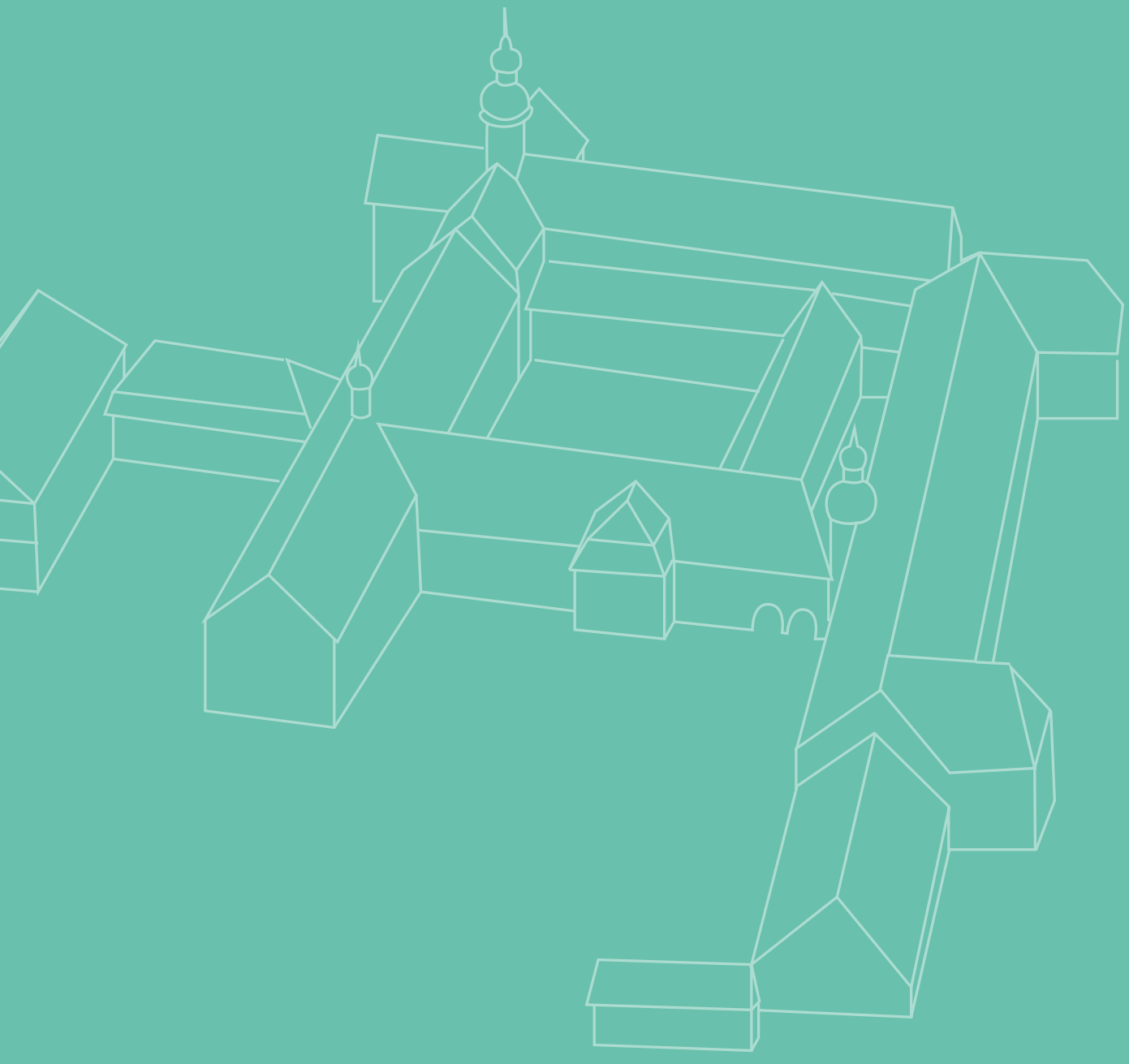
Fraunhofer SIT

FRAUNHOFER INSTITUTE FOR SECURE INFORMATION TECHNOLOGY

EBERBACHER GESPRÄCH ON »AI, SECURITY & PRIVACY«

08/2020

Präambel



Angewandte Forschung zur IT-Sicherheit braucht den Dialog zwischen Wissenschaft und Wirtschaft, um anwendungsrelevante Antworten auf die grundsätzlichen Fragestellungen zu erhalten: Was sind die aktuellen Herausforderungen für IT-Sicherheit und Privatsphärenschutz? Was ist für die Zukunft zu erwarten? Was kann und soll Technik leisten? Wo sind die Grenzen des Machbaren? Wo braucht es neue Ideen?

Die »Eberbacher Gespräche« des Fraunhofer SIT bieten ein Forum für diesen Dialog. Experten aus Wissenschaft, Wirtschaft und Verwaltung treffen sich für jeweils einen Tag und erarbeiten für ein Thema gemeinsam Antworten auf diese Fragen. Im August 2019 ging es um »AI, Security & Privacy«. Teilnehmer der Veranstaltung waren:

Experten extern

- Alfred Ermer, arago GmbH
- Gerold Hübner, SAP AG
- Stefan Römmele, Continental AG
- Dr. Ralf Schneider, Allianz Deutschland AG
- Borislav Tadic, Deutsche Telekom AG
- Dr. Michael Tagscherer, Giesecke+Devrient
- Dr. Michael von der Horst, CISCO Systems GmbH
- Robert Zehder, Deutsche Telekom AG

ATHENE / Fraunhofer SIT

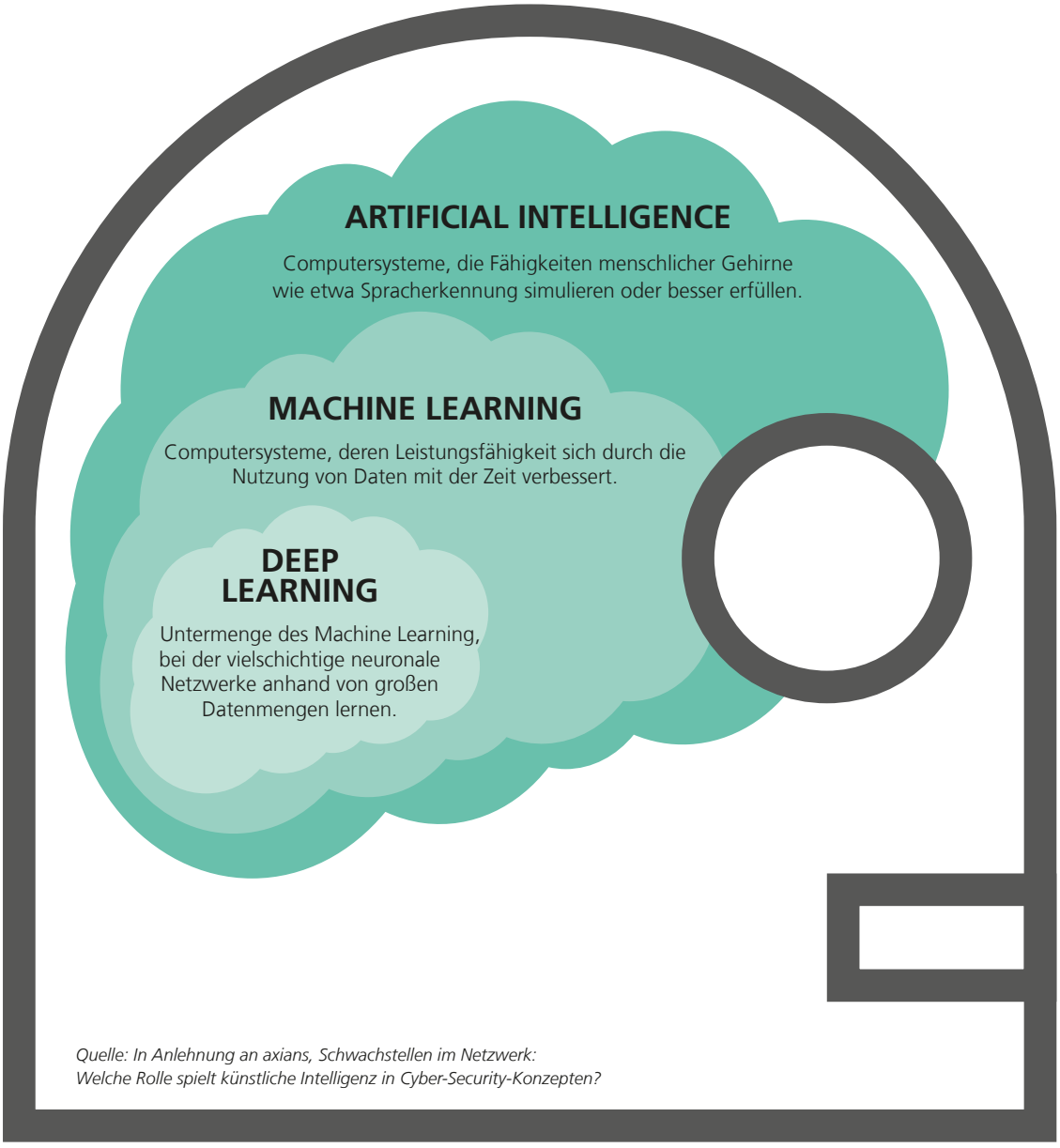
- Dr. Michael Kreutzer
- Prof. Dr. Martin Steinebach
- Prof. Dr. Michael Waidner
- Dr. Sascha Zmudzinski

Autoren

Dr. Michael Kreutzer, Oliver Küch, Prof. Dr. Martin Steinebach

Aus Gründen der leichteren Lesbarkeit wird im Text die männliche Sprachform verwendet. Dennoch beziehen sich sämtliche Aussagen gleichermaßen auf Frauen und Männer sowie Personen des dritten Geschlechts. Das Vorgehen dient lediglich der sprachlichen Vereinfachung und impliziert keine Benachteiligung einer Personengruppe. Dementsprechend sind alle männlichen Formulierungen im Text geschlechtsneutral zu verstehen.

AI im Überblick



Quelle: In Anlehnung an axians, Schwachstellen im Netzwerk: Welche Rolle spielt künstliche Intelligenz in Cyber-Security-Konzepten?

Artificial Intelligence (AI) definiert das Oxford Dictionary als „Die Theorie und Entwicklung von Computersystemen, die in der Lage sind, Aufgaben zu erfüllen, die normalerweise menschliche Intelligenz erfordern, wie visuelle Wahrnehmung, Spracherkennung, Entscheidungsfindung und Übersetzung zwischen Sprachen." Maschinelles Lernen ist ein Teilgebiet von AI. Neuronale Netze sind wiederum einen Untermenge von ML-Verfahren. Von Deep Learning spricht man, wenn das Lernen durch mehrschichtige künstliche neuronale Netze ermöglicht wird. Diese Lernform hat in den vergangenen Jahren große Fortschritte ermöglicht – etwa bei der Bilderkennung, Spracherkennung (Speech Recognition) oder der Analyse von natürlicher Sprache (Natural Language Processing).

1. MANAGEMENT SUMMARY

Der Einsatz von Artificial Intelligence (AI), insbesondere des Maschinellen Lernens (ML), eröffnet große Innovationsmöglichkeiten für zahlreiche Anwendungsfelder und bildet einen Schlüsselbereich für die erfolgreiche digitale Transformation.[1)] Im Bereich der Cybersicherheit ist die Nutzung von ML bereits heute gängige Praxis. Sie kommt etwa bei Malware- und Spam-Bekämpfung sowie bei Network Intrusion Detection zum Einsatz. Technische Fortschritte der AI können maßgeblich zur Verbesserung der Cybersicherheit beitragen.

Allerdings verwenden auch Angreifer AI für ihre Zwecke. Zudem steht die Forschung vor der großen Herausforderung, Aussagen über die Qualität von AI-Verfahren zu treffen, was für ihren breiten Einsatz in der Cybersicherheit (und in anderen Bereichen) ein Hindernis darstellt. Ungeklärte rechtliche und ethische Fragen verlangsamen ebenfalls die nutzbringende Einführung von AI in der Cybersicherheit.

Um offene Fragen mit dringendem Handlungsbedarf zum Themenfeld AI, Cybersicherheit und Privatsphärenschutz in Wirtschaft und Gesellschaft zu identifizieren, veranstaltete das Fraunhofer-Institut für Sichere Informationstechnologie SIT am 12. August 2019 die „Eberbacher Gespräche zu AI, Security & Privacy". Im Rahmen des Workshops formulierten die Teilnehmer aus Industrie und Forschung zehn Empfehlungen (siehe Tabelle auf Seite 7), die aufzeigen, welche Herausforderungen bezüglich AI, Cybersicherheit und Privatsphärenschutz vordringlich angegangen werden müssen. Diese Empfehlungen adressieren die vier wichtigen Herausforderungen, welche sich aktuell für die Anwendung von Machine Learning im Kontext von Cybersicherheit und Privatsphärenschutz stellen:

❶ **Mindeststandards und Qualitätskriterien:** Sie befördern das Grundvertrauen in AI, erhöhen die Transparenz und dadurch auch die Akzeptanz von AI in Wirtschaft und Gesellschaft. Standards und offene Schnittstellen sorgen zugleich auch für eine Kontrollierbarkeit von AI, weil sie sich im Notfall bei Bedarf auch abschalten lässt. Über die technischen Standards hinaus braucht es für eine weitere Verwendung von AI in der Cybersicherheit auch Richtlinien für Ethik und Datenschutz. Diese gilt es zu diskutieren, um das Verhältnis von Mensch und Maschine wertekonform zu gestalten.

❷ **Rechtssicherheit und Klarheit in Haftungsfragen:** AI-Systeme lernen durch Trainingsdaten, die oft Datenschutzfragen aufwerfen.[2)] In vielen Bereichen sind diese Fragen bislang nicht befriedigend zu beantworten. Unternehmen brauchen jedoch eine verlässliche juristische Einschätzung, um das wirtschaftliche Potenzial von AI-Systemen erschließen zu können. Deshalb sollte der Gesetzgeber zum einen auf nationaler und europäischer Ebene Anreize schaffen, AI-Systeme datenschutzfreundlich zu gestalten. Zum anderen gilt es, Best-Practice-Empfehlungen zu entwickeln, die Herstellern und Anwendern ausreichende Orientierung bieten, wie dies heute zum Beispiel im Bereich der Kryptografie bereits Praxis ist. Für das verbleibende Risiko empfiehlt sich das Konzept der Gefährdungshaftung, das ebenfalls durch den Gesetzgeber auszugestalten ist.

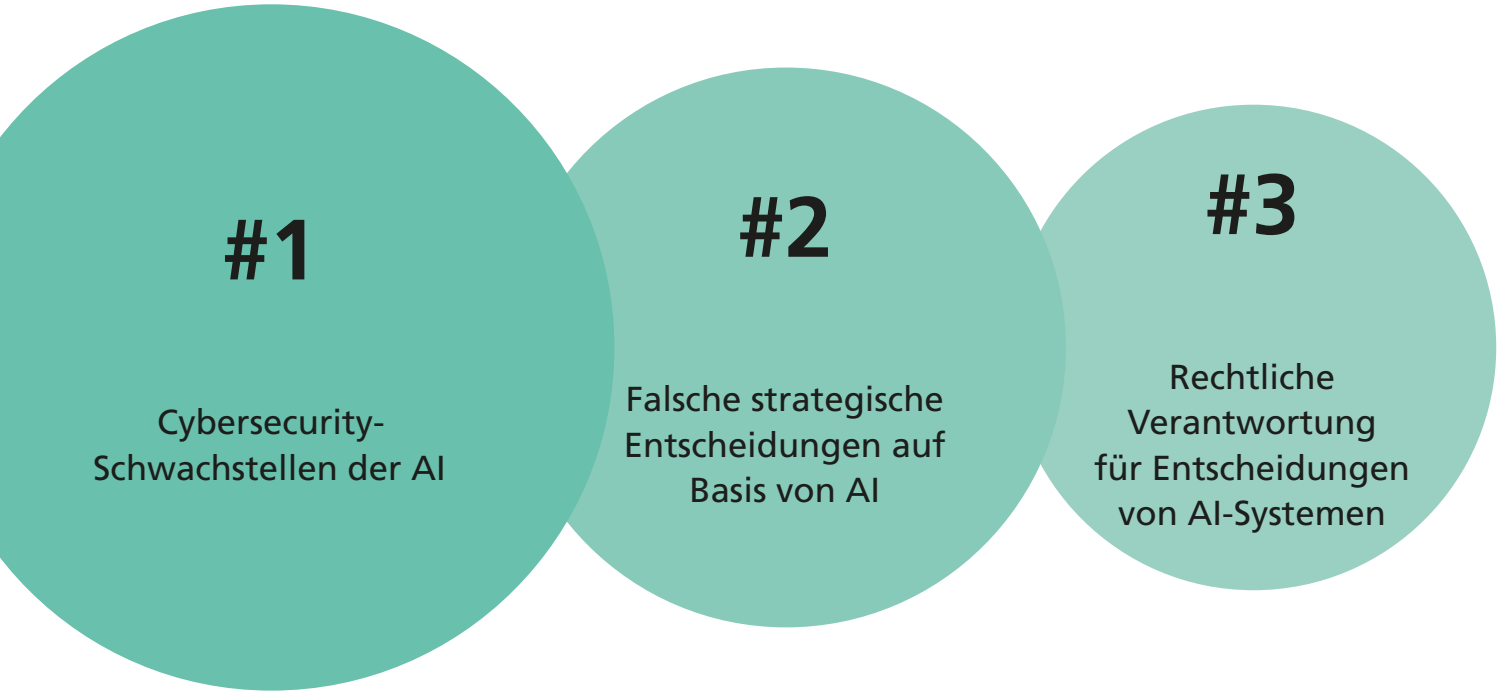
❸ **AI als Angriffswerkzeug**: Die wachsende IT-Durchdringung unseres Alltags schafft immer neue Angriffsziele für eine Vielzahl von Agressoren – staatliche Angreifer ebenso wie das organisierte Verbrechen. Die Angreifer nutzen AI für ihre Zwecke und verfügen über eine Vielzahl von Angriffswerkzeugen, deren Zahl in Zukunft noch zunehmen dürfte. Diese Risiken müssen betrachtet und soweit möglich minimiert werden. Für Letzeres ist der verantwortungsvolle Umgang mit Algorithmen und Daten besonders wichtig, der zum Beispiel durch einen Code of Conduct geregelt werden könnte, wobei dies länder- und kontinentübergreifend erfolgen muss.

❹ **AI-Fachkräfte mit Expertise für Cybersicherheit und Datenschutz:** Ohne einen Zuwachs an fachkundigem Personal werden sich die oben beschriebenen Herausforderungen nicht bewältigen lassen. Hierzu braucht es neben Informatikern auch Philosophen, Juristen, Ingenieure, Betriebswirte und viele weitere Experten, die AI ausreichend verstehen und anwenden können. Für die Cybersicherheit der nächsten Generation muss die AI-Kompetenz deshalb in allen Disziplinen und Sektoren auf- und ausgebaut werden.

Über die konkreten Empfehlungen hinaus zeigte das Eberbacher Gespräch die Wechselwirkungen zwischen Cybersicherheit, AI und Datenschutz. Diese können in der gesellschaftlichen und wirtschaftlichen Anwendung nicht getrennt betrachtet werden.

Cybersecurity ist größte Sorge

Top-3-Risiken für den AI-Einsatz in Unternehmen

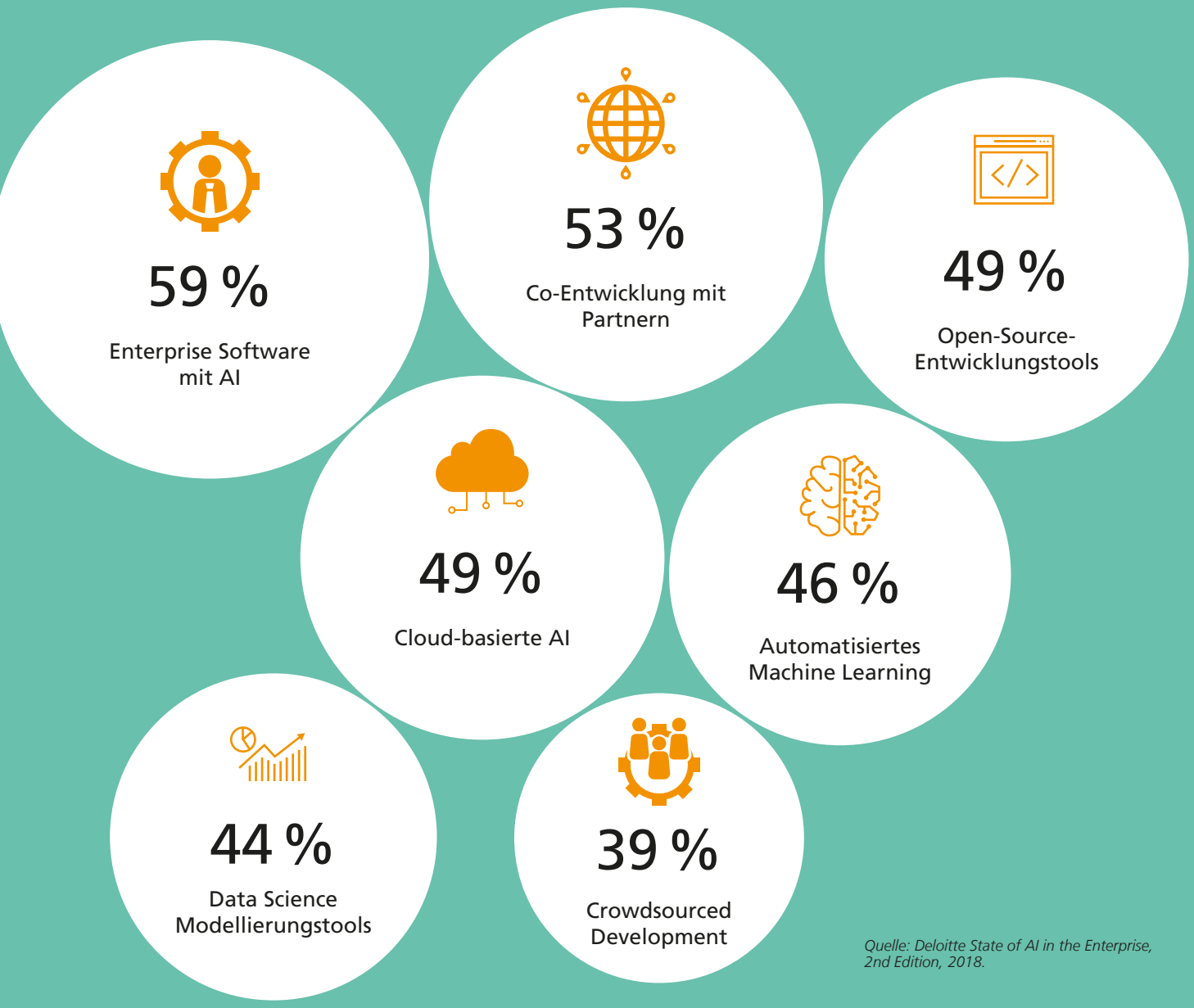


Quelle: Deloitte State of AI in the Enterprise, 2nd Edition, 2018.

2. ZUSAMMENFASSUNG DER EMPFEHLUNGEN

Herausforderung	Lösungsvorschlag	Adressat
Testierbarkeit und Mindestqualitätsniveau von AI für die Cybersicherheit	Automatisierte Prüfpunkte & Benchmarks, verifizierbar durch unabhängige Prüfeinrichtungen	Industrieverbände, Gesetzgeber (D, EU)
Schaffung von Transparenz und Akzeptanz für den AI-Einsatz in der Cybersicherheit	Standardisierung von AI-Verfahren durch unabhängige Stellen, offene Schnittstellen	Industrieverbände, Gesetzgeber (D, EU)
Kontrollierbarkeit des AI-Einsatzes in der Cybersicherheit	Modularer Aufbau, Abschaltbarkeit	Industrieverbände, Forschungseinrichtungen und Forschungsförderung
Ethischer AI-Einsatz in der Cybersicherheit	Zivilgesellschaftlicher Diskurs mit dem Ziel: Guidelines mit konkreter Prüfbarkeit	Zivilgesellschaftliche Organisationen, Medien, Forschungseinrichtungen und Forschungsförderung
Rechtssicherer Einsatz von Verfahren der Datenanonymisierung in AI-Systemen	In D und EU Rechtsklarheit in Zusammenarbeit mit der Technik schaffen; Anreize schaffen, Daten zu anonymisieren	Gesetzgeber (D und EU)
Unklarheit der Verantwortlichkeit und Haftung beim Einsatz von AI in der Cybersicherheit und für den Privatsphärenschutz	Gefährdungshaftung	Gesetzgeber (D und EU)
Neuartige Angriffsmöglichkeit: Manipulation von ML-Trainingsdaten	Embedded Trust: von der Erhebung über die Kommunikation bis zur Speicherung, Code of Conduct für ML-Einsatz	Industrieverbände: Zusammenschluss, z. B. „Charta of Trust"
Unterschiedliche Aggressoren, die diverse Methoden verwenden, darunter auch AI: Angriffswerkzeuge sind frei verfügbar, große Bandbreite der Ziele (Bsp. IoT)	Risikomanagement, insbesondere Priorisierung von Aggressoren: wer kann wo angreifen, wo kann schadensabhängig eingegriffen werden	Industrie D und EU, Politik D und EU
Nationalstaatliches Handeln ist nicht ausreichend bei AI-basierten Angriffen	Regulierung auf EU-Ebene und darüber hinaus treiben	Industrie D und EU, Politik D und EU
Nicht hinreichende Qualifikation in IT und speziell in AI	Fortbildung	Politik, Industrie, Schulungsanbieter

Business-Anwendungen beliebtester Einstieg in AI



Quelle: Deloitte State of AI in the Enterprise, 2nd Edition, 2018.

Beim **Maschinellen Lernen** erkennen Systeme nach und nach anhand von Trainingsdaten Muster und Gesetzmäßigkeiten. Beim **Supervised Learning** sind diese Daten bereits vorklassifiziert, zum Beispiel bestimmten Kategorien zugeordnet. Im Falle der SPAM-Erkennung sind E-Mails etwa schon als SPAM bzw. als NO-SPAM vorklassifiziert und das System lernt die impliziten SPAM-Kriterien anzuwenden. Beim **Unsupervised Learning** hingegen entdeckt das System selbst Muster, ohne dass diese vordefiniert wurden. Im Falle einer Angriffserkennung erkennt das System zum Beispiel, daß ein Netzwerk sich außerhalb des Normalzustands befindet, ohne daß genau definiert ist, wie der Normalzustand beschaffen ist.

3. SICHERHEIT DURCH AI – EINE ÜBERSICHT

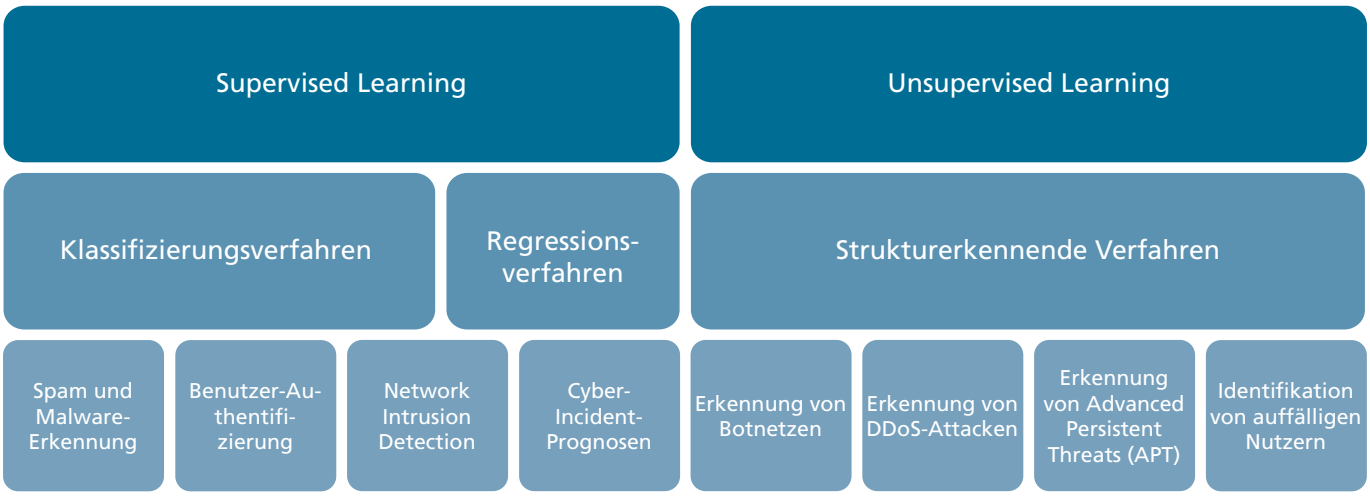
Die Begriffe „Artificial Intelligence“ (AI), „Maschinelles Lernen“ (ML) und „Big Data“ verschwimmen in der öffentlichen Diskussion häufig. ML als eine Ausprägung der Künstlichen Intelligenz wird heute üblicherweise eingesetzt, wenn von AI gesprochen wird. ML zielt darauf ab, ausgewählte Prozesse zu automatisieren, versucht aber nicht, menschliche Intelligenz in ihrer Gesamtheit zu simulieren. Wenn im Folgenden von AI gesprochen wird, handelt es sich im engeren Sinne um ML.

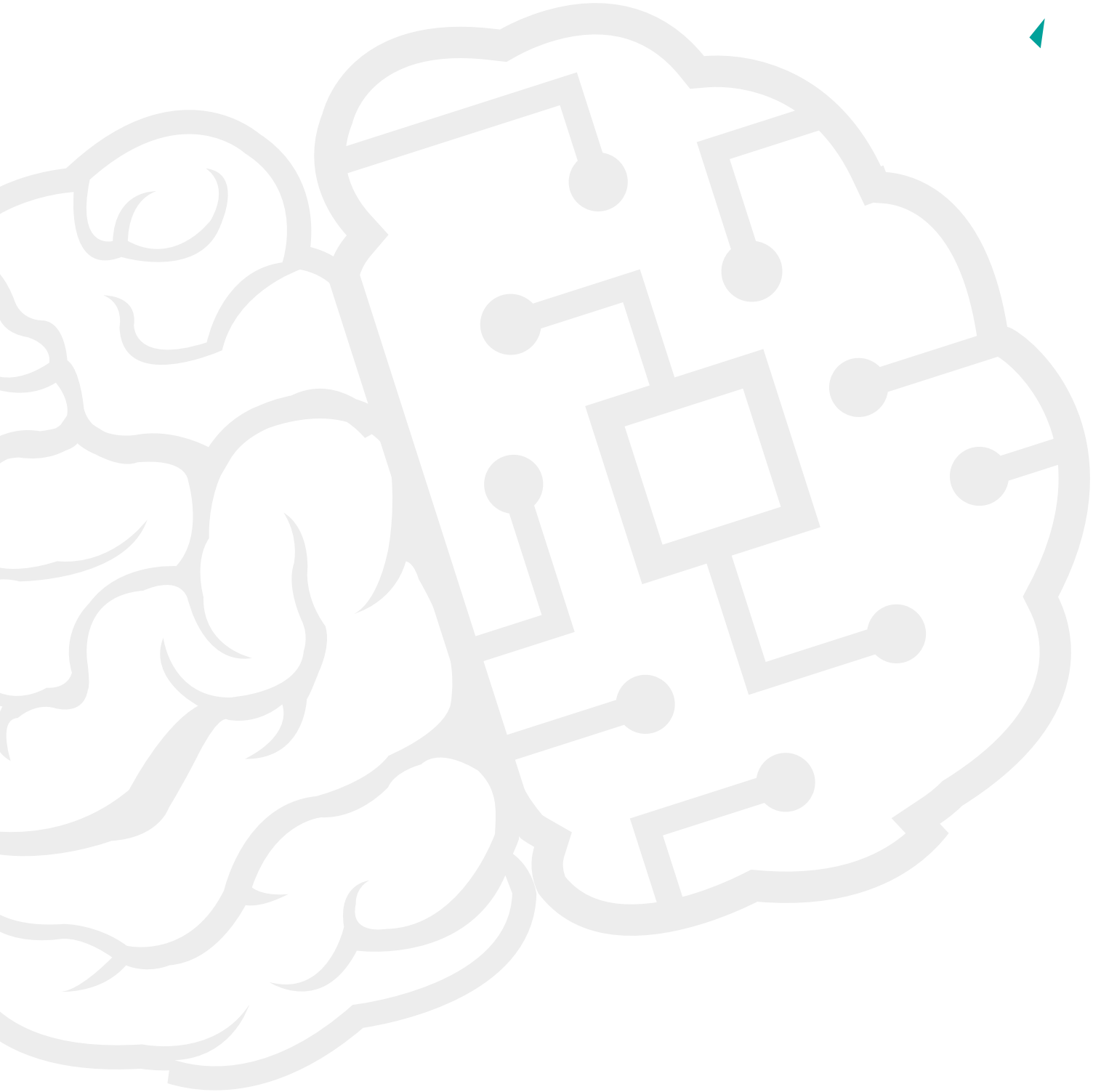
„Big Data“ beschreibt Systeme, in denen Daten, oft aus verschiedenen Quellen und heterogener Natur, verarbeitet und analysiert werden. Dabei unterstützt ML oft bei der Analyse. Auf der anderen Seite sind Big-Data-Systeme eine Quelle von großen Datenmengen, die in ML für das Erlernen von Entscheidungsprozessen und später auch für die Entscheidungen selbst eine wichtige Rolle spielen.

3.1 Einsatz von ML im engeren IT-Sicherheitskontext: Strukturierung

Spricht man von ML und IT-Sicherheit, ist die am meisten verbreitete Perspektive die, dass ein Computer bei der Erkennung und Bekämpfung von IT-Sicherheitsvorfällen helfen soll. ML kann bei der IT-Sicherheit besonders dann helfen, wenn die zu untersuchenden Daten umfangreich und unstrukturiert sind. Hier versagen Ansätze, die auf starren Regeln basieren. Mit ML können Softwaresysteme anhand von Beispielen lernen, Angriffe auf IT-Netze zu erkennen und den Normalfall davon zu unterscheiden. ML kann aber auch dabei helfen, große Datenmengen initial zu strukturieren, um Menschen einen schnellen Überblick über die aktuelle Sicherheitslage eines Systems zu verschaffen.

In der IT-Sicherheit ist der Einsatz von ML bereits heute in vielen Bereichen etabliert (siehe Grafik unten). Die SPAM-Erkennung ist ein bekanntes Beispiel dafür, dass Software mithilfe von ML erwünschte E-Mails von unerwünschten E-Mails unterscheiden kann.





3.2 Maschinelles Lernen zum Schutz von IT-Systemen: Lage- und Angriffserkennung

Lagebilderstellung und Angriffserkennung sind herausragende Beispiele für den nutzbringenden Einsatz von ML zum Schutz von IT-Systemen. ML erlaubt deutliche Verbesserungen der Erkennungsrate und Effizienzgewinne. Aus den großen Datenmengen, die beim Betrieb einer IT-Infrastruktur anfallen, können mittels ML schnell wertvolle Informationen gewonnen werden. ML automatisiert die Erkennung von Sicherheitsvorfällen teilweise und ermöglicht den verantwortlichen Sicherheitsexperten, die mit dem Schutz der Infrastruktur betraut sind, einen effizienten Überblick über den Status des Systems zu erlangen.

Die Menge der dabei betrachteten Daten ist beträchtlich. In heutigen Unternehmensnetzen werden über 100 Milliarden Events verursacht, über 100.000 Events sind davon mitunter sicherheitsrelevant. Ein Event kann dabei unterschiedlich komplex sein. Dies ist eine Aufgabe, die von einem Menschen so nicht mehr leistbar ist.

Intrusion Detection

Intrusion-Detection-Systeme (IDS) sollen Angriffe in Netzwerken selbstständig erkennen oder zumindest fachkundigen Nutzern ein Erkennen erleichtern. Sie basieren entweder auf festen Regeln, die sie anwenden, sodass das System in vorher definierten Fällen Alarm auslöst. Oder die Systeme erkennen, wenn das Systemverhalten vom Normalfall abweicht. Letzteres ist unter dem Begriff „Anomalie-Detektion“ bekannt. Zur Erkennung solcher Anomalien eignet sich sowohl überwachtes als auch unüberwachtes ML (s. Kasten 2). Beim unüberwachten ML werden mittels Clustering Aktivitäten erkannt, die stark vom Normalfall abweichen. Beim überwachten ML werden beispielsweise Daten von Angriffen als Trainingsdaten verwendet. Beim Erkennen von ähnlichen Ereignissen schlägt das System dann Alarm. Beide ML-Formen haben dabei bestimmte Nachteile: Systeme, die überwachtes ML nutzen, erkennen bekannte Angriffsmuster sehr gut. Neue Angriffsformen, die unbekannte Sicherheitslücken (Zero DayExploits) nutzen, lassen sich mit überwachtem ML nicht gut erkennen. Unüberwachtes ML ist zwar in der Lage, Zero Day Exploits zu erkennen, verursacht aber oft Fehlalarme.

Security Information and Event Management

Ein typischer Anwendungsfall für ML und Big Data ist Security Information and Event Management (SIEM). SIEM-Lösungen streben an, möglichst umfassende Informationen über die Systeme, die sie überwachen, zu aggregieren und mit ihnen Sicherheitsvorfälle zu entdecken und zu bearbeiten. Diese Informationen sind insbesondere Logdaten verschiedener Systemkomponenten. Durch ML-Verfahren können SIEM-Systeme nicht nur auf strukturierte Daten wie Logdaten zugreifen, sondern auch unstrukturierte Meldungen berücksichtigen. So lassen sich etwa auch heterogene Daten wie Absturzmeldungen einzelner Anwendungen oder E-Mail-Hinweise von Mitarbeitern für die Situationsanalyse nutzen. Typisch für Big-Data-Systeme ist auch, dass sie einen Überblick über die Datenlage in Echtzeit ermöglichen.

Managed Security

Werden Vorfälle in einem Netzwerk nicht nur auf Basis von Daten des Netzes bewertet, sondern auch anhand von Informationen aus aktuellen Angriffskampagnen oder dem Überwachen von Internetforen, wächst die Komplexität der Systeme. Entsprechende Lösungen werden von Unternehmen angeboten, die sich auf die Detektion auch komplexer Angriffe spezialisiert haben. ML spielt dann auch eine Rolle beim Erkennen von Risiken in der Kommunikation zwischen potenziellen Angreifern, die sich beispielsweise in Foren austauschen. Hier ist Natural Language Processing (NLP) eine etablierte Methode, die wiederum verschiedene Formen des ML verwendet.

Die Methoden der Angriffserkennung sind umso wirksamer, je mehr Informationen über Angriffe und Kampagnen zur Verfügung stehen. Dementsprechend können unternehmens- und behördenübergreifende Lösungen unter Umständen bessere Ergebnisse erzielen. Ob solche organisationsübergreifenden Systeme Angriffe wirklich besser erkennen, ist noch nicht erwiesen. Darüber hinaus ist noch nicht geklärt, ob und falls ja welche zusätzlichen Risiken übergreifende Systeme beinhalten und wie diesen Risiken wirksam begegnet werden kann. Kryptografisch gesicherte neutrale Datenräume könnten eine vielversprechende Lösung darstellen.

Wo Cybersecurity-Bedenken AI-Nutzung behindert

Branchen, in denen sich der AI-Einsatz aufgrund von Cybersecurity verzögert

Quelle: Capgemini Research Institute, AI in Cybersecurity executive survey, N = 850 executives

3.3 Ergebnisse des Maschinellen Lernens: weder 0 noch 1

Eine fundamentale Eigenschaft von ML ist, dass die Ergebnisse immer eine Wahrscheinlichkeit darstellen. Diese Wahrscheinlichkeit beschreibt die Übereinstimmung, die das trainierte System zwischen den vorliegenden Daten und den Trainingsdaten festellen kann. Dabei besteht immer die Chance, dass sich das System irrt und zum Beispiel einen Angriff zu erkennen glaubt, wo keiner ist. Je nach Methode und Datengrundlage sind hier Fehlerwahrscheinlichkeiten im Bereich von einem Promille bis hin zu 20 Prozent und mehr zu beobachten.[3)] Ein Fehler kann darin bestehen, einen Fehlalarm auszulösen oder fälschlicherweise keinen Alarm auszulösen – je nachdem, ob ein Schwellenwert über- oder unterschritten wird. Dabei gilt: Ein sensibles System, das möglichst alle Angriffe erkennen möchte, wird zu einem niedrigen Schwellenwert tendieren und dadurch eine größere Anzahl von Fehlalarmen verursachen.

Das Maß, in dem solche Fehler akzeptabel sind, ist stark abhängig von der Anwendung. Während im Marketing eine Fehlerrate von 20 Prozent immer noch ein erfolgreiches Instrument beschreiben kann, ist dies in einer Sicherheitsanwendung eventuell ein Ausschlusskriterium: Ist jeder fünfte Empfänger eines Werbeschreibens nicht an einem Produkt interessiert, kann die Kampagne durchaus erfolgreich sein. Wird jede fünfte Transaktion eines Kreditinstituts als Betrugsversuch angesehen und gestoppt oder jede fünfte Internetverbindung unterbunden, da die AI einen Angriff vermutet, ist dies für die Betroffenen unzumutbar. Sind entsprechende Fehlerraten unvermeidbar, taugt eine AI nicht als autonome Entscheidungsinstanz, kann jedoch als Vorfilter dienen, der einem menschlichen Entscheider die Arbeit wesentlich erleichtert.

Durch die Einführung von AI-Methoden gibt es große Fortschritte in der Cybersicherheit, allerdings wird das volle Potenzial von AI für Cybersicherheit bei Weitem noch nicht ausgeschöpft. Zudem sind auch AI-Systeme angreifbar, und AI kann selbst als Angriffswerkszeug genutzt werden.[4)] Dies wurde im Eberbacher Gespräch thematisiert: Von den Teilnehmern wurden als vordringliche Handlungsfelder die fehlenden Mindestanforderungen und Qualitätskriterien von AI, offene Fragen bezüglich Datenschutz und AI, die Bedrohung durch AI als Angriffswerkzeug sowie die geeignete Weiterbildung zur Vermeidung des Fachkräftemangels identifiziert.

Bewertungskriterien für das Qualitätsniveau von AI

Möglichkeiten zur Bewertung von Algorithmen

Kriterienlisten?

Input vs. Output?
(Benchmarks)

Explainable AI
(Reverse Engineering)

Qualität besser als Mensch
(Stichproben)

4. EMPFEHLUNGEN: AI, CYBERSICHERHEIT UND PRIVATSPHÄRENSCHUTZ

4.1 Mindestanforderungen und Qualitätskriterien von AI

1. Empfehlung: Testierbarkeit und Mindestqualitätsniveau von AI für die Cybersicherheit

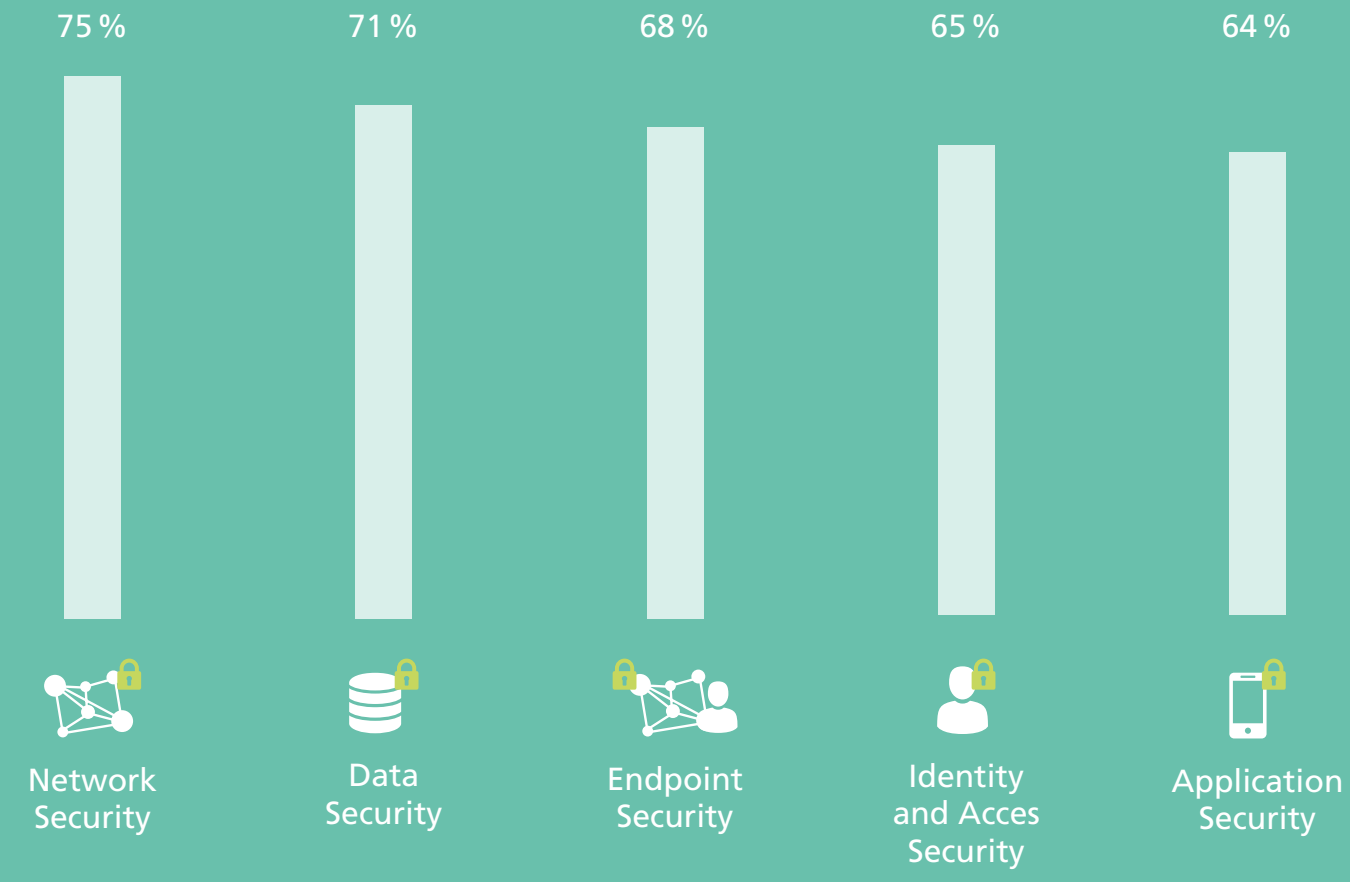
Bei ML-Systemen können die bemerkenswerten Leistungen unter Laborbedingungen, beispielsweise in der Angriffserkennung, erheblich von denen in realen Umgebungen abweichen und je nach Implementierung Ergebnisse von unterschiedlicher Qualität erzeugen. Vielfach lässt sich auch nicht präzise sagen, wie das ML zu einem spezifischen Ergebnis kommt bzw. auf welcher Grundlage entsprechende Ergebnisse erzielt werden. Zur Beurteilung einer AI-Leistung benötigt man die Fehlerraten des Systems (False Acceptance und False Positives) sowie Aussagen über die nötigen bzw. verwendeten Trainingsdaten. Aktuell ist oft unklar, welche Ergebnisse AI-Systeme erzielen. Für den Einsatz von AI für die Cybersicherheit braucht es allerdings begründetes Vertrauen, dass Entscheidungen tatsächlich zu einer Erhöhung des Sicherheitsniveaus führen und ganz sicher nicht, auch nicht im Einzelfall, zu einer Verringerung. Dort wo die Überlegenheit von AI-Systemen eindeutig nachgewiesen ist, wird sich deren Akzeptanz in Unternehmen und Gesellschaft vergrößern.

Für ein begründetes Vertrauen in AI-Systeme könnte eine wissenschaftlich fundierte Metrik die Grundlage bilden. Mit einer solchen Metrik ließe sich die Qualität von Systemen besser beurteilen und die Eignung von AI-Systemen für bestimmte Anwendungen feststellen (Eignungstreue). So könnten Mindestanforderungen an die Qualität von AI-Systemen entstehen und die Nachvollziehbarkeit von Enscheidungen (Verifizierbarkeit) sichergestellt werden. Zu diesem Zweck könnten automatisierte Prüfpunkte bzw. Benchmarks definiert werden, die durch unabhängige Prüfeinrichtungen verifizierbar sein müssen. Da diese Lösung zu einer weiteren Verbreitung hilfreicher Technologie beitragen würde, regen die Eberbach-Teilnehmer an, dass einschlägige Industrie- und Branchenverbände im Eigeninteresse Industriestandards für Testierbarkeit und Mindestqualitäten von AI in der Cybersicherheit etablieren. Auch die öffentliche Hand in Deutschland und in der EU hat ein hohes Interesse am Schutz eigener Systeme und der Absicherung von kritischen Infrastrukturen. Die öffentliche Hand könnte zum Beispiel staatliche Prüfeinrichtungen hierfür unterhalten bzw. existierende staatliche Prüfeinrichtungen mit diesen zusätzlichen Testieraufgaben betrauen.

Herausforderung	Testierbarkeit und Mindestqualitätsniveau von AI für die Cybersicherheit
Lösungsvorschlag	Automatisierte Prüfpunkte & Benchmarks, verifizierbar durch unabhängige Prüfeinrichtung(en)
Adressat	Verbände, Gesetzgeber (D, EU)

Netzwerksicherheit führt bei AI-Nutzung im Bereich Cybersecurity

Nutzung von AI nach Security-Anwendungsbereichen



Quelle: Capgemini Research Institute, AI in Cybersecurity executive survey, N = 850 executives

2. Empfehlung: Schaffung von Transparenz, Vertrauen und Akzeptanz für den AI-Einsatz in der Cybersicherheit

Aktuell wird mit dem Einsatz von AI in der Cybersicherheit von vielen Herstellern geworben. Vielfach bleibt allerdings unklar, was sich hinter dem Schlagwort AI beim jeweiligen Produkt verbirgt. In den Medien wird wiederum mitunter einseitig über Risiken beim AI-Einsatz berichtet. Auf diese Weise kann eine ambivalente oder sogar polarisierende öffentliche Wahrnehmung entstehen, in der sich Befürworter und Gegner zunehmend unversöhnlich gegenüberstehen und eine sachliche Diskussion erschwert wird. Zur Erhöhung der Akzeptanz für den AI-Einsatz in der Cybersicherheit ist es daher dringend notwendig, Verfahren, Schnittstellen und Trainingsdaten durch unabhängige Stellen zu standardisieren und so eine Transparenz hinsichtlich Algorithmen und Datenqualität zu schaffen, die eine sachliche Auseinandersetzung unterstützt. Die Standardisierung sollte parallel durch staatliche Stellen im Generellen und durch die Industrie, beispielsweise branchenspezifisch, erfolgen.

Herausforderung	Schaffung von Transparenz und Akzeptanz für den AI-Einsatz in der Cybersicherheit
Lösungsvorschlag	Standardisierung von AI-Verfahren durch unabhängige Stellen, offene Schnittstellen
Adressat	Industrieverbände, Gesetzgeber (D, EU)

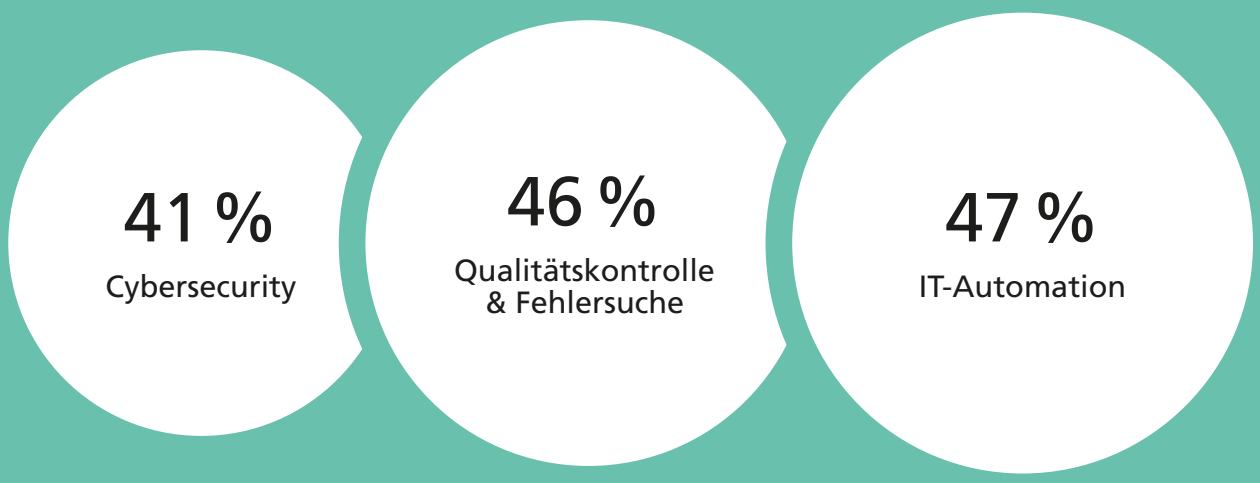
3. Empfehlung: Kontrollierbarkeit des AI-Einsatzes in der Cybersicherheit

Jeder Technologieeinsatz, der potenziell Schaden verursachen oder zulassen könnte, muss kontrollierbar sein, um unerwünschte Folgen ausschließen zu können und Risiken beherrschbar zu machen. Warum bestimmte ML-Systeme entsprechende Ergebnisse generieren, z.B. Prognosen von Deep Neural Networks, ist häufig jedoch kaum nachvollziehbar. Entsprechend stark widmet sich derzeit die Forschung diesem Aspekt. Solange die Wirkweise von AI-Systemen aber noch nicht ausreichend nachvollziehbar ist, muss das Management von Unternehmen in der Lage sein, die AI in Notfällen abzuschalten. Solche Notfall-Mechanismen können natürlich nur funktionieren, wenn das Unternehmen auf alternative Methoden und Systeme zurückgreifen kann. Für solche Fälle wäre etwa eine feingranulare Abschaltbarkeit von spezifischen AI-Modulen denkbar. Wird ein AI-Modul deaktiviert, könnte dann auf klassische Algorithmen umgeschaltet werden. Im Eigeninteresse sollte die Industrie über ihre Verbände auf solche Möglichkeiten hinwirken. Die Forschung kann hierzu wichtige Beiträge leisten, die es durch einen schnellen Wissenstransfer in Wirtschaft und Industrie zu nutzen gilt.

Herausforderung	Kontrollierbarkeit des AI-Einsatzes in der Cybersicherheit
Lösungsvorschlag	Modularer Aufbau, Abschaltbarkeit
Adressat	Industrieverbände, Forschungseinrichtungen und Forschungsförderung

Die beliebtesten AI Use Cases in der IT

Anteil der Befragten, deren Unternehmen sich auf diese Anwendungsszenarien konzentrieren



Quelle: Deloitte State of AI in the Enterprise, 2nd Edition, 2018

4. Empfehlung: Ethischer AI-Einsatz in der Cybersicherheit und für den Privatsphärenschutz

Durch den Einsatz von AI in der Cybersicherheit ergeben sich verschiedene ethische Fragestellungen, insbesondere stellt sich die Frage, ob eine AI ohne menschliche Prüfung Maßnahmen ergreifen kann. Wenn ein Rechner mittels AI in neun von zehn Fällen schneller und besser auf einen Cyberangriff reagiert als ein Mensch, werden viele Unternehmen der AI vermutlich gern den Vorrang vor menschlichen Entscheidungen einräumen. In einem Krankenhaus, in dem eine AI-Entscheidung ohne menschliche Prüfung zu einer Verzögerung oder Verschlechterung in der medizinischen Versorgung führen kann, lässt sich die Frage schon nicht mehr so einfach beantworten. Vollautomatische Entscheidungen sowie die Festlegung von Entscheidungsschwellenwerten sind nur zwei Aspekte ethischer Abwägungen, die am besten in einem gesellschaftlichen Diskurs ausgehandelt werden. In dieser notwendigen Diskussion sollten neben Gesetzgeber, Medien, Forschungseinrichtungen und Forschungsförderung auch zivilgesellschaftliche Organisationen eingebunden werden. Ein vielleicht fernes, aber absolut lohnenswertes Ziel für diese Diskussion sind ethische Richtlinien mit konkreten Kriterien und Prüfmöglichkeiten.

Herausforderung	Ethischer AI-Einsatz in der Cybersicherheit
Lösungsvorschlag	Zivilgesellschaftlicher Diskurs mit dem Ziel: Guidelines mit konkreter Prüfbarkeit
Adressat	Zivilgesellschaftliche Organisationen, Gesetzgeber, Medien, Forschungseinrichtungen und Forschungsförderung

4.2 Rechtssicherheit und Haftung beim Einsatz von AI

Die Frage der Rechtssicherheit beim Einsatz von AI stellte sich für die Teilnehmer des Eberbacher Gesprächs am drängendsten, wenn es um die Anonymisierung von personenbezogenen Daten und um Haftungsfragen geht.

5. Empfehlung: Rechtssicherer Einsatz von Verfahren der Datenanonymisierung in AI-Systemen

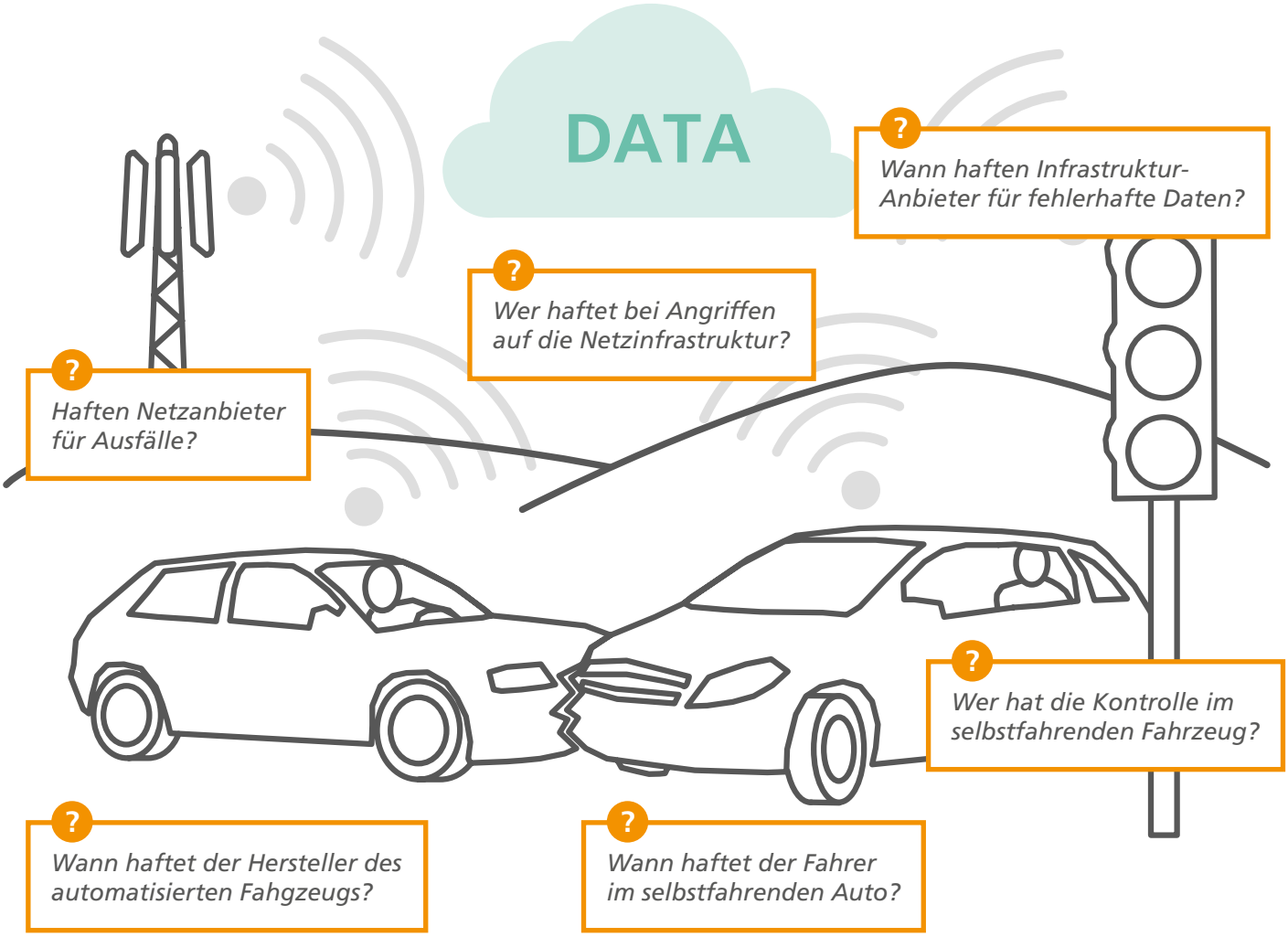
Datenschutz und Big Data werden oft als unvereinbar angesehen – ohne dass dies wirklich geprüft worden ist.[5)] Bis heute fehlt etwa eine belastbare Untersuchung dazu, welche Wirkung technische Datenschutzmechanismen auf Big-Data-Analysen haben. Unzweifelhaft ist hingegen, dass Big Data und ML eine Herausforderung für die Wirksamkeit des Datenschutzes darstellen. So lassen sich mit entsprechenden Analysen etwa Verknüpfungen zwischen Daten herstellen, die zu einem Aufheben einer vermeintlichen Anonymität führen können. Weiterhin besteht bei trainierten Netzen des ML das Risiko, dass die Trainingsdaten rekonstruiert werden können und so Bezüge zu Personen möglich sind. Das ist besonders dann eine Herausforderung, wenn das Training einer AI personenbezogene Daten erfordert, für diese Daten aber noch keine Methoden zur zuverlässigen Anonymisierung bekannt sind.

Noch ist also unklar, wie der Datenschutz beim Einsatz von AI-Verfahren nachweislich eingehalten und wie Privacy-Schutz in der Zukunft aufrechterhalten werden kann. Denn Fortschritte in der AI-Forschung können dafür sorgen, aktuell funktionierenden Datenschutz technisch aufzuheben, beispielsweise durch eine nachträgliche Verknüpfung von rückgewonnenen Trainingsdaten des ML. Dass dies keine theoretische Gefahr ist, zeigen aktuelle Angriffe auf AI-Systeme, die darauf abzielen, personenbezogene Daten aus einem trainierten neuronalen Netz zu extrahieren.Um den Datenschutz zu wahren und gleichzeitig Herstellern und Anwendern von AI-Systemen Rechtssicherheit zu verschaffen, braucht es Best Practice-Empfehlungen, die festlegen, was nach dem Stand der Technik an Schutzmaßnahmen zu verwenden ist, um personenbezogene Daten zu verschleiern (Pseudonymisierung) oder zu anonymisieren.

Herausforderung	Rechtssicherer Einsatz von Verfahren der Datenanonymisierung in AI-Systemen
Lösungsvorschlag	In D und EU Rechtsklarheit schaffen in Zusammenarbeit mit der Technik; Anreize schaffen, Daten zu anonymisieren
Adressat	Gesetzgeber (D und EU)

Haftungsfragen bei selbstfahrenden Fahrzeugen

Fehlerquellen und Beteiligte



6. Empfehlung: Klärung der Haftungsfrage

Unklare Haftungsregelungen bremsen Unternehmen und Behörden dabei, das volle AI-Potenzial zur Erhöhung der Cybersicherheit zu erschließen. Wenn Einrichtungen das Haftungsrisiko nicht abschätzen können, verzichten sie auf einen ansonsten plausiblen Einsatz.

Solange die AI-Mechanismen wiederum nicht vollständig erklärbar sind, sind sie als nicht-deterministisch anzusehen. Dadurch gibt es beim Einsatz solcher Systeme kein Ursache-Wirkung-Prinzip. In der Haftungsfrage greift hier am ehesten ein verschuldensunabhängiger Mechanismus. Die Teilnehmer des Eberbacher Gesprächs plädieren aus diesem Grund für die Einführung einer Gefährdungshaftung beim AI-Einsatz durch den Gesetzgeber, die unabhängig von schuldfähigem Handeln wie Vorsatz oder Fahrlässigkeit ist. Bei der Gefährdungshaftung erlaubt der Gesetzgeber die Nutzung einer Technologie auch dann, wenn gewisse Schäden nicht ausgeschlossen werden können. Beispiel Auto: Jeder Fahrzeughalter trägt das Risiko, dass er beim Betrieb des Autos andere Verkehrsteilnehmer verletzt. Allerdings begrenzt der Gesetzgeber bei der Gefährdungshaftung auch die Haftungspflichten – zum Beispiel für den Fall der höheren Gewalt. Ähnliches könnte auch für den AI-Einsatz geltend gemacht werden, bei dem ein schuldfähiges Verhalten nur schwer nachweisbar sein dürfte. Die entsprechenden Haftungsregelungen und Haftungsausschlüsse sind entsprechend durch den Gesetzgeber in Deutschland und der EU genauer auszuarbeiten.[6)]

Herausforderung	Unklarheit der Verantwortlichkeit und Haftung beim Einsatz von AI in der Cybersicherheit und für den Privatsphärenschutz
Lösungsvorschlag	Gefährdungshaftung
Adressat	Gesetzgeber (D und EU)

4.3 AI als Angriffswerkzeug

ML wird inzwischen auch von Angreifern als Werkzeug eingesetzt und ist somit auch ein eigenständiges Risiko für die IT-Sicherheit von Wirtschaft und Gesellschaft. Adversarial machine learning ist ein Ansatz, bei dem zwei AI-Systeme zusammengeschaltet werden und die zweite AI die erste angreift. So können sehr schnell Schwachstellen in der ersten AI, dem Verteidiger, entdeckt werden. Eine AI kann auch dazu eingesetzt werden, Angriffe auf IT-Systeme oder Nutzer zu automatisieren, beispielsweise, um sehr gut gezielte Spear-Phishing-Angriffe auf Personen anhand von Social-Media-Profilen durchzuführen.

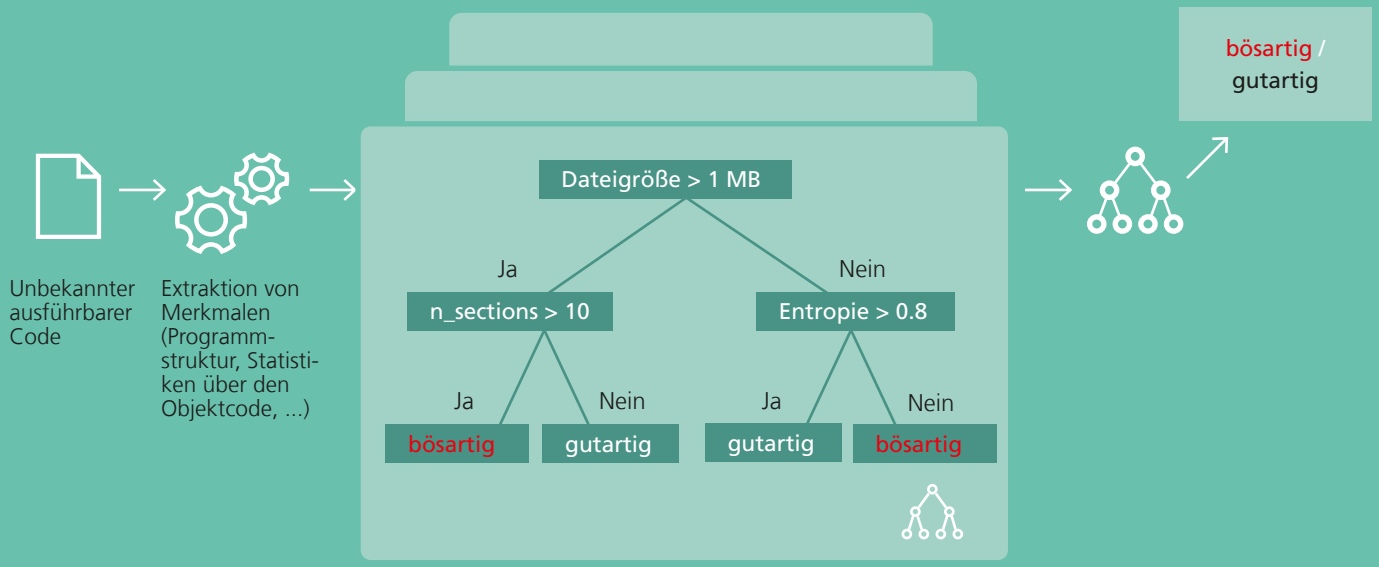
7. Empfehlung: Code of Conduc“ für sichere ML

Es wird eine Richtschnur für den Umgang mit AI benötigt, Ansätze für ein entsprechendes Regelwerk sind vorhanden, etwa die Hambacher Erklärung[7)] oder die Empfehlungen der Datenethik-Kommission beim Bundesinnenministerium[8)]. Diese gilt es auszubauen bzw. umzusetzen[9)] zu einem tragfähigen Code of Conduct für sicheres ML. Der Code of Conduct muss kodifizierbar, überprüfbar und verifizierbar sein, d. h. am Ende muss eine Güteaussage stehen, dass die Daten valide sind. Embedded Trust ist hier die Lösung, die über gesamte die Strecke – von der Erhebung über die Kommunikation bis zur Speicherung – wirksam ist und von der Industrie im Eigeninteresse verfolgt werden muss. Hierzu sollte sich die Industrie in Verbänden und anderen Zusammenschlüssen verpflichten.

Herausforderung	Neuartige Angriffsmöglichkeiten, z. B. Manipulation von ML-Trainingsdaten
Lösungsvorschlag	Embedded Trust: von der Erhebung über die Kommunikation bis zur Speicherung, Code of Conduct für ML-Einsatz
Adressat	Industrieverbände: Zusammenschluss , z. B. „Charta of Trust”

Machine Learning in Cybersecurity

Kaskade mehrerer Entscheidungsbäume zur Malware-Erkennung



Quelle: Kaspersky

Entscheidungsbäume werden zur Effizienzsteigerung bei der Malware-Erkennung eingesetzt. Ziel ist es, mittels lernender Systeme die Präzision der Erkennungsleistung stetig zu erhöhen und mutierte sowie sogar neue Malware erkennen zu können. Jeder Entscheidungsbaum wird fortwährend auf Basis von Daten bekannter Malware und gutartigen Dateien trainiert. Für jede neu zu prüfende ausführbare Datei werden Entscheidungsbäume eingesetzt, die mit geringem Rechenaufwand eine Vorauswahl treffen. Die meisten Dateien werden hierbei erwartungsgemäß als gutartig befunden. Die verbleibenden Dateien können teilweise direkt als Malware identifiziert werden, alle weiteren Verdachtsfälle werden weiteren, ressourcenintensiveren Entscheidungsbäumen zugeführt.

8. Empfehlung: Risikomanagement

Auch Angreifer können AI und ML für ihre Zwecke nutzen: Diverse Methoden und Angriffswerkzeuge stehen Angreifern zur Verfügung und die Bandbreite der Ziele ist groß – von kritischen Infrastrukturen bis hin zum Internet der Dinge. Da es sehr unterschiedliche Aggressoren gibt, neben Wirtschaftskriminellen zum Beispiel auch staatliche Angreifer, müssen die Risiken entsprechend Angriffszielen und Wirkung priorisiert werden: Wer kann wo angreifen? Wo sind hohe Schäden zu erwarten? Gefordert ist klassisches Risikomanagement. Welche Risiken gibt es, wie hoch sind die Wahrscheinlichkeiten, wie groß der angenommene Schaden. Die Schadenshöhe könnte in Euro beziffert werden, allerdings wird dies schwierig bei Reputationsschäden für Unternehmen.

Herausforderung	Unterschiedliche Aggressoren, die diverse Methoden verwenden, darunter auch AI: Angriffswerkzeuge sind frei verfügbar, große Bandbreite der Ziele (Bsp. IoT)
Lösungsvorschlag	Risikomanagement, insbesondere Priorisierung von Aggressoren: Wer kann wo angreifen, wo kann schadensabhängig eingegriffen werden
Adressat	Industrie D+EU, Politik D+EU

9. Empfehlung: Internationale Lösungen

Da Cyberangriffe – ob mit oder ohne AI – ein globales Problem sind, muss die Thematik global angegangen werden. Beispielsweise kann eine ML nur mit weltweit verteilten Messpunkten eine sich über den Globus hinweg aufbauende Angriffswelle erkennen. Je mehr qualitativ hochwertige internationale Daten, desto besser. Für die Frage der internationalen Zusammenarbeit gilt es zunächst zu klären, ob inhärente Interessen von internationalen Playern Lösungsansätzen entgegenstehen. Wie können diese Hinderungsmotive ausgeräumt werden?

Herausforderung	Nationalstaatliches Handeln ist nicht hinreichend bei AI-basierten Angriffen
Lösungsvorschlag	Koordiniertes Handeln auf EU-Ebene und darüber hinaus
Adressat	Industrie D+EU, Politik D+EU

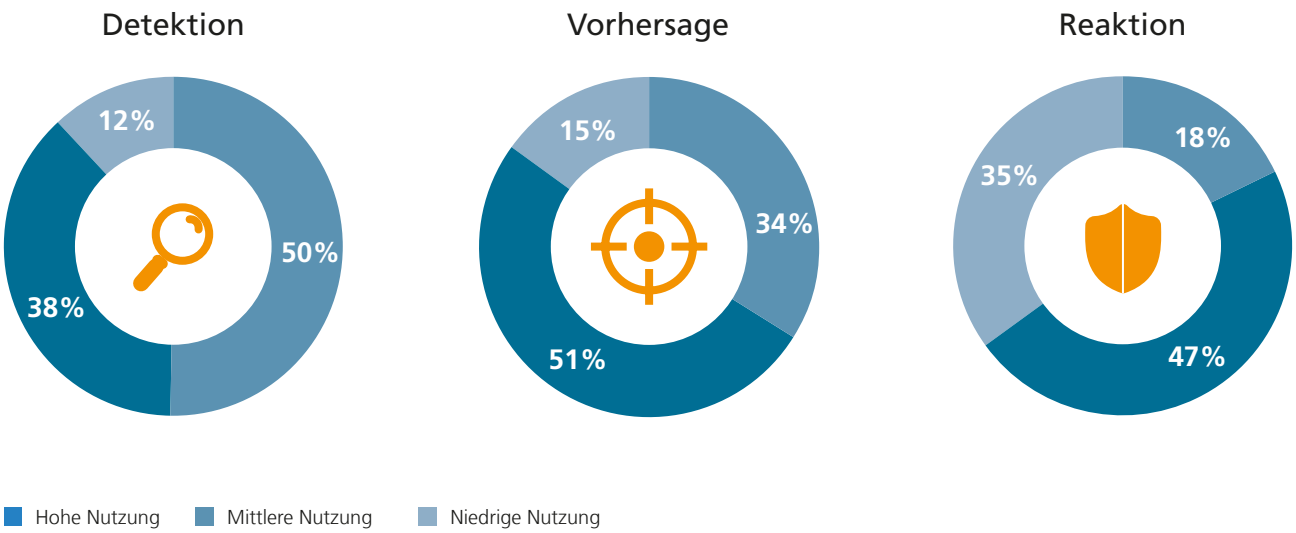
4.3 Fachkräfte AI, Cybersicherheit und Datenschutz

10. Empfehlung: Fortbildung

Für den Einsatz von AI in der Cybersicherheit ist die Investition in die Köpfe mindestens so wichtig wie die Investition in Technologie. Deutschland und Europa brauchen mehr Fachkräfte, die in diesen Bereichen qualifiziert sind. Am dringendsten werden aktuell Bildungs- und Weiterbildungsangebote benötigt, die Fachkräfte schnell und bedarfsspezifisch in der Thematik qualifizieren.

Herausforderung	Nicht hinreichende Qualifikation in IT und speziell in AI
Lösungsvorschlag	Fortbildung
Adressat	Politik, Industrie, Schulungsanbieter

AI-Nutzungsziele in der Cybersicherheit



Quelle: Capgemini Research Institute, AI in Cybersecurity executive survey, N = 850 executives

5. ZUSAMMENFASSUNG

Im Eberbacher Gespräch zu AI, Security & Privacy wurden zehn konkrete Handlungsempfehlungen für Wirtschaft und Politik entwickelt. Die Möglichkeit zur Umsetzung und damit zur konstruktiven Nutzung der Empfehlungen haben neben Politik und Gesetzgeber insbesondere die Industrieverbände. Während die Herstellung von Rechtssicherheit im Hinblick auf ausreichenden Datenschutz und Haftungsproblematik vordringliche Aufgabe des Gesetzgebers ist, können und sollten Politik und Wirtschaft gemeinsam die gesellschaftliche Akzeptanz für den AI-Einsatz erhöhen, indem sie Transparenz und notwendige Mindestanforderungen definieren und entsprechende Kontrollmöglichkeiten vorsehen. Forschung und Entwicklung können hierfür die technischen Voraussetzungen schaffen, beispielsweise Möglichkeiten zur Abschaltung von AI. Zur Klärung der ethischen Rahmenbedingungen hingegen bedarf es eines breiten gesellschaftlichen Diskurses, der bereits begonnen wurde, aber noch verstärkt werden sollte. Dies alles kann nach Meinung der Teilnehmer nur gelingen, wenn die Regulierungen international abgestimmt sind und die Gefahren durch neue Angriffsmöglichkeiten erfolgreich begrenzt werden können.

Jedes einzelne der drei im Titel angesprochenen Themen - AI, Cybersicherheit und Privatsphärenschutz - ist essenziell für eine erfolgreiche Digitalisierung von Wirtschaft und Gesellschaft. Gleichzeitig ist jedes einzelne Gebiet auch für alle Branchen als Querschnittsthema hochrelevant. Das Eberbacher Gespräch zeigte die vielfältigen Beziehungen zwischen den drei Themen auf. Von herausragender Wichtigkeit stellten sich dabei drei Abhängigkeitsbeziehungen dar:

1) „Cybersicherheit braucht AI“
2) „AI braucht Cybersicherheit“ und
3) „AI braucht Privatsphärenschutz“

Die Umsetzung der im Eberbacher Gespräch erarbeiteten zehn Empfehlungen brächte die drei Themenfelder und die drei Schlüsselbeziehungen signifikant voran und würde die Digitalisierung in Wirtschaft und Gesellschaft erheblich befördern.

Referenzen

1) Vgl. KI-Strategie der Bundesregierung https://www.ki-strategie-deutschland.de/home.html

2) Vgl. Martin Steinebach, Christian Winter, Oren Halvani, Marcel Schäfer und York Yannikos: *Big Data und Privatheit*. Darmstadt 2015. ISSN: 2192-8169. https://www.sit.fraunhofer.de/fileadmin/dokumente/studien_und_technical_reports/Big-Data-Studie2015_FraunhoferSIT.pdf

3) Lukas Graner, Oren Halvani, Verena Battis,Christian Winter, Inna Vogel, Martin Steinebach,York Yannikos: *Design-Unterstützung Big-Data-Analysen*. Fraunhofer-Institut für Sichere Informationstechnologie (SIT), Dezember 2019 (in Veröffentlichung).

4) Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff et aliter: *The Malicious Use of Artificial Intelligence:Forecasting, Prevention, and Mitigation*. Oxford 2018. https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/MaliciousUseofAI.pdf?ver=1553030594217

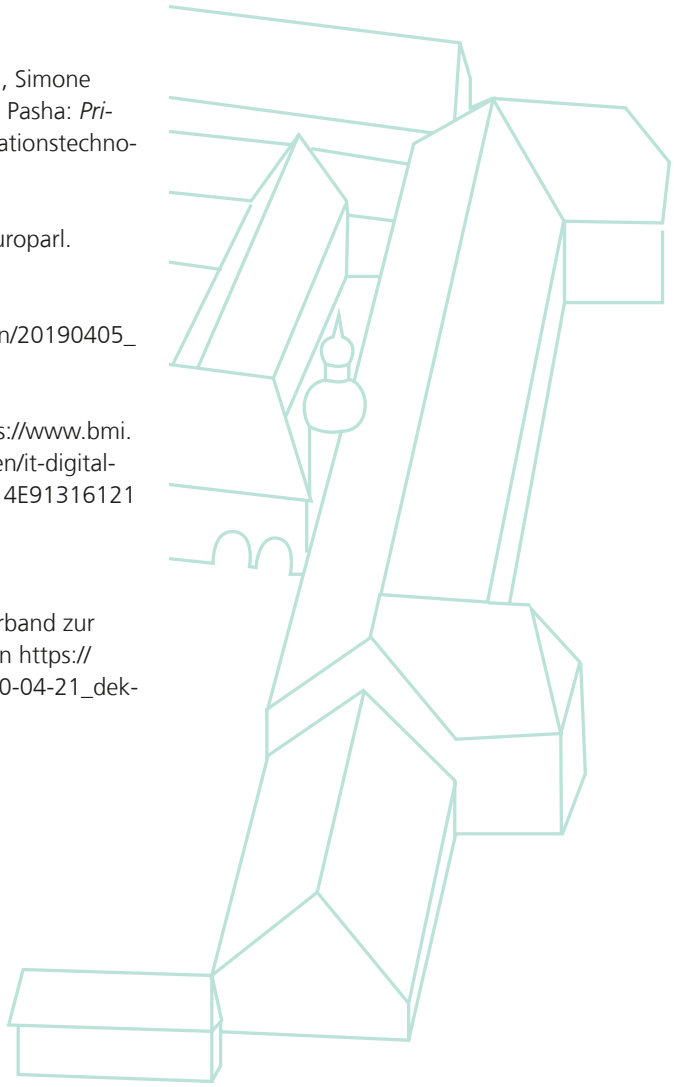
5) Christian Winter, Martin Steinebach,Wendy Heereman, Simone Steiner,Verena Battis, Oren Halvani, York Yannikos, Jamal Pasha: *Privacy und Big Data*. Fraunhofer-Institut für Sichere Informationstechnologie (SIT), Dezember 2019 (in Veröffentlichung).

6) Vgl. EU-Resolution zu Haftungsfragen: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.pdf

7) https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf

8) Gutachten der deutschen Datenethikkommission https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digital-politik/gutachten-datenethikkommission.pdf;jsessionid=14E9131612171FD37889639A5B377472.2_cid295?__blob=publicationFile&v=6

9) Monitoring-Bericht der Verbraucherzentrale Bundesverband zur Umsetzung der Empfehlungen der Datenethikkommission https://www.vzbv.de/sites/default/files/downloads/2020/04/29/20-04-21_dek-evaluierung_politikcheck_1.halbjahr_final2.pdf



Autoren

Dr. Michael Kreutzer
Oliver Küch
Prof. Dr. Martin Steinebach

Anschrift

Fraunhofer SIT
Rheinstraße 75
64295 Darmstadt
Deutschland
Telefon +49 6151 869-282
Fax +49 6151 869-224
redaktion@sit.fraunhofer.de