Software im Auslieferungszustand verlassen, sondern immer ein hohes Maß an Individualisierung erfahren und diese sollte nicht statisch sein, sondern einem dynamischen Veränderungsmechanismus ihrer Policies unterliegen und

dabei technische Fortschritte möglichst umgehend implementieren.

Die damals beschriebenen generativen feindlichen Algorithmen haben sich inzwischen auf viele neue Anwendungsfelder ausgedehnt und können heute nicht nur Bild, Ton, Video oder andere vermeintlich sicher unterscheidbare Muster perfekt fälschen. Mit weiterentwickelter GAN-Technologie können die spieltheoretisch erzeugten Gleichgewichte, die Abwehrsysteme blenden, gefunden werden, sobald erkannt wird, wie in der jeweiligen Implementierung der ZTA Anomalien aufgespürt werden. Es ist deshalb auch empfehlenswert, den Fortschritt der professionell wissenschaftlich erzeugten Technologie-Fortschritte solcher Algorithmen mittels kompetenter Beratung zu verfolgen.

## 7. Zusammenfassung und Ausblick

Zero Trust Architecture (ZTA) ist ein Sicherheitsansatz, der auf der Annahme basiert, dass kein Benutzer, Gerät oder Netzwerk a priori vertrauenswürdig ist. Daher müssen alle Zugriffsversuche auch innerhalb eines Netzwerkes von vermeintlich vertrauenswürdigen Assets authentifiziert und autorisiert werden.

Der Einsatz von KI für ZTA kann vor diesem Hintergrund dazu beitragen, die Sicherheit zu verbessern, indem komplexere und präzisere Entscheidungen über den Zugriff auf Ressourcen getroffen werden.

Die Grundprinzipien der IT-Sicherheitstechnologie befinden sich in einem grundlegenden Wandel, der schnell voranschreitet, dabei bereits große Veränderungen hervorgebracht hat und dessen erwartbares Ende in keiner Weise vorhersehbar ist. Die IT-Sicherheitstechnologie ist nach den Vorbildern, die wir aus den Entwicklungen anderer Technologiefelder seit Jahren kennen, auf dem Weg zu einer KI-basierten Automatisierung – Sicherheitssysteme werden autonom und übernehmen in wachsendem Umfang ihre eigene Steuerung.

In Analogie zu den hauptsächlichlichen Verkehrssystemen wie Automobil, Bahn, Flugzeug, Schiff oder Maschinen, die immer weniger Steuerung durch menschliche Bediener brauchen und deshalb die Fehlerursache „menschliches Versagen“ trotz stark wachsender Leistung weitgehend ausschalten, wächst die Wirtschaftlichkeit der darauf basierenden Prozesse. Vom autonomen Auto wird erwartet, dass diese Technologie verglichen mit den menschlich gesteuerten Vorgängern und dem Einsatz von Assistenzsystemen und autonomen Steuerungen ca. 10 Prozent weniger Verkehrsunfälle, viel weniger Verletzungen und Verkehrstote und signifikante Zeit- und Kosteneinsparungen bewirkt.

ZTA ist der Schritt in Richtung einer weitgehend autonomen Sicherheitsarchitektur, vor allem in den Unternehmen der Industrie und den Wertschöpfungsketten des produzierenden Gewerbes. Die Vision besteht darin, dass immer seltener Security Incidents in wochenlangen forensischen Analysen verstanden werden müssen, um den nächsten Angriff zu verhindern. Vielmehr wird erwartet, dass sich Technologie für IT-Security durch autonome Steuerungen in wenigen Jahren so weit entwickelt, dass sie nach erfolgreicher autonomer Abwehr dem Sicherheitspersonal einen Bericht über den nicht eingetretenen Zwischenfall liefert. Die Rolle des Personals verändert sich also von Analyse und Ursachenforschung der oft teuren Security Incidents auf die ständige Weiterentwicklung und Perfektionierung des autonomen Abwehrsystems, das die Zahl der aufgetretenen Zwischenfälle drastisch verringert.

Im Zuge dieser Entwicklung ist damit zu rechnen, dass die IT-Security-Software-Industrie in dem neuen ZTA- und Autonomie-bezogenen Software- und Dienstleistungsmarkt eine deutliche Spezialisierung und Konzentration erleben wird. Große Teile der Implementierungen und Funktionsbereitstellungen werden in den Märkten der dominierenden Cloud-Provider erfolgen, die hier eventuell große neue Geschäftspotenziale finden. Eine signifikante Zukunftsindustrie fokussiert sich wahrscheinlich auf Nordamerika und China.

Die weitere Entwicklung von ZTA wird voraussichtlich von folgenden Aspekten geprägt sein:

- Identifikation, Authentifikation und Autorisierung werden weiter vor dem Hintergrund zunehmender Anzahl kommunizierender Assets automatisiert.
- Beschränkte Anzahl von Anbietern für Identifikation, Authentifikation und Autorisierung, die über ausreichende Infrastrukturen verfügen. Dementsprechend ist mit einem zunehmend monopolisierten Markt zu rechnen.
- Eine Differenzierung zwischen den Anbietern von ZTA-Implementierungen wird aufgrund einer fehlenden Metrik bezüglich des Sicherheitsgewinns nicht erfolgen können.
- Dezentrale Ansätze zu verfolgen, die global auf Basis internationaler Standardisierung durchsetzbar sind, ist wünschenswert.

Werden die zukünftigen Sicherheitstechnologien so intelligent sein, dass sie industrielle Sicherheitsrisiken eliminieren und auch die „fortgeschrittensten“ Hacker keine Chance mehr haben? Natürlich nicht: Es wird immer eine Grenze für die Autonomie geben, die „der Mensch“ als solcher an eine Maschine aus rationaler Erwägung abgibt. Auch das autonome Auto wird Unfälle verursachen und die autonome ZTA wird Security Incidents erleben. Es wird also Grenzen geben, an die sich die autonome KI halten muss, Grenzen, die mit statistisch ermittelten Restrisiken für die jeweiligen Entscheidungskompetenzen rational von Menschen in einer geeigneten Sicherheitspolitik geregelt werden. Die letzte Instanz der Verantwortung kann auch in der vollkommen implementierten maschinellen Autonomie keine Maschine sein.

Das vorliegende Diskussionspapier stellt den gerade beginnenden Paradigmenwechsel zum Stand Frühjahr 2024 in seinen Grundzügen dar. Wir planen ein Folgedokument nach ca. 12 Monaten zu veröffentlichen, um so die Dynamik des ablaufenden Prozesses im Jahr 2025 zu belegen.

# 8. Anhang

## 8.1 Einführung eines Zero-Trust-Maturity-Modells

Zero Trust ist keine fertig etablierte Lösung, sondern ein andauernder Prozess, der entsprechend nicht ad hoc implementierbar ist. Dies gilt auch für viele Use Cases der Industrie 4.0.

Insbesondere in der Industrie (3.0, 4.0) sind gewachsene Strukturen und etablierte Prozesse mitunter starr und historisch gewachsen. Dadurch ergibt sich ein „Brown Field“ für die Einführung von Zero Trust Architecture, das größere (aber schrittweiser) Veränderungen der technischen Landschaft bedarf.

Ein Zero-Trust-Maturity-Modell beschreibt den Grad der Implementierung einer Zero Trust Architecture. Es gibt mehrere Abbildungen von Zero-Trust-Maturity-Modellen (z. B. das Modell der U.S. Cybersecurity and Infrastructure Security Agency <sup>[14]</sup>, das Modell von Palo Alto Networks <sup>[15]</sup> sowie Abbildungen in verschiedenen Forschungsarbeiten <sup>[16]</sup>, <sup>[17]</sup>).

Im Weiteren wird auf das Maturity-Modell der CISA<sup>[14]</sup> Bezug genommen.

Es basiert auf fünf Säulen, die sich aus Identität (Identity), Geräten (Devices), Netzwerk (Networks), Anwendungen (Applications), Arbeitsbelastung (Workloads) und Daten (Data) zusammensetzen. Innerhalb dieser fünf Säulen findet die Anpassung verschiedener Faktoren statt, die zur Optimierung des Zero-Trust-Grades führen. Dabei werden vorhandene Prozesse beleuchtet und vereinheitlicht innerhalb der jeweiligen Teilbereiche. Die Geschwindigkeit des Optimierungsgrades kann von Säule zu Säule variieren.

Der Reifegrad des Zero-Trust-Modells pro Säule wird zuzätzlich in vier verschiedene Stufen eingeteilt<sup>3</sup>, anhand derer die zu ergreifenden Maßnahmen zur Implementierung abgeleitet werden. Das Maturity-Modell ist geeignet, den Status der Umsetzung des jeweiligen Use Cases zu überprüfen.

Ein Security-Maturity-Level kann u. a. auch als eine Mindestanforderung im Standard einer Geschäftsbeziehung herangezogen werden. Allerdings sind nicht für alle Use

Cases auch alle Säulen relevant, bei welchen das Maturity-Level geprüft wird. Je nach Erfahrungswert der Stakeholder in den Use Cases empfiehlt sich die Anwendung eines Maturity-Modells zur Prüfung des Zero-Trust-Grades und zur Implementierung und andauernden Überprüfung des Konzepts.

Im Industrie 4.0-Kontext hilft ein Zero-Trust-Maturity-Modell, beispielsweise bei der Bestimmung des Abdeckungsgrades einer kontinuierlichen Prüfung von insbesondere Datenzugriffen auf Daten aus Verwaltungsschalen oder innerhalb eines Datenraumes beziehungsweise Daten-Ökosystems (z. B. Catena-X, Manufacturing-X), sowie bei der Prüfung von Kommunikationswegen innerhalb einer Industrie 4.0-Wertschöpfungskette.

## 8.2 Detailbetrachtung der Beispiele zum Industrie 4.0-Kontext von Zero-Trust-Ansätzen aus Kapitel 5

Die Detailbetrachtungen in den folgenden Unterkapiteln untersuchen für die gegebenen Industrie 4.0-Beispiele verschiedene Security-Aspekte, wie Teilnehmerprüfung, Authentisierung, Autorisierung beim Datenzugriff, Prüfung der Sicherheit der Kommunikation und Kommunikationswege sowie Sicherheit der Datenablage. Diese Betrachtungen werden durchgeführt unter Bezugnahme auf die in Kapitel 2 erläuterten Zero-Trust-Prinzipien:

- IAM (Identity and Access Management)
- Least privileged access
- Mikrosegmentierung
- MFA (Multifaktor-Authentifizierung)
- Aufklärung und Security-Analytik

Es findet des Weiteren eine Gegenüberstellung der gängigen Security-Maßnahmen mit erweiterten Security-Maßnahmen, die zur Vervollständigung von Zero-Trust-Ansätzen beitragen können, statt. Hierbei kann oder muss KI zum Einsatz kommen, um komplexe Sachverhalte grundsätzlich prüfbar zu gestalten.

3 Traditionell (traditional), initial (initial), fortgeschritten (advanced), optimal (optimal).

### 8.2.1 Beispiel der Teilnahme an einer Industrie 4.0-Wertschöpfungskette

Innerhalb der Wertschöpfungskette kommunizieren deren Teilnehmer (Menschen und Maschinen) miteinander und erlangen Zugriff auf für sie notwendige (Teil-)Daten des Produktes, der Produktion oder des Produktionsprozesses. In diesem aktiven Zustand der Wertschöpfungskette werden im Folgenden mehrere Security- und Zero-Trust-Aspekte betrachtet.

#### Generelle Prüfung der Teilnehmer im Vorfeld (Akkreditierung):

Im Vorfeld der Teilnahme an einer Wertschöpfungskette stellt sich die Frage, unter welchen Vorbedingungen ein potenzieller Teilnehmer beziehungsweise Kandidat an der Wertschöpfungskette initial und dauerhaft teilnehmen darf. Bei der Beurteilung spielen grundsätzliche Fragen zum Kandidaten eine wichtige Rolle, wie z. B.:

- Ist der Kandidat bereits bekannt?
- Ist der Kandidat vertrauenswürdig?
- Ist der Kandidat solvent?
- etc.

**Stand heute** werden diese Prüfungen teils automatisiert, teils manuell recherchiert vorgenommen. Das geschieht durch die Beauftragung von Dienstleistern oder durch eigene Recherchen seitens des Initiators der Wertschöpfungskette, z. B. dem Auftraggeber in Bezug auf den Auftragnehmer oder Zulieferer. Die Prüfung wird sehr häufig einmalig oder in regelmäßigen, größeren Abständen (z. B. jährlich) vorgenommen. Bei einigen Datenökosystemen (z. B. Catena-X und Manufacturing-X) ist nach der Akkreditierung für das Ökosystem eine grundsätzliche Hürde zur Inanspruchnahme der Services des Ökosystems überwunden, und der Teilnehmer erwirbt dadurch das Recht, grundlegende Daten des Ökosystems zu konsumieren.

**Eine Annäherung an Zero-Trust-Ansätze** ist erzielbar u. a. durch regelmäßige Aktualisierung der Prüfungsergebnisse, nicht nur zu Beginn, sondern auch während der aktiven Phase der Wertschöpfungskette. Dabei können kritischere Teilaspekte ggf. öfter geprüft werden als unkritischere. Ein Beispiel für kritischere Prüfungen ist die Überprüfung von aktualisierten beziehungsweise dynamischen Handelsverboten, die den Kandidaten betreffen könnten. Eine unkritischere Prüfung könnte z. B. die Prüfung der Gesellschaftsform sein.

#### Authentisierung/Authentifizierung während des produzierenden Zustands der Wertschöpfungskette:

Während der Teilnahme an der Wertschöpfungskette stellt sich die Frage, wie der Teilnehmer in der Wertschöpfungskette eindeutig und sicher authentisiert werden kann.

**Stand heute** geschieht dies durch ein Zertifikatsmanagement, welches durch eine entsprechende vertrauenswürdige Infrastruktur bereitgestellt wird, sowie bekannte Arten von Multifaktor-Authentisierungsmethoden.

Information: Der **Stand heute** würde einer Stufe 1 bis 2 im Rahmen eines Reifegradmodells entsprechen, die Maßnahmen zur Vervollständigung würden einer Stufe 3 entsprechen.

**Eine Annäherung an Zero-Trust-Ansätze** ist beispielsweise erzielbar durch eine aktive Bewegungsanalyse des Teilnehmers, innerhalb und außerhalb der Wertschöpfungskette (z. B. innerhalb eines Data Spaces/Datenökosystems/Marktplatzes), mithilfe von KI.

Die Prüfung im Vorfeld der Teilnahme und die Teilnahme an der Wertschöpfungskette kann den Zero-Trust-Prinzipien „IAM (Identity and Access Management)“ und „MFA (Multifaktor-Authentifizierung)“ zugeordnet werden.

#### Autorisierungsprüfung beim Datenzugriff:

Nach der Authentisierung eines Teilnehmers stellt sich sofort die Frage, für welche Datenzugriffe der Teilnehmer im Detail berechtigt ist und wie die Zugriffsprüfung erfolgt.

Berechtigungen werden zum Teil rollenbasiert vergeben, zum Teil sind sie regelbasiert.

Ein Beispiel für eine dauerhafte (rollenbasierte) Berechtigung ist Mitgliedschaft in einem Datenökosystem (Catena-X oder Manufacturing-X). Dabei ist das Mitglied dauerhaft berechtigt, auf Services und bestimmte Daten innerhalb des Ökosystems zuzugreifen. Durch Vereinbarungen über Nutzungsrechts-Erklärungen können ggf. weitere Datenzugriffe ermöglicht werden.

**Stand heute** geschieht die dauerhafte (statische) Berechtigungsprüfung über Rollenvergabe und die Prüfung auf die Rolle zum Zeitpunkt des Datenzugriffs.

**Eine Annäherung an Zero-Trust-Ansätze** ist erzielbar durch zusätzliche regelmäßige Prüfungen, die über die reine Mitgliedschaftsprüfung hinausgehen. Beispielsweise können zusätzliche Individualisierungsprüfungen der Bedingungen der Nutzungsrechts-Erklärung stattfinden. Diese Prüfung erfolgt wiederum durch regelbasierte Autorisierung (siehe unten).

Ein Beispiel für eine regelbasierte Prüfung ist, ob der Teilnehmer sich gerade in einem bestimmten Produktions-Kontext befindet und deshalb gerade das Recht hat, auf Baupläne oder Teile davon zuzugreifen.

**Stand heute** finden solche regelbasierten Prüfungen, auch unter Berücksichtigung von Kontext-Daten, statt.

**Eine Annäherung an Zero-Trust-Ansätze** ist erzielbar durch zeitnahe Anpassung von Regeln und/oder durch KI-basierte Prüfung bezüglich der Normalität/Nicht-Normalität der Abfrageparameter für Zugriffsanfragen. Die Autorisierungsprüfung kann dem Zero-Trust-Prinzip „IAM (Identity and Access Management)“ zugeordnet werden.

#### Prüfung der Sicherheit der Datenkommunikation und der Kommunikationswege

Während der Kommunikation von Daten über jegliche Art von Datenprotokollen stellt sich die Frage, ob die Daten-Verschlüsselungs- und -Übertragungstechnologie security-technisch auf dem neuesten Stand ist.

**Stand heute** geschieht dies in der Regel durch die Verwendung von Verschlüsselungsalgorithmen und Datenübertragung auf dem neuesten Stand der Technik (z. B. neueste Versionen des Protokolls „Transport Layer Security“). Updates auf die entsprechenden Protokoll-Versionen werden meist nach und nach und häufig manuell durchgeführt, u. a. wegen der Heterogenität der unterschiedlichen Umgebungen (IT, OT, verschiedene Protokolle, verschiedene Software-Stände, verschiedene nationale oder industriespezifische Regularien). Schlüssel werden regelmäßig ausgetauscht, teilweise manuell, teils semi-automatisch.

**Eine Annäherung an Zero-Trust-Ansätze** ist erzielbar durch striktere Begrenzung der Gültigkeit von Schlüsseln und Zertifikaten (höher getakteter und automatisierter Austausch) sowie KI-basierte/regelbasierte Prüfung von angesprochenen Endpunkten, Kommunikationsfehlern und Häufigkeiten. Um auf dem neuesten Stand der Technik bezüglich der Kommunikationsprotokolle zu bleiben, wären geeignete Kommunikations-Zwischenlayer einzuführen, die die aktuelle Heterogenität der Landschaften bei der Kommunika-

tion zwischen verschiedenen Teilnehmern vereinheitlichen und als auch zentrale Protokoll- und Schlüssel-Updates erfahren können.

Eine weitere Frage, die sich stellt, ist, ob die Start- und Endpunkte für Anfragen erlaubt/normal sind, beziehungsweise ggf. verdächtig/suspekt sind.

**Stand heute** geschieht diese Prüfung in der Regel mithilfe von Perimeter-Schutzmechanismen, wie z. B. durch Firewalls, Proxys und Gateways, die regelbasiert den Datenverkehr erlauben oder blockieren und aktuelle Security-Daten berücksichtigen (z. B. Prüfung gegen eine aktuelle Liste verdächtiger IP-Adressen im Internet). Häufig werden auf dem Kommunikationsweg auch Protokollwechsel durchgeführt, sodass eine Kommunikation mit dem Startprotokoll des Senders unter Umgehung des Protokollumsetzers nicht möglich ist.

**Eine Annäherung an Zero-Trust-Ansätze** erfolgt durch KI-basierte Prüfung von Häufigkeiten oder plötzlichen/allmählichen Verschiebungen der erlaubten Kommunikationswege/Endpunkte.

Die Kommunikationsprüfung kann den Zero-Trust-Prinzipien „Mikrosegmentierung“ und „Aufklärung und Security-Analytik“ zugeordnet werden.

Generell wird die Datenkommunikation über die verschiedenen Gesichtspunkte hinweg mit einem Echtzeit-Monitoring überwacht, welches im einfachsten Fall regelbasierte Prüfungen durchführt und Alarme erzeugt, falls verdächtige Zugriffe oder Zugriffsversuche erfolgen. Das Regelwerk kann im Rahmen eines Zero-Trust-Ansatzes mit KI-basierten Regeln ergänzt werden, die ungewöhnliche Kommunikation (z. B. Kommunikationswege, Protokolle, Header- und Dateninhalte, gewählte Security-Einstellungen etc.) erkennen und alarmieren.

#### Technische und semantische Prüfung des Dateninhaltes von ausgetauschten Nachrichten

Die nächste Frage ist, ob eine Dateninhaltsprüfung der Anfragen und Antworten auf technische und semantische Korrektheit erfolgt und wie diese Prüfungen möglichst performant und vollständig ausgeführt werden können.

**Stand heute** geschieht dies in der Regel durch feste Regeln am Endpunkt (beziehungsweise beim Anfrager), die die Daten auf die Konsistenz des Datenformats (z. B. gültiges Datumsformat, wohlgeformtes JSON-Dokument) prüfen sowie für ausgewählte (und durchaus viele) Datenkombinationen die

Gültigkeit prüfen (z. B. wenn der Auftrag vom Typ „Kleinmenge“ ist, darf die Menge nicht mehr als 10 sein).

**Eine Annäherung an Zero-Trust-Ansätze** ist erzielbar durch die KI-basierte Prüfung aller Datenabfrage-Parameter-Kombinationen sowie die KI-basierte Prüfung der Antwortdaten bezüglich aller zurückgegebenen Attribute. Außerdem kann die Prüfung weiterer Kontext-Daten, die auf anderen Kanälen zwischen den Teilnehmern ausgetauscht werden (z.B. bekannter statischer Zustand einer I40-Entität oder aktuelle Produktionskapazität, gegen die die tatsächliche Produktionsleistung abgeglichen werden könnte), durchgeführt werden

Die Dateninhaltsprüfung kann den Zero-Trust-Prinzipien „Least privileged access“ und „Aufklärung und Security-Analytik“ zugeordnet werden.

#### Prüfung der Sicherheit der Datenablage

Eine weitere Frage ist, ob persistierte Daten sicher (d.h. nicht löschar) und verfälschungssicher aufbewahrt werden und ob sicherstellbar ist, dass keine Manipulation der Daten auf anderen Kommunikationskanälen stattfindet (z. B. durch direkten Zugriff auf das File-System oder die Datenbank).

**Stand heute** geschieht die Sicherstellung in der Regel durch Verschlüsselung der Persistenz (der File-Ablage oder der Datenbank) sowie durch rollen- oder regelbasierte Autorisierungsprüfungen beim Datenzugriff über alle Zugriffskanäle hinweg.

**Eine Annäherung an Zero-Trust-Ansätze** ist erzielbar durch die bereits oben genannten, dynamischeren und KI-basierten Prüfungen für die Identität, die Datenkommunikation, die Datenkommunikations-Inhalte in Kombination mit den Zugriffskanälen. Die KI kann Fragen dieser Art beantworten: „Greifen Identitäten dieses Identitäts-Typs normalerweise auf diese Art von Daten mit dieser Anfrage-Parametrisierung über diese Zugriffskanäle auf diese Art von Daten und diese Menge von Daten zu, oder ist dies eher nicht normal beziehungsweise verdächtig?“

Die Datenablageprüfung kann den Zero-Trust-Prinzipien „Mikrosegmentierung“ und „Aufklärung und Security-Analytik“ zugeordnet werden.

## 9. Literaturverzeichnis

- 1 J. Kindervag, „Build Security into Your Network’s DNA: The Zero Trust Network Architecture“, 2010.
- 2 „BeyondCorp: A new approach to Enterprise Security“, 01.12.2014 [Online]. Available: <https://static.googleusercontent.com/media/research.google.com/de//pubs/archive/43231.pdf> [Zugriff am 29.02.2024].
- 3 AG „Sicherheit vernetzter Systeme“, *IT-Security in der Industrie 4.0 – Handlungsfelder für Betreiber*, Berlin: Plattform Industrie 4.0, 2016.
- 4 AG „Sicherheit vernetzter Systeme“, *Zugriffsteuerung für Industrie 4.0*, Berlin: Plattform Industrie 4.0, 2018.
- 5 AG „Sicherheit vernetzter Systeme“, *Sichere Kommunikation*, Berlin: Plattform Industrie 4.0, 2018.
- 6 AG „Sicherheit vernetzter Systeme“, *Vertrauensinfrastrukturen für Industrie 4.0*, Berlin: Plattform Industrie 4.0, 2021.
- 7 AG „Sicherheit vernetzter Systeme“, *Sichere digitale Identitäten*, Berlin: Plattform Industrie 4.0, 2016.
- 8 D. E. Rumelhart, G. E. Hinton und R. J. Williams, „Learning representations by back-propagating errors“, in *Nature* 323, New York, Nature, 1986, pp. 533–536.
- 9 J. K. Ramezanzpour, „Intelligent Zero Trust Architecture for 5G/6G Tactical Networks: Principles, Challenges and the Role of Machine Learning“.
- 10 Y. Caoa, S. R. Pokhrel, Y. Zhu, R. Doss und G. Li, „Automation and Orchestration of Zero Trust Architecture: Potential Solutions and Challenges“, *TechRxiv*, 2022.
- 11 S. Jacob, Y. Qiao, Y. Ye und B. Lee, „Anomalous distributed traffic: Detecting cyber security attacks amongst microservices using graph convolutional networks“, *Computers & Security*, 2022.
- 12 AG „Sicherheit vernetzter Systeme“, *KI in Sicherheitsaspekten von Industrie 4.0*, Berlin: Plattform Industrie 4.0, 2019.
- 13 AG „Sicherheit vernetzter Systeme“, *Industrial security and the development of AI applications in the edge*, Berlin: Plattform Industrie 4.0, 2022.
- 14 CISA, „Cybersecurity and Infrastructure Security Agency“, Cybersecurity and Infrastructure Security Agency, 01.04.2023. [Online]. Available: [https://www.cisa.gov/sites/default/files/2023-04/zero\\_trust\\_maturity\\_model\\_v2\\_508.pdf](https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf) [Zugriff am 20.11.2023].
- 15 Palo\_Alto\_Networks, „Federal News Network“, Palo Alto Networks, 01.01.2020. [Online]. Available: <https://federalnewsnetwork.com/wp-content/uploads/2020/01/zero-trust-maturity-model.pdf> [Zugriff am 20.11.2023].
- 16 H. C. G. L. Ali, „A Maturity Framework for Zero-Trust Security in Multiaccess Edge Computing“, Bd. 2022, Nr. 3178760, 2022.
- 17 L. S. J. Yeoh, „Zero trust cybersecurity: Critical success factors and A maturity assessment framework“, Bd. 133, Nr. 103412, 2023.
- 18 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser und I. Polosukhin, „Attention Is All You Need“, 01 Juni 2017. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017arXiv170603762V/abstract> [Zugriff am 29.02.2024].

## AUTOREN

Björn Flubacher (Bundesamt für Sicherheit in der Informationstechnik) | Dr. Detlef Houdeau (Infineon Technologies AG) |  
Dr. Bernd Kosch (Industrie KI GmbH) | Dr. Michael Schmitt (SAP SE) | Thomas Walloschke (Leitung, secon trust consult)

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

