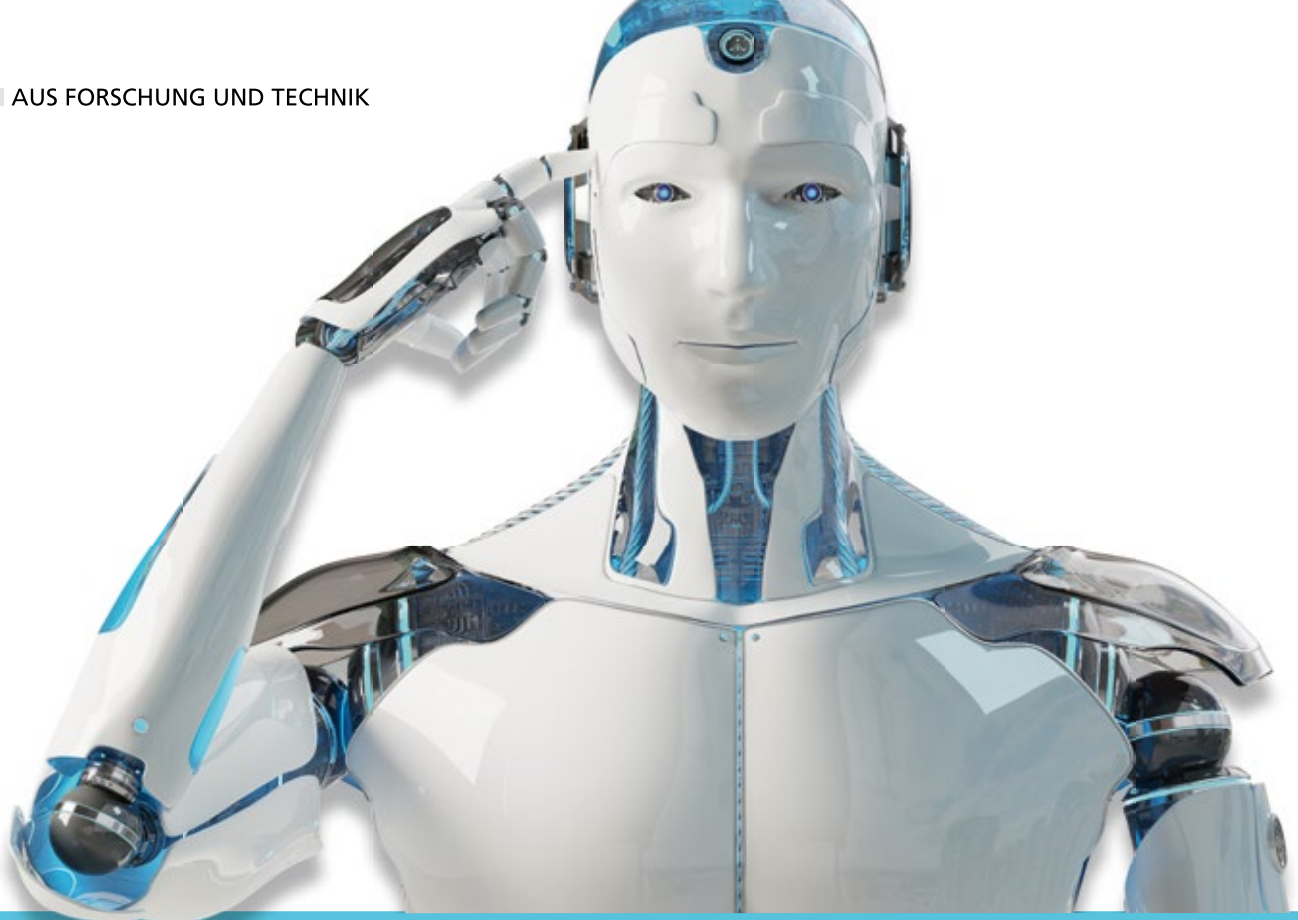

Künstliche Intelligenz und Cybersicherheit

Unausgegoren aber notwendig

Cybersicherheitssysteme, die künstliche Intelligenz (KI) berücksichtigen, werden in der Zukunft helfen, deutlich besser die intelligenten Hacker und deren Angriffe zu entdecken, Schäden zu vermeiden und Risiken im gewünschten Digitalisierungsprozess zu minimieren. Mithilfe von künstlicher Intelligenz kann die Erkennungsrate von Angriffen im Netzwerk und in ubiquitären IT-Endgeräten (Smartphone, Notebook, Server, IoT etc.) deutlich erhöht werden. Das bedeutet aber auch, dass IT-Systeme ohne KI-gestützte Verteidigung zur leichten Beute für Angreifer werden können, die ihrerseits vermehrt KI einsetzen. Somit hat künstliche Intelligenz vermehrt Auswirkungen auf die Cybersicherheitslage, die durch aktuelle Lagebilder aufzeigbar gemacht werden muss[1].

Eine große Herausforderung für die Verteidiger ist, für welche der sehr vielen erkannten sicherheitsrelevanten Ereignisse zusätzlich noch menschliche Analysten notwendig sind. Nicht alle Ereignisse können durch Cybersicherheitsexperten verarbeitet werden, da die Anzahl der Ereignisse die Verarbeitungsfähigkeit und Verarbeitungskapazitäten menschlicher Analysten an ihre Grenzen bringen. Diesen Umstand können Angreifer ausnutzen und die Verteidiger gezielt ablenken, um unbemerkt in das IT-System einzudringen. Künstliche Intelligenz kann dabei helfen, die Ereignisse in Echtzeit zu analysieren und situationsgerecht zu entscheiden, ob ein menschliches Eingreifen überhaupt noch notwendig ist. In anderen Einsatzszenarien, bei denen eine Teilautonomie technisch nicht möglich ist und der Mensch zwingend eingebunden werden muss, kann der Einsatz von KI die Aufgaben und Tätigkeiten des Menschen wesentlich unterstützen. Damit werden die wenigen vorhandenen Ressourcen gezielter eingesetzt und das Cybersicherheitsniveau insgesamt erhöht. Situationsgerecht bedeutet dabei, dass klassische Verfahren auf Basis von Signaturen nur noch unterstützend eingesetzt werden und neuartige, verhaltensbasierte Verfahren, wie fortgeschrittene Anomalie-Erkennung oder Predictive Analysis, Einzug halten.

Einordnung der künstlichen Intelligenz

Die Wissenschaft „Data Science", ein Fachgebiet der Informatik, beschäftigt sich mit der Extraktion von Wissen aus den Informationen in Daten. Da es immer mehr Daten mit Informationen gibt, kann auch immer mehr Wissen aus den Informationen der Daten abgeleitet werden, insbesondere auch im Bereich der Cybersicherheit (Bild 1). Dabei setzt „künstliche Intelligenz" intelligentes Verhalten in Algorithmen um, mit der Zielsetzung, automatisiert „menschenähnliche Intelligenz" so gut wie möglich nachzubilden.

Bei künstlichen Intelligenzen kann zwischen schwacher und starker KI unterschieden werden. Eine starke KI soll eine Intelligenz schaffen, die dem Menschen gleicht kommt oder sogar übertrifft, während die schwache KI sich in der Regel mit konkreten Anwendungsproblemen des menschlichen Denkens beschäftigt.

Maschinelles Lernen (Machine Learning/ML) ist ein Begriff im Bereich der künstli-

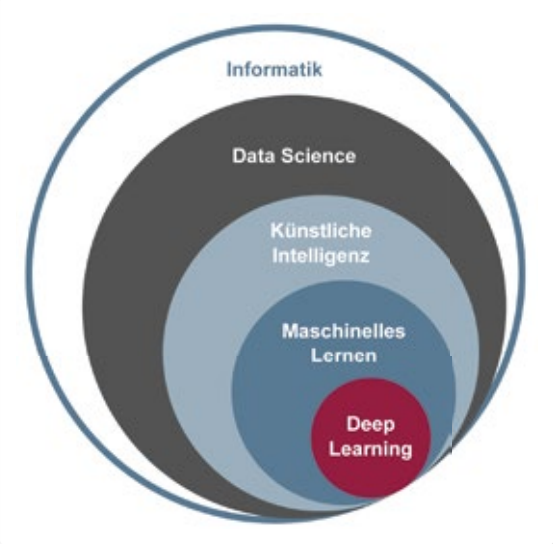

Bild 1: Einordnung der künstlichen Intelligenz

chen Intelligenz für die „künstliche" Generierung von Wissen aus den Informationen in Daten mit der Hilfe von IT-Systemen. Mithilfe der Algorithmen des maschinellen Lernens werden mit vorhandenen Datenbeständen Muster und Gesetzmäßigkeiten erkannt und verallgemeinert, um damit neue Problemlösungen umzusetzen. In Lernphasen gewinnen entsprechende ML-Algorithmen die Befähigung, aus vielen, diversen Beispielen simple Muster und Strukturen, hin zu komplexen Merkmalen und Gesetzmäßigkeiten zu erkennen. Daraus entstehende Regeln können auf neue Daten und ähnliche Situationen angewendet werden, in denen die KI beispielsweise entscheiden muss, ob es sich um einen Angriff oder eine legitime Nutzeraktion handelt.

Maschinelles Lernen wird noch effektiver durch Deep Learning. Deep Learning ist eine Spezialisierung des maschinellen Lernens und nutzt vorwiegend künstliche neuronale Netze (KNN).

Erfolgsfaktoren der künstlichen Intelligenz

Die Erfolgsfaktoren der künstlichen Intelligenz sind vielfältig. Die Entwicklung sowie die Zeit werden diesen Trend weiter fördern. Folgende Aspekte spielen eine Rolle:

1.) Leistungsfähigkeit der IT-Systeme

Die Leistung von IT-Systemen (CPU, RAM …) hat einen Stand erreicht, der die zentrale Speicherung und Verarbeitung von massenhaften Input-Daten sehr einfach macht. Viele umfangreiche Prozesse von maschinellem Lernen sind heute in akzeptabler Zeit durchführbar. Parallelisierung steigert diese Leistung nochmals deutlich. Hohe Geschwindigkeiten in der Datenübertragung erlauben ein Auslagern verschiedener Prozesse und Aufgaben auf weitere Server. Spezielle Software-Frameworks helfen, die Umsetzung zu optimieren.

Unterstützung gibt es außerdem von gut skalierbaren Ressourcen aus der Public Cloud, wie Amazon Web Services, Microsoft Azure, Google Cloud Platform und die IBM Cloud.

2.) Immer mehr verwertbare Daten

Neben der Entwicklung der technischen Leistungsfähigkeit sind die mit der fortschreitenden Digitalisierung stetig zunehmenden Daten ein relevanter Faktor für den Erfolg von KI. Auf der einen Seite ist die Quantität der verwertbaren Daten durch sehr viele Sensoren (in IT-Systemen, in Diensten, in Gebäuden, am Körper, im Auto …) rasant gestiegen. Auch der Austausch von Daten hat zugenommen.

Auf der anderen Seite ist die Qualität der Daten durch weitere Individualisierung der (persönlichen) Daten von IT-Systemen (PC, Notebook, Smartphone, Smartwatch, Automobile …) gestiegen. Außerdem werden zunehmend gezielt sicherheitsrelevante Informationen in Daten durch spezielle Sensoren im Cybersicherheitsbereich zur Verfügung gestellt.

3.) Immer bessere Algorithmen

Hinzu kommt die Verfügbarkeit sehr effizienter Algorithmen, die optimiertes Machine Learning ermöglichen. Der Gesamtablauf des ML lässt sich dabei durch einfache Methoden verbessern, zum Beispiel führen intelligent gewählte Eingangsdaten zu einer Reduktion der Komplexität. Es findet ein iterativer Lernprozess der Algorithmen statt. Algorithmen des maschinellen Lernens werden durch diese stetigen Verbesserungen praktisch umsetzbar gemacht und auch für komplexere Daten effizient. In der weiteren Entwicklung wird ML durch Deep Learning noch effektiver. Deep Learning ist eine Spezialisierung des maschinellen Lernens und nutzt vorwiegend komplexere künstliche neuronale Netze. Dabei werden zusammenhängende Schichten aus künstlichen Neuronen zur Datenverarbeitung genutzt. Das Potenzial von Deep Learning besteht darin, dass im Vergleich zu traditioneller KI nicht nur effektiver analysiert wird, sondern durch den effektiveren Lernprozess der KI auch mit unvollständigen Daten eine Analyse erfolgreich umgesetzt werden kann. So kann durch den ständigen Lernprozess des Deep Learnings eine KI in bis dahin unbekannten Situationen angewandt werden.

4.) Immer mehr Erfahrungen mit den Umgang

Durch die immer häufigere Nutzung von KI werden die Erfahrungen im Umgang mit Daten und Algorithmen immer größer.

5.) Immer einfacherer Zugang

Algorithmen werden immer zugänglicher und nutzbarer durch bessere Frameworks, Dokumentationen, Bibliotheken.

Das Prinzip des maschinellen Lernens

Die Algorithmen des maschinellen Lernens haben als Input Eingangsdaten mit Informationen, berechnen mit einem Algorithmus nach einem vorgegebenen Verfahren und liefern als Output die Ergebnisse. Die Anwendung entscheidet, wie die Ergebnisse verwendet werden sollen (Bild 2).

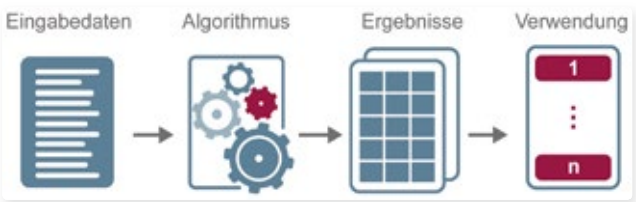

Bild 2: Workflow des Maschinellen Lernens

Eingangsdaten
Die Eingangsdaten können sehr vielfältig sein. Beispiele aus dem Cybersicherheitsbereich:
- Smartphone: Lage- und Beschleunigungssensoren, GPS-Daten, Nutzereingaben …
- Notebook, PC: Nutzereingaben, Log-Daten …
- Netzwerke: Bandbreite, Verzögerung …, Header- und Kommunikationsdaten …
- IoT: Sensorik- und Aktorik-Daten
- Allgemein IT-Systeme: CPU-Aktivitäten, RAM-Verbrauch, SW-Aufrufe, Kommunikation …

Algorithmen
- Support-Vector-Machine (SVM)
- k-Nearest-Neighbor (kNN)
- k-Means-Algorithmus
- hierarchische Clustering-Verfahren
- Convolutional Neural Network

und viele mehr.

Ergebnisse
Ergebnisse aus der Verarbeitung der Eingangsdaten mit den Algorithmen können sein:
- Klassifizierung der Eingangsdaten, wie Erkennung von Angriffen
- numerische Werte, wie Hinweise zur Verbesserung eines Produkts
- binäre Werte, wie eine erfolgreiche biometrischer Authentifizierung

Verwendung
Die Anwendung entscheidet, wie die Ergebnisse verwendet werden.

Kategorien und Algorithmen des maschinellen Lernens

1. Überwachtes Lernen
Beim überwachten Lernen wird ein Algorithmus mithilfe von Eingabedaten und Ergebnissen trainiert. Dadurch kann der Algorithmus lernen, ob das Ergebnis mit den Eingabedaten den Erwartungen entspricht. Zum Aufgabenfeld des überwachten Lernens gehört das Regressions- und Klassifizierungsproblem. Mit der Regressionsanalyse ist es möglich, Werte von abhängigen Variablen zu prognostizieren. Aufgaben der Klassifikation befassen sich damit, Daten in verschiedene Klassen mit ähnlichen Ausprägungen einzuteilen.

Ziele des überwachten Lernens sind:
- Regression: Vorhersagen von numerischen Werten
- Klassifizierung: Einteilung von Eingabedaten in Klassen

Beispiele im Bereich der Cybersicherheit sind:
- Erkennung von Spam-Mails
- Erkennen von Angriffen in Intrusion Dectection Systems (IDS)

ML-Algorithmen aus dem Bereich des überwachten Lernens sind zum Beispiel:
- Support-Vector-Machine (SVM)
- k-Nearest-Neighbor (kNN)

Support-Vector-Machine (SVM)

Eine Support Vector Machine ist ein mathematisches Verfahren zur Klassifizierung von Eingabedaten (Objekte). Eine SVM arbeitet mit Trainingsdaten, für die bereits definiert ist, welcher Klasse sie zugehören. Jedes Eingabedatum wird dabei durch einen Vektor in einem n-dimensionalen Vektorraum repräsentiert. Für diesen Vektorraum versucht die SVM eine optimale Hyperebene zu berechnen, um damit die Daten in zwei Klassen zu unterteilen. In Bild 3 ist exemplarisch dargestellt, wie in einem zweidimensionalen Raum nach einer optimalen Hyperebene zu den gegebenen Eingabedaten gesucht wird. In einem n-dimensionalen Raum hat die Hyperebene die Dimension n-1. Aus diesem Grund ist jede der betrachteten Hyperebenen in dem dargestellten Beispiel eine Linie.

Eine Hyperebene ist optimal, wenn der Abstand zu den sogenannten „Support Vectors" am höchsten ist. Ein „Support Vector" ist der nächste Vektor einer Klasse zu der betrachteten Hyperebene. In dem dargestellten Beispiel sind es jeweils die nächsten Punkte einer Klasse (rot oder blau) zu der betrachteten Linie.

Trainieren einer Support Vector Machine:
Eingabedaten: (1) in Bild 3
- klassifizierte Objekte (Trainingsdaten, für die bereits definiert ist, welcher Klasse sie zugehören)
- Abstandsmaß der Objekte untereinander (durch Beschreibung als Vektor)

ML-Algorithmus: (2) in Bild 3
- Ermitteln von Geraden zur Trennung der klassifizierten Objekte
- Bewertung durch Abstand zu den Punkten
- Wahl der Geraden mit maximalem Abstand zu beiden Klassen

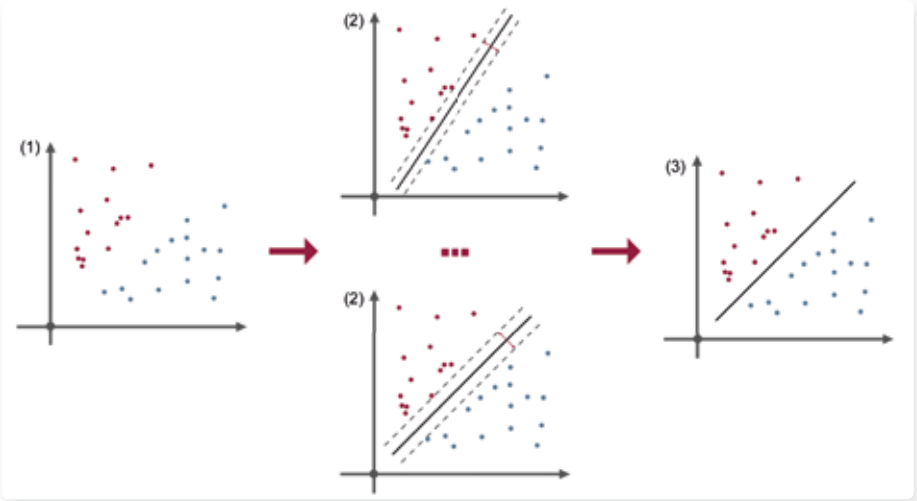

Bild 3: Support-Vector-Machine (SVM) – Training

Ergebnis: (3) in Bild 3
- Gerade als Modell zur Klassifizierung

Danach klassifiziert das Modell mithilfe der Lage der Punkte der Eingabewerte in eine Klasse.

k-Nearest-Neighbor (kNN)

Der k-Nearest-Neighbor-Algorithmus ist ein Klassifikationsverfahren, bei dem eine Klassenzuordnung auf Basis seiner k nächsten Nachbarn durchgeführt wird. Auch bei diesem Klassifikationsverfahren müssen bereits klassifizierte Objekte vorhanden sein. Die Klassifikation eines neuen Objektes erfolgt im einfachsten Fall durch Mehrheitsentscheidung. Für die Mehrheitsentscheidung werden die k nächsten bereits klassifizierten Objekte herangezogen. Als Maß für den Abstand der Objekte zueinander kann zum Beispiel die euklidische Distanz verwendet werden. Die euklidische Distanz beschreibt eine räumliche Distanz zwischen zwei Objekten und ist nachfolgend definiert:

$$dist(v,w) = \sqrt{(\sum(v_i - w_i)^2}, \text{ wobei } v,w \in R$$

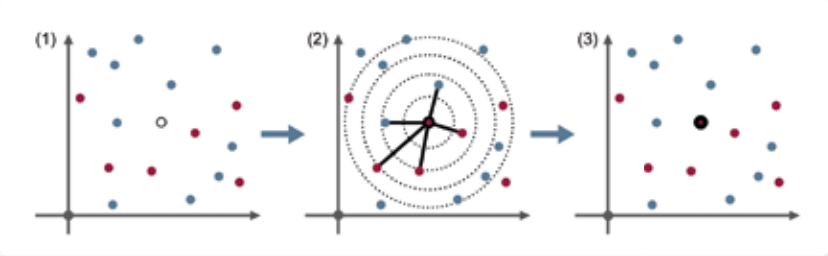

Bild 4: k-Nearest-Neighbor (kNN)

Eingabedaten: (1) in Bild 4
- bereits klassifizierte Objekte
- Anzahl der zu betrachtenden Nachbarobjekte k
- unklassifiziertes Objekt, das klassifiziert werden soll

ML-Algorithmus: (2) in Bild 4
- Berechnung der Distanz zu allen anderen Objekten
- Betrachtung der k nächsten Nachbarobjekte
- Zuordnung zur am häufigsten vorkommenden Klasse

Ergebnis: (3) in Bild 4
- Klassifizierung des neuen Objekts durch Mehrheitsentscheidung

2. Unüberwachtes Lernen

Beim unüberwachten Lernen werden Muster und Gesetzmäßigkeiten in unklassifizierten Objekten gesucht.

Die Stärke im unüberwachten Ansatz liegt darin, nach Mustern auch in unklassifizierten Daten zu suchen, um sie nach vorheriger Aufbereitung besser beschreiben zu können. Mittels Clustering werden ähnliche Datengruppen miteinander in Verbindung gesetzt. Die Erwartungshaltung an diesen Ansatz liegt unter anderem darin, Dinge zu erkennen, die vorher anderweitig nicht sichtbar waren. Dieser Ansatz ist im Weiteren auch gut geeignet, um unüberschaubare Datenmengen auf die wichtigsten Eigenschaften sowie Kriterien zu reduzieren. Da der Algorithmus selbstständig lernt, werden klassische Fehler im üblichen Sinne nicht produziert. Dies kann jedoch zu einem anderen Problem führen: Lernt der Algorithmus auch in die gewünschte Richtung? Zur Überprüfung des unüberwachten Lernens müssen folglich alle relevanten Gegebenheiten miteinander abgeglichen werden, um so Korrelationen zu finden.

ML-Algorithmus

Clustering setzt ähnliche Datengruppen miteinander in Verbindung
- k-Means-Algorithmus
- hierarchische Clustering-Verfahren

k-Means-Algorithmus

Der k-Means-Algorithmus ist ein Verfahren zur Clusteranalyse. Mit vorhandenen Eingangsdaten wird zufällig aus den gebildeten Mittelwerten für jeden Cluster ein Zentrum (Zentroid) ausgewählt. Die Elemente werden initial (zum Beispiel zufällig) zu den Clustern zugeordnet. Im nächsten Schritt werden die Abstände der einzelnen Punkte zum Beispiel mithilfe der euklidischen Distanz zu den Zentroiden neu berechnet. Dann werden die Elemente zu den am nächsten befindlichen Zentroid und seinem Cluster zugeordnet. Im nächsten Schritt werden die Zentroide erneut berechnet und die Elemente dementsprechend zugeordnet. Diese Schritte wiederholen sich iterativ so lange, bis kein Punkt mehr zu einem anderen Cluster zugeordnet werden kann. Der k-Means-Algorithmus ist einfach umzusetzen, er besteht im Prinzip nur aus Abstandsberechnungen und Neuzuordnungen und kommt iterativ zu einem stabilen und effektiven Cluster. Die Anzahl der gewünschten Cluster muss als Eingabewert (k) bestimmt werden.

Eingabedaten: (1) in Bild 5
- beliebige Objekte
- Abstandsmaß
- Anzahl k Cluster
- Initiale Zuordnung der Elemente zu Clustern (zum Beispiel zufällig)

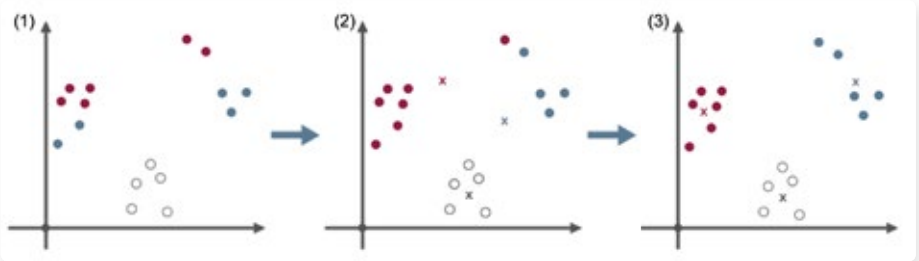

Bild 5: k-Means-Algorithmus

ML-Algorithmus: (2) in Bild 5
- Berechnung der Schwerpunkte (Zentroide)
- Zuordnung der Elemente zu Cluster mit dem nächsten Zentroid
- Neuberechnung der Zentroide und erneute Zuordnung

Ergebnis: (3) in Bild 5
- Einteilung der Objekte in k Cluster

Hierarchische Clustering-Verfahren

Bei hierarchischen Cluster-Verfahren entstehen geschachtelte Cluster, die wiederum aus Clustern entstehen. Hierbei werden zu Anfang viele kleine Cluster gebildet, die im weiteren Verlauf zur größeren Clustern zusammengeführt werden. Das Ergebnis wird in einem Dendrogramm dargestellt.

Jedes Objekt der Eingabedaten ist zu Beginn ein eigenes Cluster. Durch das gewählte Ähnlichkeitsmaß werden ähnliche Cluster zu einem größeren Cluster zusammengeführt. Die zusammengeführten Cluster werden wiederum als Eingabedaten verwendet und weiter zusammengeführt. So entsteht nach jeder Iteration eine hierarchische Struktur.

Eingabedaten: (1) in Bild 6
- beliebige Daten
- Ähnlichkeitsmaß

ML-Algorithmus: (1) bis (5) in Bildern 6 und 7
- jeder Datenpunkt ist ein eigenes Cluster
- ähnliche Cluster werden zuerst zusammengeführt
- entstandene Cluster werden erneut als Eingabedaten verwendet
- iteratives Zusammenführen der Cluster induziert eine hierarchische Struktur

Ergebnis: (6) in Bild 7
- hierarchische Beziehungen zueinander in Form eines Binärbaums (Dendrogramm)

Künstliche neuronale Netze (KNN)

Die Vorlage von künstlichen neuronalen Netzen (KNN) ist die biologische Struktur des Gehirns und seiner Neuronen. Dabei werden Gewichte, mathematische Funktionen und miteinander verbundene Schichten aus künstlichen Neuronen für die Informationsverarbeitung genutzt. Die Struktur eines künstlichen Neuronalen Netzes besteht aus einer Eingabeschicht, verdeckten Schichten und einer Ausgabeschicht. Die Schichten selbst bestehen wiederrum aus einer Vielzahl an künstlichen Neuronen.

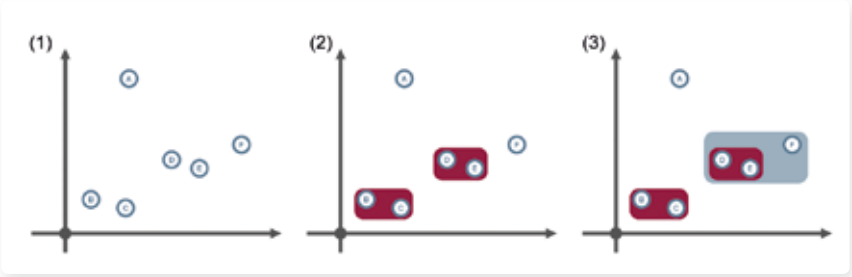

Bild 6: Hierarchische Clustering-Verfahren – Teil 1

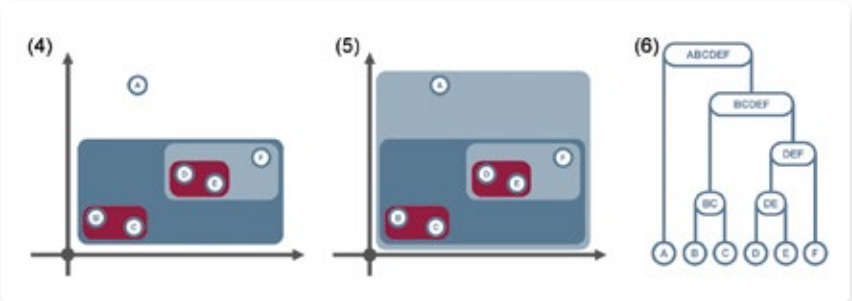

Bild 7: Hierarchische Clustering-Verfahren – Teil 2

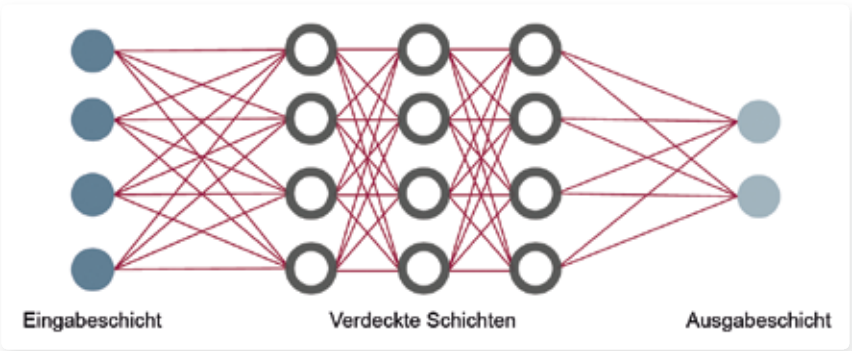

Bild 8: Künstliche neuronale Netze (KNN)

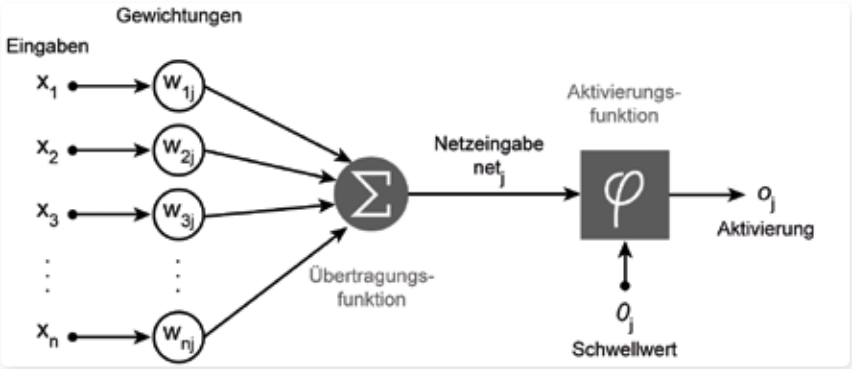

Bild 9: Künstliches Neuron

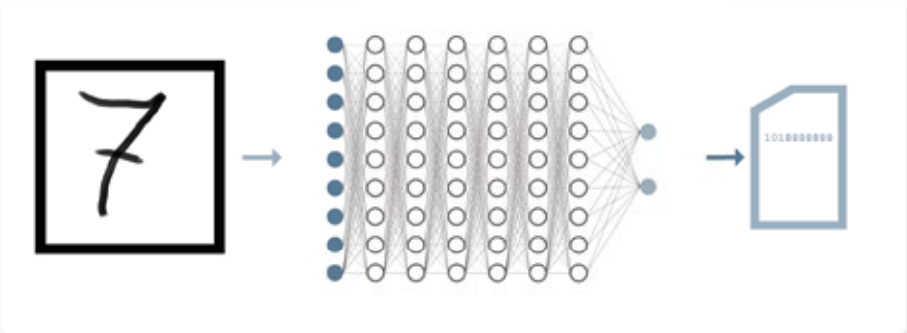

Bild 10: Beispiel Handschriftenerkennung

Die Eingabeschicht dient der Informationsaufnahme. Die Eingabedaten werden in verwendbare Repräsentationen transformiert. Je nach Komplexität und Problematik der Aufgabe kann ein KNN bis zu einer beträchtlichen Anzahl miteinander verknüpfter, verdeckter Schichten besitzen, welche mit jeder weiteren Schicht immer komplexere Merkmale und Strukturen herausfiltern soll. Am Ende gibt die Ausgabeschicht die Ergebnisse in sämtlichen möglichen Repräsentationen aus.

Bei einem künstlichen Neuron berechnet die Übertragungsfunktion anhand der Summe der Gewichtungen der Eingaben die Netzeingabe. Jedes Neuron hat einen individuellen Schwellenwert, der durch eine Schwellenfunktion berechnet wurde. Im Lernvorgang wird der Schwellenwert anhand der Eingaben stetig optimiert. Übersteigt die Netzeingabe den Schwellenwert, wird das Neuron aktiviert. Biologisch stellt der Schwellenwert die Reizschwelle dar, ab der das Neuron aktiviert wird. Als Aktivierungsfunktion kann zum Beispiel die Sigmoidfunktion verwendet werden, welche die Summe der Eingabe für die Ausgabe auf einen Wertebereich zwischen 0 und 1 abbildet. Im Gegensatz zum Schwellenwert, der bei jedem Neuron individuell sein kann, gilt die Aktivierungsfunktion für jedes Neuron.

Deep Learning

Die KI-Forschung hat in den letzten Jahren und Jahrzehnten verschiedene Architekturen hervorgebracht, die bestimmte Aufgabentypen besonders gut lösen. Komplexe Schichtenarchitekturen, wie Convulutional Neural Nets (CNN) oder Long Short-Term Memory Networks (LSTM), profitieren von enormen Datenmengen und haben daher bessere Ergebnisse im Big-Data-Bereich. Bei Deep Learning ist kein händisches Feature Engineering mehr notwendig, daher ist weniger Domänenwissen erforderlich.

Deep Learning am Beispiel Handschriftenerkennung:
Die Eingabe besteht aus einem Bild mit einer handschriftlichen Ziffer (Bild 10). Je nach der Größe des Bildes könnte ein Eingabeneuron genau ein Bildpixel repräsentieren. Bei diesem Vorgang werden die Helligkeitswerte der Pixel an die Eingabeneuronen übergeben. Die Helligkeitswerte aus der Eingabeschicht werden in den Neuronen der verdeckten Schichten weiterverarbeitet. Die Ausgabeschicht besteht aus zehn Neuronen, wobei ein Neuron jeweils eine Ziffer von 0 bis 9 repräsentiert. Die Ausgabe liefert als Ergebnis eine Tabelle mit den Ziffern und deren Wahrscheinlichkeit einer Übereinstimmung. Folglich ist aus der Tabelle zu entnehmen, dass es sich zu 85 Prozent um eine Sieben handeln könnte.

Eingangsdaten:
- Bilddatei mit einer Zahl, die klassifiziert werden soll

ML-Algorithmus (Deep Learning)
- Eingabedaten werden in den künstlichen Neuronen in den Schichten verarbeitet

Ergebnis:
- Tabelle mit einer Verteilung der Wahrscheinlichkeiten für eine Übereinstimmung mit einer Ziffer

Anwendungsszenarien von KI und Cybersicherheit

Im Folgenden werden einige ausgewählte Anwendungsszenarien von KI und Cybersicherheit aufgezeigt, um die Anwendungsvielfalt zu demonstrieren.

1.) Betrugsschutz im Online-Banking

Im Bereich des Online-Bankings kann zum Beispiel mithilfe von KI ermittelt werden, ob eine erhöhte Bedrohungslage herrscht. Dazu werden verschiedene Datenquellen herangezogen und beispielsweise ermittelt, wie viele Banking-Trojaner aktuell aktiv sind, ob es aktuell bekannte Software-Schwachstellen im Umfeld von Online-Banking gibt, die für einen Angriff auf Bankkunden verwendet werden könnten oder ob derzeit vermehrt versucht wird, mit Phishing-Mails Zugangsdaten zu Online-Konten abzugreifen. Diese und andere Indikatoren, wie identifizierte Betrugs- oder Betrugsversuchsfälle der Bank, können dann verwendet werden, um mit verschiedensten Algorithmen aus dem Bereich des maschinellen Lernens ein Bedrohungslagebild zu erstellen und den Bankkunden bei hoher Bedrohung zu warnen und entsprechend aufzuklären, um die Schäden zu verhindern.

2.) Erkennen von Angriffen über das Internet und Kommunikationslagebild

Durch die Analyse der Kommunikationsdaten können mithilfe von KI Angriffe über das Internet erkannt werden. Dadurch können die Kommunikationsmöglichkeiten entsprechend reduziert werden, um den Angriff abzuwehren. Die Reduzierung kann sich zum Beispiel auf einen bestimmten Port oder die ganze Internet-Kommunikation beziehen. Ob ein IT-Sicherheitsexperte bei der Entscheidung eingebunden wird oder das Cybersicherheitssystem dies automatisiert durchführt, ist ein wichtiger Aspekt für die Effektivität und Kosten des Systems. Die Ergebnisse können dann in ein Security Information and Event Management (SIEM)-System einfließen und zum besseren Management von Vorfällen führen. Zusätzlich kann auch ein Kommunikationslagebild erstellt werden, um Angriffe, Bedrohungen und Schwachstellen eines Netzwerks auszuwerten und Handlungsempfehlungen zu geben.

3.) Authentifikationsverfahren

Passive, kontinuierliche Authentifizierung ist besonders bei der zunehmenden Verbreitung mobiler Endgeräte ein Zukunftsfeld für KI-Algorithmen. Sensordaten aus Beschleunigungsmessgeräten oder Gyroskopen können während der Nutzung des Geräts erhoben und ausgewertet werden. Die KI kann folglich unberechtigte Nutzer von der Gerätenutzung ausschließen. Solche Authentifizierungsverfahren sind ein weiterer Schritt zur Usability von robusten und sicheren Cybersicherheitsmechanismen. Diese sind außerdem inklusiv, da sie keine zusätzliche Nutzerinteraktion erfordern und auch von Nutzern mit (beispielsweise kognitiven) Einschränkungen genutzt werden können. Neben der Analyse von Sensordaten ist auch eine verbesserte Authentifizierung anhand von Bild- oder Spracherkennung möglich, da die Hardware zum Aufnehmen in den Endgeräten vorhanden ist, und die Algorithmen zur Auswertung besser geworden sind.

4.) Malware-Erkennung

Die konventionelle Malware-Erkennung basiert zumeist auf signaturorientierten Detektoren, die bei einer Überprüfung die Signaturen von Dateien und Programmen mit bekannten Signaturen von Malware vergleicht. Wird Malware jedoch nur minimal verändert, kann die Signatur nicht mehr zur Erkennung genutzt werden. Heutige Malware verändert sich daher dynamisch. Dies hat zur Folge, dass immer neuere Varianten erscheinen und die Analyse und Aktualisierungen der Signatur-Datenbanken kaum noch effizient zu bewältigen ist. KI-basierte Detektoren können genutzt werden, um in Echtzeit verdächtige Aktivitäten zu erkennen. Anomalie-Erkennung oder Predictive Malware Analysis sind Verfahren, die durch den Einsatz von KI deutlich verbessert werden können.

5.) IT-Forensik

Im Bereich der IT-Forensik werden KI-Systeme ebenfalls ein relevanter Faktor. Durch die vermehrte Verlagerung von Lebensbereichen in die digitale Welt werden auch zunehmend Straftaten im digitalen Raum begangen, deren Spuren in den gewaltigen Datenmengen der alltäglichen Nutzung gefunden werden müssen. Dabei stoßen klassische Analysewerkzeuge immer schneller an ihre Grenzen, da IT-Systeme prinzipiell heterogener Natur sind. Verschiedenste IT-Geräte mit unterschiedlichen Betriebssystemen, Installationen und Konfigurationen können unzählige Fragmente aufweisen, die im Kontext von Ermittlungen vielfältige Relevanz besitzen. KI-Anwendungen können hier beispielsweise dabei helfen, zu entscheiden, ob bestimmte „Adressen" von einer verdächtigten Person kontaktiert wurden, oder ob es sich um Fragmente handelt, die von Software-Entwicklern standardmäßig in ihr Programm eingebunden wurden – wie es unter anderem bei Support-Adressen häufig der Fall ist.

Weitere Anwendungsszenarien

Weitere Anwendungsszenarien sind: sichere Softwareentwicklung, Erkennen von FakeNews, Bilderkennung von Ausweisen, VideoIdent, biometrische Verfahren wie Tippverhalten, Gestik-Erkennung, Seitenkanalanalyse, Kryptoanalyse, Advanced Persistent Threats & Cyber-Crime Threat Intelligence, identifizieren von Spam-Mails etc.

Manipulierbarkeit von künstlicher Intelligenz

In diesem Abschnitt wird diskutiert, wie und an welchen Stellen die künstliche Intelligenz mit ihren Algorithmen manipuliert werden kann.

1.) Eingabedaten

Die Qualität der Eingabedaten bestimmt sich auch die Güte der Ergebnisse. Hierbei

gilt es, einige Faktoren zu beachten. So ist es beispielsweise bei Verwendung der persönlichen Daten eines Nutzers wichtig, dass diese auch Eigenschaften und Interessen der jeweiligen Person beschreiben. Wenn beispielsweise diese Daten aus dem Surfverhalten eines Browsers auf einem Smartphone resultieren, werden die Ergebnisse nicht optimal sein können, da sich nicht garantieren lässt, dass die Recherche des Nutzers ausschließlich seinem Informationsbedarf entspricht und nicht zufällig auch dem von Freunden oder Kollegen. Über eine Parametrisierung des Algorithmus ist der Betreiber zudem in der Lage, durch die Festlegung etwa von Schwellenwerten oder Grenzwerten die Ergebnisse zu beeinflussen. Die Eingabedaten, die Wissen und Erfahrungen in einem bestimmten Bereich dokumentieren, haben ebenso Einfluss auf die Ergebnisse. Daher ist die Kenntnis darüber, was davon genutzt wird für die Bewertung sehr relevant. Denn wenn in den Eingabedaten Vorurteile und diskriminierende Ansichten enthalten sind, werden die modernen neuronalen Netze auch entsprechende Ergebnisse erzeugen. Heute ist es schwierig, die Eingabedaten darauf hin zu überprüfen, weil dafür ein gewünschtes Abbild bezüglich definierter Werte einer Gesellschaft vorhanden sein müsste, das jedoch (noch) nicht existiert.

2.) Manipulieren von Trainingsdaten

Die Eingangsdaten werden so manipuliert, dass Angriffe nicht mehr oder nicht mehr so gut erkannt werden. Zum Beispiel werden bei der Support-Vector-Machine die klassifizierten Eingangsdaten so modifiziert, dass die Hyperebene zur Trennung der klassifizierten Objekte so verändert wird, dass dadurch gezielt unerkannte Angriffe möglich sind.

3.) Algorithmus

Der Umgang mit maschinellem Lernen ist oftmals durch Ausprobieren geprägt und benötigt viel Erfahrung. Es lässt sich vorab nicht eindeutig bestimmen, welcher Ansatz der bestmögliche für eine bestimmte Aufgabenstellung ist. Gerade im unüberwachten Ansatz besteht die reale Möglichkeit, dass Korrelationen in den Input-Daten gefunden werden, die in die Irre führen können. Die Herausforderungen in diesem Bereich liegen darin, eine geeignete Skalierbarkeit der Dateninfrastruktur und eine passende Architektur sowie Algorithmen der automatisierten Entscheidungsfindung abzuleiten. Die Architekten (Zielsetzungsgeber) und Programmierer (Umsetzer) können somit im Prinzip die Ergebnisse durch die konkreten Methoden und deren Umsetzung beeinflussen. Aus diesem Grund wird es zunehmend essenzieller, dass die Richtigkeit der Nutzung von Algorithmen validiert werden kann.

4.) Ergebnisse

Die Ergebnisse sind erst einmal (theoretisch) neutral, weil diese durch den Algorithmus berechnet worden sind. Abhängig von der konkreten Problemstellung, können die gelernten Ergebnisse in der Praxis als schützenswerte Ressource betrachtet werden, da sie beispielsweise Rückschlüsse auf sensible Eingabe- oder Ausgabedaten aus der Lernphase ermöglichen können (Model Inversion Attack).

Im Kontext eines Cybersicherheitsmechanismus auf Basis einer künstlichen Intelligenz, könnten Cyberkriminelle die gelernten Ergebnisse verwenden, um beispielsweise den Erfolg von Angriffen im Vorfeld zu simulieren. Darauf aufbauend könnten von den Cyberkriminellen gegebenenfalls weitere Schutzvorkehrungen gegen eine Erkennung im Produktivumfeld implementiert werden.

Für derartige Angriffsszenarios könnten die Cyberkriminellen entweder direkt auf die gespeicherten Datenstrukturen einer künstlichen Intelligenz zugreifen oder bereitgestellte Funktionen in der Anwendung einer künstlichen Intelligenz verwenden (zum Beispiel API-Aufrufe), um anschließend die Ergebnisse zu rekonstruieren (Model Extraction Attack).

5.) Verwendung

Bei Verwendung der Ergebnisse sind die Einflussmöglichkeiten am größten. So kommen etwa bei der Google-Suchmaschine basierend auf dem Algorithmus die relevantesten Einträge in einer entsprechenden Reihenfolge heraus. Bei der Auflistung von Suchergebnissen jeglicher Art setzt Google jedoch an die erste Stelle Werbung, was eine Manipulation der Resultate darstellt, wie beispielsweise über den Hinweis „Anzeige" dokumentiert wird. Aufgrund dessen ist es leicht vorstellbar, dass jegliche Ergebnisse mithilfe eines weiteren Algorithmus gemäß der Zielsetzung von Google manipuliert werden können, und damit nicht mehr die „relevantesten Einträge" des eigentlichen Algorithmus sind, sondern die von Google präferierten. Diese Art der Manipulation lässt sich bei jedem automatisierten Entscheidungssystemen anwenden.

Zusammenfassung

Künstliche Intelligenz im Bereich Cybersicherheit wird helfen, Angriffe besser zu identifizieren, die wenigen Cybersicherheitsexperten zu unterstützen und die Wirkung von Cybersicherheitslösungen zu erhöhen. Außerdem wird Cybersicherheit benötigt, um den Schutz von künstlicher Intelligenz und deren Ergebnisse zu gewährleisten. ■

NORBERT POHLMANN,
Informatikprofessor für Informationssicherheit und Leiter des Instituts für Internet-Sicherheit – if(is) an der Westfälischen Hochschule in Gelsenkirchen sowie Vorstandsvorsitzender des Bundesverbands IT-Sicherheit – TeleTrusT und im Vorstand des Internetverbandes – eco.

Literatur

[1] *N. Pohlmann: Lehrbuch „Cyber-Sicherheit", Springer Vieweg Verlag, Wiesbaden 2019*