Künstliche Intelligenz und IT-Security

Der Einsatz von künstlicher Intelligenz konnte bereits in vielen Anwendungen die Leistung eines Produktes deutlich steigern, auch im Bereich der IT- Security kann KI neue Leistungsmerkmale setzen. Dieses Paper verschafft zunächst einen Überblick über die aktuelle Gefährdungslage in der IT-Security. Im Rahmen dieser wird das Potential und der Einsatz von künstlicher Intelligenz als Maßnahme zur Bekämpfung von Angriffen auf Informationssysteme eruiert. Im Gegensatz dazu wird anschließend die entgegengesetzte Sicht beleuchtet: Die aktuelle Lage der Sicherheit in der künstlichen Intelligenz. Hierbei werden verschiedene Schwachstellen moderner KI-Systeme aufgezeigt.

Inhalt

Einführung

Seit der Zunahme und Verbreitung von Informationsstrukturen sind diese nicht mehr aus der Infrastruktur von Unternehmen wegzudenken. Neben des immer weiteren Wachstums der IT-Landschaft werden Angriffe aus dem Cyber-Raum auf diese häufiger, komplexer und professioneller [1]. Gerade durch die fortschreitende Vernetzung der Systeme und Endgeräte ist es dabei möglich, durch einen erfolgreichen Angriff ein gesamtes Unternehmen oder Infrastruktur vorrübergehend auszuschalten. Vor allem kritische Infrastrukturen, wie die Versorgung von Wasser und Energie sind zunehmend immer mehr von informationstechnischen Infrastrukturen abhängig. Ein Ausfall dieser

aufgrund eines Angriffs aus dem Cyber-Raum kann dementsprechend große Folgen auf die gesamte Gesellschaft haben. Nicht zuletzt stufte das Bundesamt für Sicherheit in der Informationstechnik (BSI) die aktuelle Gefährdungslage als *hoch* ein, wie aus dem aktuellsten Bericht des Amtes hervorgeht [2]. Eine große Bedrohung geht dabei für Unternehmen von Ransomware aus, die Dateien auf dem Zielsystem verschlüsseln und für das Entschlüsseln Lösegeld verlangt. Dabei steigt die Zahl der bekannten Schadprogramme (Malware) immer weiter an. Neben bereits 800 Millionen bekannten, kommen täglich ca. 400.000 neue dazu. Die Zahl der Identitätsdiebstähle erreicht ebenfalls eine neue Größenordnung. Kriminelle versuchen dabei in den Besitz von Passwörtern, Kreditkartennummern und anderen vertraulichen Daten von Firmen sowie auch Privatpersonen zu gelangen. Auf dem Schwarzmarkt existieren Datenbanken mit gestohlenen Identitäten im Milliardenbereich. Gerade auch das Aufstreben des *Internet of Things* (IoT) sorgt für Probleme: Durch mangelnde Sicherheit der Produkte sind diese automatisiert kompromittierbar und werden so Teile von riesigen Botnets. Mit Hilfe derer können Angreifer dann gezielt Infrastrukturen so stark überlasten, dass diese vorrübergehend nicht erreichbar sind. Diese Art des Angriffs wird auch als *Denial of Service* (DoS) bezeichnet [2].

Herkömmliche Systeme zur Bekämpfung von Schadsoftware und deren Schwächen

Seit es Computer gibt, die beliebige Software ausführen können, gibt es auch unerwünschte Software. Dabei wird in der Regel zwischen unerwünschter und unerwünschter bösartiger unterschieden. Vor allem unerwünscht bösartige Software, auch Malware genannt, stellt für ein System ein großes Sicherheitsrisiko dar. Um ein System erfolgreich mit Schadsoftware zu kompromittieren, ist oft nicht mal eine Schwachstelle in einer genutzten Software von Nöten. Häufig reicht es bereits aus, den Nutzer mittels *Social Engineering* zu täuschen, damit er die Schadsoftware selbst ausführt. Im Laufe der Zeit entstanden verschiedene Softwarelösungen um diese Risiken zu minimieren oder gänzlich zu beheben. Dieses Kapitel soll einen kleinen Überblick über traditionelle Software für IT-Security verschaffen.

Spamfilter

Mehr als die Hälfte des E-Mail-Aufkommens weltweit besteht aus unerwünschten E-Mails, auch als Spam bezeichnet. Viele sind jedoch nicht nur einfach lästige Werbung, sondern enthalten auch gefährliche Schadprogramme [3]. Eine aktuelle Spamwelle mit dem Namen *Emotet* nutzt automatisiertes Social Engineering, um Nutzer beispielsweise dazu zu bringen Schadcode auszuführen [4]. Diese Art von E-Mails sind eine Untergruppe von Spam und werden auch Phishing-E-Mails genannt [3].

Als Gegenmaßnahme existieren Spamfilter – Programme, die automatisiert E-Mails in die Klasse *erwünscht* und *unerwünscht* einteilen. Dabei gibt es unterschiedliche Herangehensweisen dies umzusetzen. Die einfachste Möglichkeit ist der regelbasierte Ansatz. Dabei wird zum Beispiel geprüft, ob ein bestimmtes Wort in der E-Mail vorkommt oder der Absender in einer Liste unerwünschter Absender steht. Dieser Ansatz hat jedoch das Problem, auf diese Weise auch erwünschte E-Mails als Spam zu klassifizieren, sollte ein sonst unerwünschtes Wort in einer erwünschten E-Mail vorkommen. Ein weiterer in der Praxis häufig eingesetzter Ansatz ist der *Naive Bayes Classifier*. Im Training lernt der Algorithmus die bedingte Wahrscheinlichkeit jedes Wortes einer E-Mail, und kann danach aufgrund der summierten bedingten Wahrscheinlichkeiten aller Wörter einer E-Mail auf die Klassifizierung schließen. Dieser Ansatz wird bereits zum Einsatz künstlicher Intelligenz in der IT-Security gezählt, aufgrund des selbständigen Lernens welches Wort mit welcher Wahrscheinlichkeit zu welcher Klasse gehört [5]. Diese Ansätze funktionieren gut für Werbe-E-Mails. Bei jedoch so genannten *Spear-Phishing-E-Mails*, bei denen Empfänger gezielt mit realen Tatsachen getäuscht werden, versagen diese Ansätze.

Virenschutzprogramme

Bereits seit den 80er Jahren existieren Computerprogramme mit der Aufgabe Schadsoftware auf einem Computersystem aufzuspüren. Die Funktionsweise ist dabei recht einfach. Das Virenschutzprogramm verfügt über eine Datenbank mit allen bekannten Signaturen von Schadsoftware und vergleicht diese mit jeder Datei des Computersystems. Sind zwei Signaturen identisch meldet das Virenschutzprogramm einen Virenfund. Die Signatur einer Datei wird über den sogenannten *Hashwert* ermittelt. Mit einem Hashing-Algorithmus ist es möglich, aus einer beliebigen Datensequenz, eine immer gleiche und eindeutige Prüfsumme zu berechnen. Diese wird verwendet um Dateien miteinander zu vergleichen [6]. Um dieses Funktionsprinzip zu ermöglichen ist es daher unumgänglich, dass Hersteller regelmäßig die Datenbanken über bekannte Schadsoftware aktualisieren und Anwender ebenfalls ein stets aktualisiertes Virenschutzprogramm vorliegen haben. Ist jedoch die Schadsoftware minimal abgewandelt oder komplett neu, ist ihre Signatur anders und kann nicht mehr auf diese Weise als Schadsoftware erkannt werden [6].

Intrusion Detection System (IDS)

Mit dem Einsatz eines IDS können Angriffe auf ein Computersystem automatisch musterbasiert erkannt werden. Dadurch ergänzt die Software die Funktionalität üblicher Firewalls. Die Funktionalität eines IDS bezieht sich hauptsächlich auf die Datenanalyse des gesamten Netzwerkverkehrs eines informationstechnischen Systems. Dabei sollte sichergestellt sein, dass alle zu analysierenden Daten aus vertrauenswürdigen Quellen stammen, beziehungsweise ausgeschlossen werden kann, dass diese nicht schon zuvor korrumpiert worden sind. Für die Datenanalyse prüft das System die Daten auf vordefinierte Muster, die durch Datenbanken vorgegeben werden. Auf diese Weise können bekannte Angriffe erkannt werden. Ebenfalls sollen durch die Analyse des Netzwerkverkehrs Anomalien, also Abweichungen vom Normalbetrieb erkannt werden. Vereinzelt kommt dabei bereits heute schon künstliche Intelligenz zum Einsatz. Ein bestehendes Problem ist jedoch die falsch-positiv Erkennungsrate. Da lieber ein Angriff zu viel als zu wenig erkannt werden soll, muss ein Mensch nachfolgend bewerten, ob es sich um einen tatsächlichen Angriff handelt. Ebenfalls kann das System gänzlich neue oder gezielte Angriffe nur schwer erkennen, da diese in der Datenbank als Mustervorgabe nicht enthalten sind [7].

Potentiale für künstliche Intelligenz in der IT-Security

Zwei von drei aller Unternehmen geben an, ohne den Einsatz von künstlicher Intelligenz keine Chance mehr in der IT-Security zu sehen. Dies geht aus dem aktuellen Cybersecurity-Report der Firma *Capgemini* hervor [8]. Die Gründe dafür sind verschiedener Natur. Bereits im Jahr 2021 werden 25 Milliarden verschiedener Geräte im Internet erwartet. Durch immer mehr Geräte im Internet wird ein dreifacher Anstieg des weltweiten Datenverkehrs im Internet in den Jahren 2017 bis 2021 erwartet. Somit bietet dies auch einen stetig wachsenden Markt für Internetkriminelle. Demnach steigen die Angriffe aus dem Cyber-Raum auf Unternehmen – die angestellten Security-Analysten sind bereits jetzt schon überlastet aufgrund der großen Datenmengen. Die Überlastung der Angestellten kommt vor allem auch durch den Einsatz von regelbasierter Sicherheitssoftware zu Stande, die viele *false positives* generiert. Dadurch verschwenden Analysten viel Zeit damit, um zu überprüfen, ob tatsächlich ein Angriff oder schadhafter E-Mail-Anhang vorliegt. Ebenfalls durch die fortschreitende Digitalisierung, den Einsatz von IoT-Geräten und Smart Devices werden immer mehr Geräte zu potenziellen Angriffszielen. Die Umstellung auf Industrie 4.0 und der Einsatz von Informationssystemen in beispielsweise Transportmitteln bieten Kriminellen auch eine immer größer wachsende Angriffsfläche. Neben der Häufung von Angriffen auf Unternehmen, werden diese auch professioneller. Denn Kriminelle nutzen auch künstliche Intelligenzen, um Angriffe durchzuführen. Diese kommt beispielsweise zum Einsatz, um automatisch nach Schwachstellen zu suchen oder regelbasierte

Abwehrsysteme zu umgehen. Dabei ist die künstliche Intelligenz soweit im Stande, die Schadsoftware automatisch so zu mutieren, dass diese nicht mehr von Virenscannern erkannt werden kann [8].

Wenn KI eines sehr gut beherrscht, dann ist dies der Umgang mit großen Datenmengen. Wie bereits zuvor angesprochen liegt dies auch in der IT-Security-Branche vor und bietet dort demnach gute Grundvoraussetzungen für den Einsatz von KI. Häufig scheitern aktuell eingesetzte regelbasierte Systeme aufgrund mehrerer Probleme. Das größte Problem ist die mangelnde Möglichkeit Muster in den Daten zu erkennen, um so auf einen semantischen Zusammenhang in den Daten zu schließen. Außerdem müssen die Regeln zunächst manuell erstellt werden und in regelmäßigen Abständen gewartet werden, um den neusten Stand der Bedrohungen zu erkennen. Ein weiteres Problem dabei ist die Zeit verstreicht, wenn ein Angriff erkannt wird (*Detection*), bis auch tatsächlich reagiert wird (*Response*). Diese Problemstellungen könnten durch den Einsatz verschiedener künstlicher Intelligenzen bearbeitet werden. Da eine künstliche Intelligenz deutlich schneller als ein Mensch arbeiten kann, würde diese vor allem in Bereichen wo viele Daten in Echtzeit geprüft werden müssen, erhebliche Verbesserungen bringen (siehe Siemens) [8, 9].

Jedoch darf KI nicht selbst zur Schwachstelle bei der Bekämpfung von Sicherheitsrisiken werden. Da Angreifer auch auf Automatisierung setzen, ist es wichtig, dass diese sich nicht gezielt von Angreifern überlasten und ausschalten lässt. Eine Angriffs-KI könnte so beispielsweise automatisierte kleine jeweils leicht veränderte Angriffe senden. Ein wichtiges Leistungsmerkmal für die Security-KI wäre es demnach das Muster zu erkennen und *alle* Angriffe direkt abzublocken. Eine weitere Herausforderung für eine KI im Security Umfeld sind die großen Abweichungen der Einsatzfelder in jedem Unternehmen. Demnach ist es schwer möglich eine standardisierte Security-KI für verschiedene Unternehmen zu entwerfen, da jedes eine andere informationstechnische Infrastruktur besitzt und andere Geschäftsprozesse implementiert. Entsprechend würden Angreifer auch verschiedene Wege suchen, um einen erfolgreichen Angriff durchzuführen. Was daraus abgeleitet werden kann, ist dass es unumgänglich ist, eine Security-KI in ihrem jeweiligen Einsatzumfeld zu trainieren. Dies schließt dabei auch aus, dass mehrere Unternehmen für die gleichen standardisierten Angriffe verwundbar sind [8, 9].

Die Potentiale für die Arbeit mit künstlicher Intelligenz sind auf jeden Fall gegeben. Nicht nur die Angreifer werden professioneller, auch die Unternehmen müssen es werden. Um dies jedoch umzusetzen ist es nötig, die Investitionen für die Entwicklung entsprechender Technologien zu erhöhen. Wie aus dem Cybersecurity-Report der Firma *Capgemini* auch hervorgeht, sehen viele Unternehmen große Potentiale, sogar die einzige Chance, ihre IT-Systeme vor Angriffen zu schützen durch die Hilfe einer KI. Firmen, die bereits mit künstlicher Intelligenz auf diesem Gebiet arbeiten, geben eine deutliche Verbesserung in mehreren Punkten an. Nicht nur die Kosten sinken nachhaltig, sondern auch die Qualität steigt deutlich. 71% der befragten Unternehmen geben an, durch den Einsatz von KI in der IT-Security schneller auf Angriffe reagieren und diese einschränken oder sogar verhindern zu können [8].

Einsätze von künstlichen Intelligenzen in der Praxis

In der Industrie und freien Wirtschaft sind bereits einige Systeme als Maßnahme zum Schutz von IT-Infrastrukturen vor Angriffen im Einsatz. Im Rahmen der Literaturrecherche für dieses Paper werden in diesem Kapitel einige Beispiele aufgeführt und genauer erläutert.

PerimeterX

Das Startup PerimeterX aus San Francisco nutzt künstliche Intelligenz für die Erkennung von Angriffen auf Webserver. Um einen erfolgreichen Angriff auf einen Webserver durchzuführen, nutzen professionelle Angreifer hauptsächlich automatisierte Bots, die automatisiert eine Webseite nach der anderen auf gängige Sicherheitslücken überprüfen und gegebenenfalls ausnutzen [10]. Gerade bei großen Web-Plattformen, die auf verteilte Systeme wie Microservices setzen, ist es sehr schwierig den Überblick zu behalten und verdächtigen Netzwerkverkehr wie diesen aufzuspüren. Regelbasierte Systeme stoßen dabei schnell an ihre Grenzen und Menschen müssen viele Ressourcen investieren um die Meldungen solcher Systeme zu bewerten [8].

Mit Hilfe künstlicher Intelligenz will das Unternehmen PerimeterX Anomalien, die durch verdächtigen Netzwerkverkehr wie zum Beispiel Bots entstehen, automatisiert erkennen. Dabei lernt die KI das Verhalten eines normalen Nutzers auf einer Webseite von dem eines Bots zu unterscheiden. Die Datengrundlage für die KI bilden dabei unter anderem die Bewegung der Maus und das Scroll-Verhalten des Nutzers. Auch die Hardwareressourcen des Nutzers werden dabei mit einbezogen. Werden zum Beispiel keine Mausbewegungen in Kombination von Hardwareressourcen, die dem eines Servers in einem Rechenzentrum gleichen, festgestellt, kann die KI beispielsweise einen Bot erkennen und automatisiert den Netzwerkverkehr mit dem Ziel einstellen [10, 11].

Siemens

In einem Bericht von Amazon Cloud Services (AWS) berichtet das Unternehmen über seine KI-basierte Cyber-Defence Strategie in der Amazon Cloud. Das Siemens Cyber Defence Center betreibt diese und hat damit die Aufgabe das Unternehmen vor Schadsoftware, Internetkriminalität und Identitätsdiebstählen zu beschützen. Laut Angaben des globalen Unternehmens fallen dabei weltweit 200.000 Schadsoftware-Dateien an, weit mehr als von Menschenhand verarbeitet werden könnten. Aufgrund solch massiv großer Datenmengen, entschied sich das Unternehmen für eine KI-basierte Lösung der Amazon Cloud [12].

Der Zweck der künstlichen Intelligenz liegt in der Erkennung von Angriffen, Einstufung von Schadsoftware und dem automatischen Eingreifen im Ernstfall [12]. Für das Vorbereiten der Daten für das Training der KI wird der Service *SageMaker* von Amazon benutzt. SageMaker ist eine Softwarelösung für die automatische Unterstützung bei der Erstellung von Machine-Learning-Algorithmen, insbesondere dem Datenaufbereiten und dem Kennzeichnen von Daten [13]. Für die großen Datenmengen und Bereitstellung der Daten für die KI nutzt das Unternehmen den Amazon Simple Storage Service (Amazon S3). Dadurch ist es möglich die 6 Terrabyte Log-Dateien, die im Unternehmen täglich anfallen zu verwalten und für die Analyse bereitzustellen. Schneller als jeder Mensch schafft die KI es, 60.000 potenziell gefährliche Datensätze zu analysieren und gegebenenfalls Alarm zu schlagen. Die genaue Erkennungsrate hält Siemens geheim, gibt jedoch an, durch den KI-basierten Ansatz die falsch erkannten Bedrohungen drastisch reduziert zu haben [12].

General Electric: Digital Ghost

Auch das US-amerikanische Unternehmen General Electrics (GE) arbeitet mit einer künstlichen Intelligenz zur Bekämpfung von Bedrohungen aus dem Cyber-Raum. Dabei setzt das Unternehmen auf sein System mit dem Namen *Digital Ghost.* Das Ziel von Digital Ghost ist es, Anomalien im Verhalten von Systemen, insbesondere großen Industriesystemen, zu erkennen. Dabei soll vor allem erkannt werden, ob das System kompromittiert wurde. Einsatzgebiete sieht das Unternehmen hauptsächlich in der IT-Infrastruktur von großen Transportmitteln wie Zügen oder Flugzeugen, aber auch in den Kontrollinfrastrukturen kritischer Infrastrukturen wie Kraftwerken [14].

Die Funktionsweise setzt direkt auf der Hardware-Ebene eines Systems an. Basierend auf dieser wird durch eine Hardware-Simulation ein digitaler Zwilling des Systems erzeugt. Die KI wird mit den Daten, wie Sensordaten des Echt-Systems trainiert und lernt so das Verhalten des Systems. Im Echtzeitbetrieb erhält die KI Daten des Echt-Systems und Daten der Simulation und kann darauf basierend Abweichungen erkennen. Digital Ghost soll dabei vier große Funktionalitäten erfüllen können. Wie bereits angesprochen steht die Erkennung von Anomalien im Fokus. Darüber hinaus kann erkannt werden, wo im System die Anomalie auftritt. Zusätzlich gibt Digital Ghost eine Möglichkeit des Echtzeit-Monitorings in das System und kann bei Angriffen selbständig versuchen diese abzublocken, beziehungsweise gefährdete Teilsysteme direkt zu schützen [14].

Laut eigenen Angaben des Unternehmens erkennt die KI 98% der Cyber-Attacken auf das System und das auch wenn sogar mehr als die Hälfte des Systems bereits kompromittiert wurde. Dies bedeutet auch wenn 50% der Daten fehlerhaft sind, kann die KI trotzdem das richtige Ergebnis vorhersagen [14].

Security in der Künstlichen Intelligenz

In diesem Paper wurde zunächst die Seite beleuchtet, wie künstliche Intelligenz als Werkzeug einen Mehrwert in der IT-Security Branche erzielen kann. In diesem Kapitel soll der Blickwinkel sich jedoch darauf richten, wo künstliche Intelligenz aktuell massive Sicherheitsprobleme aufweist. Denn bereits mehrere Teams von Wissenschaftlern weltweit haben veröffentlicht, wie einfach künstliche Intelligenzen ausgetrickst werden können [15]. Aufgrund des unaufhaltsamen Vormarschs der künstlichen Intelligenz in immer mehr Aufgabenbereiche ist dabei jedoch Vorsicht geboten. Zu leicht könnten Kriminelle die KI für Ihre Zwecke manipulieren. Anhand von mehreren aktuellen Beispielen soll dies nun erläutert werden.

Gezielte Manipulation der Eingaben

Künstliche Intelligenzen haben in den letzten Jahren in kürzester Zeit immer neue Bestmarken gebrochen, was die Bildverarbeitung und -Erkennung angeht. Dabei kamen verschiedenste Architekturen tiefer neuronaler Netzte (DNN) zum Einsatz. Doch wie nun verschiedene Forscherteams unabhängig herausgefunden haben, genügt bereits eine kleine Veränderung der Ausgabe, um das neuronale Netzt zu täuschen und eine falsche Entscheidung herbeizuführen [16].

Man stelle sich vor, ein autonom-fahrendes Auto steuert auf ein Stoppschild zu. Anstatt jedoch zu bremsen, beschleunigt es immer mehr. Wie kann dies sein? Die KI des Autos hat aufgrund mehrerer Rechtecke auf dem Stoppschild, es fälschlicherweise als Tempo 45 Schild klassifiziert. Dies ist zwar nur ein fiktives Beispielszenario, jedoch weißt es eine erschreckende Realitätsnähe auf. Im Straßenverkehr sind häufig Verunreinigungen auf Straßenschildern zu finden [15]. Des Weiteren zeigten die Wissenschaftler auf Abbildung 1 auf, dass zufallsgenerierte Eingaben bereits die KI dazu verleiten, darauf etwas erkennen zu wollen, obwohl dort kein sinnvolles Objekt abgebildet ist. Beispielsweise wurde dort ein Königspinguin und ein Seestern erkannt [16].

Ein weiteres Beispiel wie einfach moderne Bilderkennungssysteme basierend auf tiefen neuronalen Netzen manipuliert werden können, zeigen die Autoren der „One Pixel Attack“ [17]. Dabei entwickelten die Wissenschaftler eine Methode, um gezielt den Farbwert von genau einem Pixel des Bildes so zu verändern, dass das trainierte Modell eine falsche Aussage trifft. Wie auf Abbildung 2 zu sehen ist, genügt jeweils ein bestimmter Pixel (rot umrandet), um die Wahrscheinlichkeit der Objektvorhersage zu verändern.

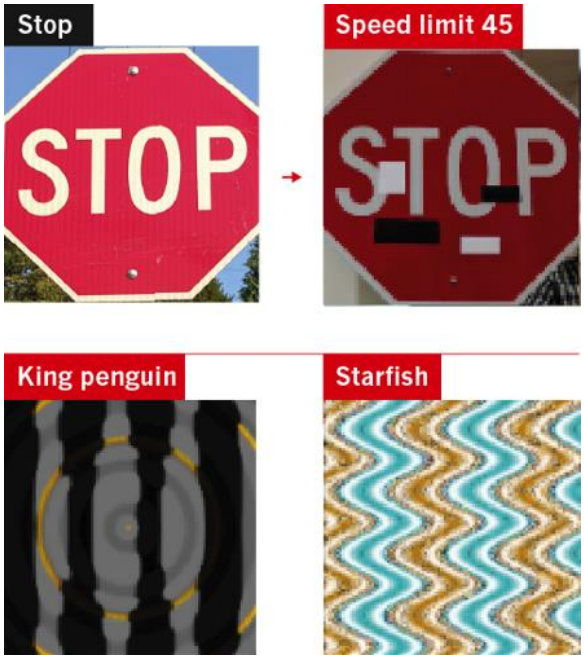

Abbildung 1: Manipulation der Eingabe [16]

Abbildung 2: One Pixel Attack [17]

Täuschen durch unerwartetes Verhalten

Jedoch nicht nur die künstlichen Intelligenzen in der Bildverarbeitung haben Sicherheitsprobleme. Auch durch Deep Reinforcement Learning trainierte neuronale Netze weisen Probleme auf. Beim Reinforcement Learning versucht der Algorithmus anhand von sporadischen positiven oder negativen „Belohnungen“ (Rewards) sich bestmöglich an die Umgebung und die gestellte Aufgabe anzupassen. In dem hier als Quelle vorliegenden Versuch wurde in einer dreidimensionalen Computergrafik-Umgebung ein Agent so trainiert, bei einem sich immer gleichverhaltenden Torwart ein Tor zu schießen. Wenn der Torwart sich jedoch plötzlich anders verhielt, sich zum Beispiel gezielt auf den Boden fallen ließ, konnte der Agent plötzlich sein Gelerntes nicht mehr anwenden und kein Tor mehr schießen [18].

Fazit

Künstliche Intelligenzen sind immer weiter auf dem Vormarsch und haben bereits begonnen immer mehr Bereiche neu zu erfinden. Auch die IT-Security Branche kann davon nicht verschont bleiben. Die immer weiter ausgeprägte Nutzung von informationstechnischen Endgeräten und die damit steigende Zahl der Nutzer hat einen massiven Anstieg des Datenverkehrs im Internet zur Folge. Dies bietet Kriminellen eine stetig wachsende Angriffsfläche. Doch nicht nur das, auch diese haben künstliche Intelligenzen für sich entdeckt, um gezielte Schwachstellen in IT-Infrastrukturen aufzuspüren, auszunutzen oder um Social Engineering durchzuführen. Unternehmen müssen darauf reagieren und mit der Leistung der Technik mitziehen. Wie in diesem Paper aufgezeigt, gibt es viele Aufgabenbereiche in der IT-Security, die vor allem aufgrund des hohen Datenaufkommens in Kombination mit künstlicher Intelligenz profitieren können. Bereits heute sind viele KI-basierte Systeme bei Unternehmen im Einsatz, diese geben eine deutliche Verbesserung der Gesamtleistung

an. Doch wird KI die IT-Security revolutionieren und sicherer machen? Davon ist nicht auszugehen, da Hacker ebenfalls künstliche Intelligenzen einsetzten und verbessern werden. Es bleibt also nach wie vor ein wetteifernder Rüstungskrieg beider Seiten, der nun KI als neue Waffe für sein Arsenal gewinnen konnte.

Doch das was es der IT-Security Branche noch an KI fehlt, das fehlt es der KI an Security. Moderne künstliche Intelligenzen erreichen immer weitere Bestmarken für Klassifizierungsaufgaben (beispielsweise Bilder), sind jedoch durch simple Manipulationen bereits zu täuschen. Die Recherche hat gezeigt, sobald es sich um unrealistische, verrauschte oder gezielt manipulierte Eingaben handelt, ist die KI überfordert und liefert ein falsches Ergebnis. Dies liegt vor allem an dem fehlenden Weltbild der KI. Sie ist heutzutage nicht im Stande zu improvisieren oder eine abstrakte Semantik abzuleiten, sondern lediglich in der Lage statistische Muster zu erkennen. Auf diesem Gebiet muss noch viel getan werden, um künstliche Intelligenz nicht nur als „Hype“ sondern auch als zukunftsweisende Schlüsseltechnologie zu etablieren.

Literatur

[1] BSI, „Cyber-Sicherheit: Gefährdungslage“, 2020. [Online]. Verfügbar unter: https://www.bsi.bund.de/DE/Themen/Cyber-Sicherheit/Gefaehrdungslage/cs_Gefaehrdungslage_node.html.

[2] „Die Lage der IT-Sicherheit in Deutschland 2018“, 2018. [Online]. Verfügbar unter: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2018.pdf?__blob=publicationFile&v=6.

[3] BSI, Hg., „Spam, Phishing & Co“, 2020. [Online]. Verfügbar unter: https://www.bsi-fuer-buerger.de/BSIFB/DE/Risiken/SpamPhishingCo/spamPhishingCo_node.html. Zugriff am: 4. März 2020.

[4] „Aktuelle Information zur Schadsoftware Emotet“, 2020. [Online]. Verfügbar unter: https://www.bsi-fuer-buerger.de/BSIFB/DE/Service/Aktuell/Informationen/Artikel/emotet.html. Zugriff am: 4. März 2020.

[5] P. Graham, *Better Bayesian Filtering.* [Online]. Verfügbar unter: http://www.paulgraham.com/better.html.

[6] BSI, Hg., „Virenschutzprogramme“, 2020. [Online]. Verfügbar unter: https://www.bsi-fuer-buerger.de/BSIFB/DE/Empfehlungen/Schutzprogramme/Virenschutzprogramme/virenschutzprogramme_node.html. Zugriff am: 5. März 2020.

[7] S. Luber, *Was ist ein Intrusion Detection System (IDS)?* [Online]. Verfügbar unter: https://www.security-insider.de/was-ist-ein-intrusion-detection-system-ids-a-612870/. Zugriff am: 9. März 2020.

[8] G. van der Linden, J. Xu, M. Markus und D. Morefield, „Reinventing Cybersecurity with Artificial Intelligence: The new frontier in digital security“, 2019. [Online]. Verfügbar unter: https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_20190711_V06.pdf.

[9] P. Schmitz, *Bad Bots vs. Security-KI: Erfolgsfaktoren beim Einsatz von KI.* [Online]. Verfügbar unter: https://www.security-insider.de/bad-bots-vs-security-ki-a-904608/. Zugriff am: 2. März 2020.

[10] R. Miller, *PerimeterX secures $43M to protect web apps from bot attacks.* [Online]. Verfügbar unter: https://techcrunch.com/2019/02/11/perimeterx-secures-43m-to-protect-web-apps-from-bot-attacks/.

[11] *PerimeterX*. [Online]. Verfügbar unter: https://www.perimeterx.com/. Zugriff am: 2. März 2020.

[12] J. Pospisil, „Siemens Handles 60,000 Cyber Threats per Second Using AWS Machine Learning“, 2019. [Online]. Verfügbar unter: https://aws.amazon.com/de/solutions/case-studies/siemens-cybersecurity/. Zugriff am: 2. März 2020.

[13] AWS, *Amazon SageMaker: Machine Learning für jeden Entwickler und Daten-Wissenschaftler.* [Online]. Verfügbar unter: https://aws.amazon.com/de/sagemaker/. Zugriff am: 2. März 2020.

[14] J. John, „Digital Ghost: Real-Time, Active Cyber Defense“. [Online]. Verfügbar unter: https://www.ge.com/research/offering/digital-ghost-real-time-active-cyber-defense. Zugriff am: 2. März 2020.

[15] D. Haeven, „Why deep-learning AIs are so easy to fool“, 2020. [Online]. Verfügbar unter: https://www.nature.com/articles/d41586-019-03013-5. Zugriff am: 5. März 2020.

[16] A. Nguyen, J. Yosinski und J. Clune, „Deep neural networks are easily fooled: High confidence predictions for unrecognizable images“ in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 07.06.2015 - 12.06.2015, S. 427–436, doi: 10.1109/CVPR.2015.7298640.

[17] J. Su, D. V. Vargas und S. Kouichi, „One pixel attack for fooling deep neural networks“, *IEEE Trans. Evol. Computat.*, Jg. 23, Nr. 5, S. 828–841, 2019, doi: 10.1109/TEVC.2019.2890858.

[18] A. Gleave *et al.*, „Adversarial Policies: Attacking Deep Reinforcement Learning“, 25. Mai 2019. [Online]. Verfügbar unter: http://arxiv.org/pdf/1905.10615v2.