WHITEPAPER

Cybersicherheit im Zeitalter Künstlicher Intelligenz

Ondrej Kubovič und René Holt

mit Unterstützung von:

Juraj Jánošík

Filip Mazán

Peter Košinár

Jakub Debski

Peter Stančík

Digital Security
Progress. Protected.

April 2024

Inhaltsverzeichnis

Digital Security
Progress. Protected.

Einführung

Künstliche Intelligenz (KI) ist aktuell eines der am meisten gehypten Themen – sowohl Presse als auch Verkaufs- und Marketingmaterialien sind voll davon und eine schier endlose Zahl an Online-Diensten setzt ebenfalls auf den Trend. Führte 2019 die Online-Suche nach dem englischen Äquivalent „AI" (Artificial Intelligence) noch zu über 2 Milliarden Treffern, sind es 5 Jahre später fast 18 Milliarden Ergebnisse. Offensichtlich ist also das öffentliche Interesse gewachsen.

Der größte Teil des Hypes lässt sich auf große Sprachmodelle (large language models, LLMs) zurückführen, wodurch die Kommunikationsfähigkeit und die visuelle Kreativität von KI-Systemen bedenklich nahe an die menschlichen Fähigkeiten gekommen sind. Einige sehen darin lukrative Geschäftsmöglichkeiten und eine blühende Zukunft voller KI-Lösungen, andere wiederum fürchten um Millionen von Arbeitsplätzen und sogar die Menschheit selbst.

In diesem Whitepaper werden wir uns von großen Visionen und Untergangsszenarien fernhalten. Wir konzentrieren uns stattdessen auf die **tatsächlichen Vorteile und Risiken dieser Technologie für die Cybersicherheit**.

Wir werden zeigen, wie wir **präzise ausgewählte KI-Algorithmen in unsere Erkennungssysteme integriert** und sie so zu hocheffektiven Schutzlösungen mit maximalen Erkennungsraten und minimalen Fehlalarmen entwickelt haben.

Darüber hinaus erklären wir, wie unsere **Threat Intelligence- und Threat Hunting-Services von KI profitieren**, indem sie unsere Experten u. a. auf interessante Angriffsszenarien und Malware-Merkmale aufmerksam macht.

Uns ist auch bewusst, welchen Schaden KI in den Händen von Cyberkriminellen und staatlich gelenkten Bedrohungsakteuren anrichten kann. Wir werden uns also auch **mit den Gefahren befassen, die mit dieser Technologie einhergehen**. Unser Schwerpunkt liegt dabei auf den generativen Fähigkeiten der KI, also der Möglichkeit, neue Malware zu entwickeln, die Finesse bei der Adressierung der Opfer bei Social Engineering-Kampagnen zu verbessern oder die Qualität und Quantität von Spam-Kampagnen zu steigern.

Im Kampf gegen solche Bedrohungen spielt KI eine zentrale Rolle. **Sie kann riesige Datensätze in Echtzeit analysieren und dabei schnell Muster und Anomalien erkennen**, die auf neue Bedrohungen oder Sicherheitslücken hindeuten. Das wiederum ermöglicht eine rasche Behebung. Für uns als Sicherheitsexperten ist hierbei besonders wichtig, dass **KI zur Gefahrenprävention beiträgt und gleichzeitig eine proaktive und adaptive Verteidigungsstrategie** unterstützt.

KI im Dienste des Guten

Angesichts komplexer Bedrohungsszenarien sowie einer immer größer werdenden Zahl an Cyberkriminellen und staatlich organisierten Bedrohungsakteuren einerseits und dem anhaltenden Fachkräftemangel in der IT Security-Branche andererseits können KI-basierte Innovationen in der Cybersicherheit einen wichtigen Beitrag leisten.

Bereits seit vielen Jahren wird KI bei der Erkennung von Bedrohungen, Filterung und Analyse entsprechender Daten sowie in verschiedenen Tools eingesetzt. Die Möglichkeiten sind aber bei Weitem noch nicht ausgeschöpft. Besonders deutlich wird das am Aufkommen generativer KI und großer Sprachmodelle (LLMs), die eine am Nutzer orientierte Kommunikation ermöglichen und auch im Bereich der IT-Sicherheit dessen Fragen angemessen und verständlich beantworten können. Davon profitieren Entwickler und Spezialisten genauso wie Anwender mit weniger Wissen über Cybersicherheit.

Aktueller Einsatz von KI bei der Bedrohungsabwehr:

- Verarbeitung riesiger Datenmengen zur Erkennung von Angriffen durch Korrelation verschiedener Indikatoren.
- Identifizierung und Analyse von Programmen mit verdächtigem oder schadhaftem Code und auffälligem Verhalten.
- Überwachung und Analyse des Netzwerkverkehrs auf bösartige oder auffällige Muster.
- Beschreibung und Erläuterung komplexer Bedrohungsinformationen, um sie leichter zugänglich zu machen.
- Priorisierung von Warnmeldungen, damit sich die Sicherheitsexperten auf die kritischsten Vorfälle konzentrieren können.
- Bereitstellung ergänzender Funktionen für andere Schutzschichten.

Einsatz von KI bei der Bedrohungsabwehr, der derzeit entwickelt wird oder geplant ist:

- Erstellung neuer Erkennungen anhand der Beschreibungen von Bedrohungen.
- Analyse, Kontextualisierung und Erklärung von vergangenen oder aktuellen Ereignissen in einer konkreten Umgebung.
- Prüfung der Unternehmensumgebung auf versteckte oder unbekannte Schwachstellen.

Was IT-Sicherheitsexperten über KI wissen müssen:

- Hohe Raten an Fehlalarmen einer allein auf KI gestützten IT-Sicherheit schränken ihre Nutzbarkeit ein.
- Ohne Updates und menschliche Aufsicht wird sich ein KI-basierter Schutz aller Wahrscheinlichkeit nach stetig verschlechtern.
- Qualitativ hochwertige Trainingssätze führen zu qualitativ hochwertigen KI-Modellen.

Ausgewählte Beispiele:

Erläuterungen zu Bedrohungsdaten

LLMs sind in der Lage, aus langen und komplizierten Inhalten die wichtigsten Informationen herauszufiltern und zusammenzufassen. Das kann insbesondere für Endnutzer, IT-Administratoren und Entscheidungsträger sehr nützlich sein. In der Regel fällt es diesen Gruppen schwer, die hochtechnischen Berichte von Sicherheitsexperten zu verstehen. Mithilfe generativer KI-Modelle können im Handumdrehen kurze, leicht verständliche Texte mit allen relevanten Bedrohungsinformationen und passenden Handlungsempfehlungen erstellt werden.

KI-Assistenten in Produkten

Immer wieder entstehen Schwachstellen durch falsch konfigurierte Systeme und Sicherheitsprodukte in Unternehmensumgebungen und bieten Kriminellen ein potenzielles Einfallstor. Integrierte KI-Assistenten können Organisationen bei der Einrichtung ihrer Umgebung und passenden Konfiguration der Schutzlösungen unterstützen. Das führt nicht nur zu einer verbesserten Benutzererfahrung, sondern steigert unter Umständen auch die Schutzfunktion. Darüber hinaus können KI-Assistenten definierte Aufgaben übernehmen und z. B. eine Ausnahme zur Firewall-Regel hinzufügen oder bestimmte Einstellungen ändern. Zudem können sie dem IT-Administrator bei der Behebung von Sicherheits- oder Systemmeldungen helfen.

KI für den richtigen Fokus

Auch IT-Sicherheitsexperten können von KI-Modellen enorm profitieren. Heutzutage sind solche Teams konfrontiert mit zunehmend komplexen Umgebungen und einer entsprechend großen potenziellen Angriffsfläche, einer wachsenden Anzahl an Sicherheitstools und vor allem einer Flut an Daten. Die Verwaltung aller Aufgaben ist mühsam, zeitaufwändig und erfordert umfassendes Fachwissen. KI kann einen Teil der Belastung abfangen. Sie kann erkannte Ereignisse nach Priorität ordnen und dringende Probleme hervorheben, den Kontext bereitstellen und weniger relevante Alarme selbst lösen. Darüber hinaus kann die KI verschiedene Ereignisse in einen Zusammenhang bringen und dynamische Karten erstellen, die einen umfassenden Überblick über Vorfälle bieten und deren Untersuchung erleichtern. Dadurch können sich Mitarbeitende auf die relevanten Dinge fokussieren und effektiver arbeiten.

Erstellung neuer Erkennungen

Täglich werden unzählige neue Bedrohungsakteure und Angriffstechniken aufgedeckt. Ein KI-Modell kann öffentliche Informationen und Indicators of compromise (IoCs) mit Data Feeds sowie dem Input der Experten kombinieren und Erkennungen generieren. So kann die Unternehmensumgebung auf diese neuesten Bedrohungen gescannt werden. Zur Vermeidung von Fehlalarmen und zur besseren Einordnung von Vorfällen müssen solche Erkennungen jedoch noch von Experten auf ihre Genauigkeit und Wirksamkeit überprüft werden.

Verbesserung des Sicherheitsbewusstseins

Technische Maßnahmen allein reichen für eine gute IT-Sicherheit nicht aus, geschulte Mitarbeiter bzw. Nutzer mit einem entsprechenden Bewusstsein tragen ebenfalls dazu bei. Auch hier kann KI in vielfältiger Weise unterstützen – angefangen bei der Erstellung von kurzen Zusammenfassungen oder Infografiken über Medienartikel zu den neuesten Bedrohungen. Sie können Phishing-Mails für interne Tests erstellen sowie Ergebnisse und Risiken darlegen. Für eine spielerische Sensibilisierung lassen sich Quizfragen und andere Schulungsmaterialien entwerfen, die die interne Kultur, Produkte, Prozesse und andere Besonderheiten des jeweiligen Unternehmens berücksichtigen.

Mit KI erweitertes Sandboxing

Generative KI kann bei der Untersuchung von potenziellen Bedrohungen in Sandbox-Umgebungen nützlich sein. Sie kann beispielsweise Texte in Screenshots oder Low-Level-Ereignisse wie Änderungen in der Windows Registry analysieren und leicht verständliche Beschreibungen des Verhaltens und der Fähigkeiten einer Bedrohung verfassen.

Erweiterter Schutz vor Phishing und Spam

Spam- und Phishing-Mails sind jedem Nutzer und Sicherheitsverantwortlichen ein Dorn im Auge. KI-Modelle könnten genutzt werden, um anhand der bisherigen E-Mail-Kommunikation eines bestimmten Benutzers trainiert zu werden und anschließend auffälliges Verhalten zu erkennen. Ist die KI in der Lage, eine plötzliche Änderung der Schreibgewohnheiten oder des Nachrichteninhalts zu erkennen, können z. B. Angriffe via E-Mail-Antwortketten (reply-chain attack) verhindert werden. Bei solchen Angriffen lassen Kriminelle ihre schädlichen E-Mails vertrauenswürdig aussehen, indem sie auf eine bereits bestehende E-Mail-Kommunikation ihres Opfers antworten.

Eine derartige Schutzlösung ist für die breite Masse aufgrund des hohen Trainingsaufwands des KI-Modells zwar zu kostspielig, aber für ausgewählte Personen könnte solch eine Lösung sinnvoll sein. Insbesondere Mitarbeitende, die mit besonders sensiblen Daten arbeiten oder spezielle Befugnisse haben und entsprechend häufiger potenzielle Ziele von Spearphishing- und Whaling-Angriffen sind, könnten davon profitieren.

ESET setzt seit über 25 Jahren auf KI

Die meisten der zuvor erwähnten Anwendungsmöglichkeiten von KI werden bereits in den Sicherheitsprodukten vieler Anbieter eingesetzt oder sind in Planung. Im Folgenden erfahren Sie, wie wir im Laufe der Jahre auf verschiedene Weisen KI-Modelle in unsere Technologien implementiert haben. Und natürlich ist diese Entwicklung noch lange nicht am Ende.

1997: Erste Experimente – Erkennung von Makroviren

Bereits in den frühen Jahren unseres Unternehmens experimentierten unsere Entwickler mit Künstlicher Intelligenz und Machine Learning (ML) Modellen. 1997 setzten wir erstmals neuronale Netze in unseren Produkten ein, um die Erkennung von Makroviren zu verbessern. Das war nur der Startschuss, um zu testen, ob die Nutzung von KI und ML beim Schutz vor digitalen Bedrohungen überhaupt sinnvoll ist.

2005: DNA-Erkennungen

Der nächste Meilenstein war die Entwicklung sogenannter DNA-Erkennungen. Angelehnt an die Biologie werden verschiedene Eigenschaften eines Samples im Sinne von Genen herausgearbeitet und zu einem DNA-Profil bzw. einer DNA-Erkennung zusammengefasst. Je nach Eigenschaften und Verhaltensweisen werden diese Profile in gutartig und bösartig unterteilt und in einer Datenbank gespeichert. Alle weiteren geprüften Samples werden anschließend mit den vorhandenen DNA-Erkennungen abgeglichen und entsprechend als gutartig oder schädlich eingestuft. Sofern es keine Übereinstimmung gibt, wird eine neue Erkennung angelegt. Die Aktualisierung des Modells erfolgt sowohl automatisiert als auch mithilfe unserer Malware-Experten.

2006: Backend-Systeme zur Massenverarbeitung

Kurz nach der Einführung der DNA-Erkennungen haben wir KI-gestützte Systeme im Backend eingeführt, die in der Lage sind, Hunderttausende von Samples täglich zu verarbeiten. Auch heute noch bilden diese Systeme das Rückgrat unserer Technologien und unterstützen die Malware-Experten in den meisten Fällen bei der Sichtung, Filterung und Kennzeichnung des eingehenden Materials.

2010: ESET LiveGrid®

Der nächste Schritt bei der Weiterentwicklung unserer Erkennungsmethoden bestand darin, die KI-Systeme mit Cloud-Technologien zu verbinden. So entstand unser cloudbasiertes Reputationssystem ESET LiveGrid®, das Daten in Echtzeit erhältund verarbeitet, um Nutzer innerhalb von wenigen Minuten mit Updates zu versorgen.

Meilensteine unserer KI-basierten Entwicklung

1997

Erster Einsatz von neuronalen Netzen in ESET Produkten zur Erkennung von Makroviren.

2005

DNA-Erkennungen ermöglichen die Identifizierung von aktuellen und neuartigen Bedrohungen anhand ihres Verhaltens.

2010

Unser Reputationssystem ESET LiveGrid® verbindet KI mit Cloud-Technologien und beschleunigt die Bereitstellung von Updates.

2017

Erweitertes Machine Learning in der Cloud nutzt KI-Technologien für unsere automatisierten Erkennungssysteme.

2018

Unsere KI-gestützte Cloud Sandbox ESET LiveGuard® ermöglicht automatisierte und On-Demand-Analysen von unbekannten Samples innerhalb weniger Minuten.

2019

Erweitertes Machine Learning auf den Endpoints verwendet ebenfalls KI-Technologien für unsere automatisierten Erkennungssysteme.

2020-2021

Transformer-basierte Modelle werden in unseren Cloud- und Endpoint-Lösungen eingesetzt.

2023

Der Incident Creator in ESET Inspect nutzt KI, um verschiedene Ereignisse miteinander zu korrelieren und zu Vorfällen zusammenzufassen.

2024

Der generative KI-Assistent ESET AI Advisor erstellt bei Bedarf detaillierte Beschreibungen von Vorfällen und beantwortet alle Fragen rund um erkannte Ereignisse und Vorfälle.

2017–2019: ESET LIVESENSE®

In dieser Zeit eroberten Deep Learning-Algorithmen die digitale Welt im Sturm. Viele aufstrebende IT-Sicherheitsanbieter nutzten den Hype und sahen in dieser Technologie das Allheilmittel zur Lösung aller Probleme in der Cybersecurity. Doch schon bald erkannte man, dass diese Deep Learning-Systeme ohne menschliche Übersicht zwar alle möglichen Angriffsszenarien identifizieren und stoppen konnten, aber mit der Zeit immer mehr Fehlalarme generierten, bis die IT-Teams damit überflutet wurden.

Natürlich haben auch wir mit diesem neuen Zweig innerhalb der KI experimentiert. Wir haben neuronale Netze mit Long short-term memory (LSTM) in Kombination mit Entscheidungsbäumen und anderen Algorithmen getestet und so neue Schutzschichten für unsere Erkennungs-Engine entwickelt – das erweiterte Machine Learning. 2017 haben wir es in unsere Cloud implementiert, 2019 in die Produkte auf den Endpoints. Diese Technologie sorgt für hohe Erkennungsraten bei minimalen Fehlalarmen und ist Teil unserer Kerntechnologie namens ESET LiveSense®.

2018: ESET LIVEGUARD®

Unsere bis dato gesammelten Erfahrungen ebneten den Weg für eine neue, hochleistungsfähige Cloud Sandbox – ESET LiveGuard® (ehemals ESET Dynamic Threat Defense). Sie kombiniert vier Analysestufen: intelligentes Entpacken und Scannen, erweitertes Machine Learning, experimentelles Erkennungssystem und tiefgehende Verhaltensanalyse. Mit dieser KI-gestützten Technologie werden bis dahin unbekannte Bedrohungen innerhalb weniger Minuten oder gar Sekunden identifiziert und gestoppt.

2020–2021: Einzug von transformer-basierten Modellen

Unsere Experten verfolgten das Aufkommen der Transformer-Modelle und nahmen sie unter die Lupe. Insbesondere testeten sie den Nutzen dieser Systeme für die Erkennung schädlicher Objekte. Aufgrund der Ergebnisse wurde die Technologie 2020 bzw. 2021 in unsere Cloud- und Endpoint-Lösungen aufgenommen.

2023: Der Incident Creator in ESET INSPECT

Auch in unserer Detection and Response-Lösung ESET Inspect kommt KI zum Einsatz – und zwar in Form des sogenannten Incident Creators. Dieses Tool unterstützt IT-Sicherheitsverantwortliche bei der täglichen Arbeit, indem es die Ereignisse auf verschiedenen Endpoints miteinander korreliert und nach Priorität ordnet. Außerdem werden sie in einer visuellen Darstellung zu umfassenden Vorfällen zusammengefasst. Hierdurch lässt sich die Zeit zur Bearbeitung eines Vorfalls enorm verkürzen und die „Alarm Fatigue" (also die Desensibilisierung aufgrund einer Vielzahl von Alarmen) verringern.

Transformer-basierte generative KI

Die Idee, mithilfe von KI-basierten Modellen neue Inhalte zu generieren, gibt es schon seit Jahren und wurde auch in bestimmten Bereichen bereits angewandt, z. B. der computergestützten Chemie.

2017 veröffentlichte Google einen Artikel mit dem Titel „Attention Is All You Need". Darin wurde eine neue Architektur für ML-Modelle vorgestellt, die auf Aufmerksamkeitsmechanismen basiert. Diese Architektur mit dem Namen „Transformer" erwies sich als sehr effektiv bei der Verarbeitung natürlicher Sprache und der Erstellung einer Vielzahl von für Menschen verständlichen Inhalten.

Im Jahr 2022 zogen Modelle wie ChatGPT, Midjourney und DALL-E die Aufmerksamkeit der Öffentlichkeit auf sich, indem sie zeigten, dass Transformer-basierte Modelle mit einer einfachen Benutzereingabe in Form einer Textaufforderung einen kompletten Artikel schreiben, ein realistisches Foto erzeugen und neue Videos produzieren können. Natürlich ist dies nur die berühmte Spitze des Eisbergs, die es bisher in die Schlagzeilen der Medien geschafft hat.

2024: ESET AI Advisor

Mit dem ESET AI Advisor haben wir eine einzigartige Schnittstelle zwischen unseren Sicherheitslösungen und Nutzern geschaffen. Hierbei handelt es sich um einen generativen KI-Assistenten, der dem IT-Sicherheitspersonal für verschiedene Anliegen zur Verfügung steht. Über einfache Texteingaben erhalten Nutzer über das Tool leicht verständliche Zusammenfassungen und Erklärungen zu Bedrohungen, wie z. B. Kontextinformationen und den während eines Vorfalls erkannten Taktiken, Techniken und Verfahren (tactics, techniques and procedures, TTPs).

Mitarbeiter ohne spezifische Fachkenntnisse können sich von ESET AI Advisor sicherheitsrelevante Fragen beantworten lassen. Erfahrene Mitarbeiter können verständliche Übersichten erstellen lassen, um mit anderen Teams und Personen mit unterschiedlichem technischen Verständnis in den Austausch zu treten. ESET AI Advisor kann sogar Anleitungen verfassen, damit bestimmte Mitarbeiter(-gruppen) bei der Vorbeugung oder Behebung von Sicherheitsvorfällen mitwirken können.

Auch für unseren Threat Intelligence Service wird der ESET AI Advisor erfolgreich eingesetzt. Mit diesem Service erhalten Kunden umfangreiche Informationen über Bedrohungsakteure, genutzte TTPs, gängige Angriffsszenarien sowie deren Kontext und Indicators of compromise (IoCs). ESET AI Advisor erleichtert die Suche nach der berühmten Nadel im Heuhaufen, indem Sicherheitsexperten im Handumdrehen einen Überblick über alle wichtigen Informationen erhalten und spezifische Daten schnell finden.

ESET AI Advisor durchsucht IoCs, TTPs und spezifische Daten zu Zeit, Ort sowie Branche und verknüpft sie mit bestimmten Bedrohungsakteuren, um umfassende und gleichzeitig leicht verständliche Zusammenfassungen zu erstellen. Das Tool kann zudem Berichte für bestimmte Zielgruppen wie IT-Mitarbeiter, CISOs oder Mitarbeiter auf Vorstandsebene verfassen. Um möglichen Halluzinationen vorzubeugen (siehe Abschnitt „Halluzinationen in generativen Modellen"), verweist ESET AI Advisor stets auf die Quelldokumente.

Mithilfe von Retrieval-Augmented Generation (RAG) greift ESET AI Advisor auf die Leistung unserer internen Tools und Daten zu, um umfassende Bedrohungs- und Vorfallberichte zu erstellen.

Mit dieser umfassenden Nutzung von KI wird nicht nur der proaktive Schutz vor Bedrohungen gestärkt, sondern auch die Handhabung der Sicherheitslösungen vereinfacht.

Retrieval-Augmented Generation (RAG) ist eine Methode zur Verbesserung der Genauigkeit von Ergebnissen großer Sprachmodelle (LLMs). Hierbei erhält das zugrundeliegende LLM Zugang zu einer Auswahl von Tools, die auf externe Informationsquellen zugreifen können. Das Modell hat damit eine umfangreichere Informationsbasis und kann entsprechend bessere, aktuellere und zuverlässigere Antworten auf bestimmte Aufforderungen und Fragen formulieren.

ChatGPT könnte zum Beispiel mithilfe einer Suchmaschine die neuesten Informationen sammeln, die zum Zeitpunkt einer Anfrage zur Verfügung stehen, anstatt die Antwort nur auf Informationen aufzubauen, die während des Trainings in der Vergangenheit verfügbar waren.

Showstopper für KI

Neuronale Netze, Deep Learning, Verarbeitung natürlicher Sprache, Entscheidungsbäume, Transformer-basierte Modelle, LLMs und im Grunde alle anderen KI-Technologien können in bestimmten Rahmen zur Verbesserung der Cybersicherheit beitragen. Aufgrund unserer langjährigen Erfahrung wissen wir, dass der nutzbringende Einsatz von KI einer Menge Fachwissen bedarf und seine Grenzen hat. Hier sind ein paar Beispiele dafür, die einen erheblichen Einfluss auf den Schutz haben können.

Die Grenzen der KI:

- Fehlalarme und Warnungen mit niedriger Priorität können zu „Alarm Fatigue" führen und Fehlkonfigurationen von Sicherheitsprodukten zur Folgen haben.
- Ohne Überwachung und Optimierung durch Experten können sich ML-Modelle verschlechtern.
- Langfristige Zuverlässigkeit ist für Sicherheitslösungen unerlässlich, aber bei reinen KI-Modellen nicht garantiert.
- Bei generativen KI-Modellen besteht die Gefahr sogenannter Halluzinationen.
- Angemessene IT-Sicherheit erfordert neben KI weitere Schutzebenen und Werkzeuge.

Fehlalarme sind nach wie vor relevant

Stufen Sicherheitsexperten oder KI-gestützte Tools gutartige Dateien oder Ereignisse fälschlicherweise als bösartig ein – „False Positive" genannt – kann das fatale Folgen haben. Unter Umständen kann das für ein Unternehmen sogar schlimmer sein als das Übersehen einer schädlichen Malware – ein „False Negative". Im produzierenden Gewerbe könnte es zum Beispiel zu Produktionsunterbrechungen und Verzögerungen, Schäden am Produkt oder an der Produktionslinie und damit zu finanziellen Verlusten kommen.

Viele Fehlalarme können zudem zu „Alarm Fatigue" beim IT-Sicherheitspersonal führen. Das wiederum kann zur Folge haben, dass die Mitarbeiter entweder sehr viel Zeit mit der Problembehebung verbringen oder aber die Schutzvorkehrungen lockern und damit die Erkennungsraten verringern. Beides hat negative Auswirkungen auf die gesamte Sicherheit einer Organisation und öffnet Bedrohungsakteuren neue Angriffsmöglichkeiten.

Trainingsbedarf und Aktualität von ML- und LLM-Modellen

Als Machine Learning in den 2010er Jahren zum Standardrepertoire der meisten Sicherheitslösungen avancierte, behaupteten einige aufstrebende Anbieter, ihre Modelle könnten aktuelle und künftige Bedrohungen ohne jegliches Update erkennen. In der Praxis zeigte sich jedoch schnell, dass dieser Ansatz zu einer sehr hohen Zahl an Fehlalarmen führte und die Leistungsfähigkeit dieser Lösungen mit der Zeit abnahm. Unserer Erfahrung nach bedarf es einer kontinuierlichen Überwachung und Optimierung solcher ML-Modelle. Die Trainingsdaten sind hierbei der Schlüssel für den positiven Effekt, den sie für die IT-Sicherheit haben.

Bei LLMs sieht das ein bisschen anders aus. Die Sprache als Basis entwickelt sich nicht so schnell weiter, sodass auch die Modelle nicht ständig neu trainiert werden müssen wie ML-Systeme, die für die Erkennung von Schadsoftware genutzt werden. Um allerdings in der Lage zu sein, aktuelle und detaillierte Antworten auf Benutzerfragen zu geben, sollte ein LLM RAG-Methoden nutzen, um die benötigten Informationen online oder aus proprietären Quellen zu

beziehen. Ist diese Schnittstelle fehlkonfiguriert oder manipuliert worden, kann das Modell mit falschen Daten gefüttert werden und verzerrte oder problematische Ergebnisse liefern. Der Verzicht auf solche Methoden hingegen hätte zur Folge, dass bestimmte Antworten oder Details nicht bereitgestellt werden können.

Qualität und langfristige Zuverlässigkeit

Für Cybersicherheit sind konstante Leistung und Zuverlässigkeit entscheidend. Eine KI-gestützte Lösung, die in der einen Woche hervorragende Erkennungsergebnisse und nur wenige Fehlalarme erzielt, in der nächsten Woche aber keine Malware erkennt oder eine Flut von Fehlalarmen verursacht, erhöht die Belastung für das IT-Sicherheitsteam. Eine fachkundige Überprüfung der Sicherheitslösung durch die Entwickler ist daher von entscheidender Bedeutung, um langfristig hohe Erkennungs- und niedrige Fehlalarmraten aufrechtzuerhalten. Auch wenn es mehr Aufwand bedeutet, das Modell vor dem Einsatz anhand der Besonderheiten einer Organisation richtig zu trainieren, ist das einer Flut von Fehlalarmen oder übersehenen Bedrohungen vorzuziehen.

Halluzinationen bei generativen KI-Modellen

Glauben Sie nicht alles, was Sie (online) sehen – diese Regel gilt insbesondere für Inhalte, die eine generative KI erstellt hat. Viele der heutigen KI-Modelle sind in der Lage, das perfekte Wort oder Pixel für eine bestimmte Anfrage zu berechnen, um ein logisch klingendes und glaubwürdiges Ergebnis zu erzeugen. In manchen Fällen kann das aber dazu führen, dass ein Nutzer plausibel erscheinende Resultate erhält, die falsch sind und auf halluzinierten – also ausgedachten – Referenzen, Quellen, Daten, Autoren, Aussagen oder URLs beruhen. Das ist ein weiterer Grund, weshalb eine kontinuierliche Überprüfung durch Experten wichtig ist.

Halluzinationen von generativen KI-Modellen stellen in vielen Bereichen eine Herausforderung dar, insbesondere aber bei der Cybersicherheit. Schließlich können die Ergebnisse von Sample-Analysen, die auf erfundenen Daten beruhen, zu einer falschen Bewertung führen. Ebenso kann eine auf Halluzinationen basierende Interpretation von Bedrohungsdaten schlechte oder gar gefährliche Ratschläge und Entscheidungen zur Folge haben, die möglicherweise die Sicherheit ganzer Organisationen gefährden.

HINWEIS: Für bestimmte Anwendungsfälle sind Halluzinationen bei generativen KI-Modellen von Vorteil. Wenn das Ziel darin besteht, komplett neue Audio-, Video- oder grafische Inhalte zu generieren, braucht der Algorithmus die „kreative Freiheit“, neue Ideen zu produzieren, die über die Trainingsdaten hinausgehen.

(Generative) KI allein wird nicht ausreichen

Der Einsatz generativer KI – oder anderer Modelle – kann in bestimmten Fällen sehr aufwändig sein. Das liegt in der Regel am Trainingsaufbau und den hierfür genutzten Daten, die genau ausgewählt und markiert werden müssen, um die gewünschten Ergebnisse zu erzielen. Es gibt viele Beispiele, wo fehlende Markierungen und Rahmenbedingungen zu schlechten und verzerrten Ergebnissen geführt haben. Auch bei der Cybersicherheit ist das ein wichtiger Knackpunkt: Werden Trainingsdaten nicht sorgfältig ausgewählt und markiert, kann das Modell entweder überempfindlich werden und eine Flut an Fehlalarmen generieren oder aber wichtige Aspekte missachten und tatsächliche Malware übersehen.

Erschwerend kommt hinzu, dass Bedrohungsakteure in der Regel versuchen, ihre Schädlinge unsichtbar zu machen oder harmlos aussehen zu lassen, indem sie sie z. B. verpacken, verschleiern oder verschlüsseln. Ohne geeignete zusätzliche Werkzeuge, Trainings und die Aufsicht von Experten können KI-Modelle mit diesen Schwierigkeiten nicht umgehen. Sie sind nicht ohne weiteres in der Lage, Tarnschichten zu entfernen, um den schädlichen Kern eines Samples freizulegen.

Eine weitere beliebte Methode zur Verschleierung von Malware ist die Aufteilung in mehrere Module. Jedes Modul für sich genommen scheint sauber und erst wenn sich die einzelnen Teile zusammensetzen, tritt die schädliche Wirkung zutage. In diesen Fällen gibt es vor der Ausführung keinerlei Warnsignale und selbst eine gut aufgesetzte KI-Lösung wird diese Dateien aller Wahrscheinlichkeit nach als harmlos einstufen.

Intelligente und anpassungsfähige Angreifer

Moderne Computer können Menschen beim Schach und Go besiegen und werden auch bei der Lösung anderer Aufgaben immer effektiver. Häufig handelt es sich aber um Aufgaben in definierten Umgebungen mit festen Regeln. Bedrohungsakteure hingegen interessieren sich nicht für Vorgaben oder Grenzen und werden ohne Vorwarnung betrügen, manipulieren und das Spielfeld neu definieren.

Aufgrund dieser sich ständig wandelnden Bedrohungslage ist es unmöglich, eine universelle Sicherheitslösung zu entwickeln, die alle aktuellen und künftigen Bedrohungen abwehrt. Daran ändern auch die neuesten KI-Modelle nichts.

Ein gutes Beispiel sind selbstfahrende Autos. Trotz massiver Investitionen in ihre Entwicklung sind diese Fahrzeuge angewiesen auf markierte Objekte wie Verkehrsschilder und Ampeln. Ein Krimineller könnte diese fahrerlosen Fahrzeuge angreifen, indem er Verkehrsschilder verdeckt oder Ampeln in einer für das menschliche Auge nicht erkennbaren Geschwindigkeit blinken lässt. Durch diese Manipulation wären die Autos nicht mehr in der Lage, die richtigen Entscheidungen zu treffen und könnten gar schlimme Unfälle verursachen.

KI im Dienste des Bösen

Aktuelle KI-gestützte Bedrohungen

Unterschätzt man die Möglichkeiten, die KI-Technologien Cyberkriminellen und anderen Bedrohungsakteuren bieten, kann das für Unternehmen und IT-Sicherheitsspezialisten fatale Folgen haben. Deshalb haben wir bereits 2018 die erste Version dieses Whitepapers veröffentlicht und einige potenzielle Angriffsszenarien aufgezeigt, von denen manche mittlerweile zur alltäglichen Realität gehören.

Spam und Betrug

Die Erstellung neuer schädlicher Spam- oder Betrugs-Mails nimmt in dieser Liste einen wichtigen Platz ein. 2018 waren es vor allem KI-gestützte Übersetzungen, die hier einen entscheidenden Beitrag leisteten. Mittlerweile nutzen Angreifer LLMs, um den Schreibstil einer beliebigen Person zu imitieren oder anspruchsvolle Spam- und Betrugskampagnen zu entwerfen, die sich nur schwer allein anhand des Nachrichteninhalts als solche erkennen lassen.

Desinformationskampagnen

Gleiches gilt für Desinformationskampagnen. Früher waren sie ein mühsames Unterfangen, für das man ganze „Troll-Armeen" mit Dutzenden, wenn nicht Hunderten von Menschen benötigte. Mithilfe generativer KI-Modelle ist es heutzutage sehr viel einfacher, einen Online-Artikel mit falschen Informationen, manipulierten Fotos oder Deepfake-Videos anzureichern und zu verbreiten. Dafür werden nunmehr eine Handvoll ausgebildeter Personen benötigt. Nicht zuletzt über Soziale Netzwerke, wo die Menschen oft nur Schlagzeilen und die dazugehörigen Bilder überfliegen, lassen sich solche Kampagnen schnell und leicht verbreiten.

Tarnung krimineller Aktivitäten

Bedrohungsakteure können mithilfe von KI und ML schädliche Infrastrukturen schützen bzw. tarnen. So geschehen ist dies bereits bei Emotet – einem berüchtigten Botnetz. Um eine Erkennung zu vermeiden, wurden alle potenziellen Opfer dahingehend überprüft, ob eine Sicherheitslösung vorhanden ist. Man kann davon ausgehen, dass die Angreifer dafür von ML-Modellen Gebrauch machten, denn ansonsten wäre diese Herangehensweise enorm aufwändig gewesen.

Gefälschte E-Mail-Antworten

Auch das Spearphishing wird deutlich lukrativer, wenn man ein LLM zu Hilfe nimmt. Füttert man ein solches Tool mit den E-Mails und anderen Informationen eines potenziellen Opfers, kann es im Handumdrehen eine täuschend echt aussehende Nachricht verfassen. Wird diese Nachricht dann in eine bestehende Konversation des Opfers eingeschleust – diese Technik nennt man reply-chain attack – ist die Wahrscheinlichkeit für den Erfolg des Angriffs relativ hoch.

KI-gestützte Bedrohungen, die 2018 erwartet wurden:

- Erstellung von Social Engineering-Kampagnen und Spear-Phishing
- Optimierung von Malware, einschließlich ihrer Anpassung an spezifische Umgebungen
- Implementierung und Verbreitung falscher Hinweise
- Optimierung der Opferauswahl und des Targetings
- Suche nach neuen Schwachstellen in Software und Smart Devices
- Erstellung neuer Malware oder Übertragung in verschiedene Programmiersprachen
- Auslösen selbstzerstörerischer Mechanismen in der Malware, um Untersuchungen und Analysen zu vereiteln
- Verkürzung der Angriffszeit, um die Reaktionszeit der Opfer zu verkürzen
- Kollektives Lernen von (IoT-)Botnets

Weitere KI-gestützte Bedrohungen, die heute und künftig erwartet werden:

- Erstellung einer großen Menge an hochwertigen Spam-, Betrugs- und Phishing-Kampagnen
- Generierung einer großen Menge an Falsch- und Desinformationen, einschließlich Bildern und Deepfake-Videos, um Opfer zu beeinflussen, zu betrügen oder zu erpressen
- Analyse des Netzwerkverkehrs und der Tastatureingaben von kompromittierten Geräten, um anschließend schädliche Infrastrukturen, Codes und Operationen zu verschleiern
- Extrahieren rechtlich geschützter oder anderweitig sensibler Informationen aus generativen KI-Modellen mithilfe spezieller Eingabeaufforderungen
- Weitere Optimierungen von Social Engineering-Kampagnen durch Ausnutzung der menschenähnlichen Kommunikationsfähigkeit von LLMs

Erstellung neuer Malware

Nicht alle Bedrohungsszenarien sind tatsächlich so real, wie es in manchen Schlagzeilen anmutet – ein Beispiel dafür ist das Erstellen von Malware durch eine KI von Grund auf. Bislang sind die Programmierfähigkeiten von KI-Modellen begrenzt. Einige der aktuellen generativen KI-Modelle können für spezifische Aufgaben wie die Übersetzung von Bibliotheken in andere Sprachen, Debugging, Code-Optimierung und vielleicht sogar die Erstellung einer einfachen, konkret spezifizierten Funktion nützlich sein. Bei der Erstellung komplexer Tools oder Software – auch jene, die für destruktive Zwecke genutzt werden könnten – sind die Ergebnisse der KI jedoch nicht optimal.

Selbst wenn es künftig KI-Modelle geben wird, die hochwertige Malware schreiben, bedeutet das nicht gleich unseren Untergang. Schließlich handelt es sich hierbei nur um einen von vielen Schritten auf dem Weg zu einer effektiven und für die Kriminellen profitablen Bedrohung. Angreifer müssen Strategien für die Verbreitung und die Vermeidung einer Erkennung durch Sicherheitslösungen ausarbeiten. Sie müssen sich überlegen, wie sie die Malware nutzen können, um an Geld zu gelangen. Unter Umständen bedarf es einer weiteren Kommunikation mit dem Opfer. KI kann zwar bei einigen dieser Schritte nützlich sein, aber sie kann den intelligenten menschlichen Angreifer nicht vollständig ersetzen – zumindest bislang.

Science-Fiction oder nahe Zukunft?

Wir möchten betonen, dass bei der derzeitigen Entwicklungsgeschwindigkeit im Bereich der KI die Modelle in den kommenden Jahren oder sogar Monaten vermutlich in allen oben genannten Bereichen besser werden. Das führt uns zu den Science-Fiction-Szenarien, die bislang noch nicht eingetreten sind, aber in absehbarer Zeit Realität werden könnten.

Platzierung falscher Indizien

Cyberkriminelle könnten ihre generativen KI-Modelle mit den veröffentlichten Informationen von Sicherheitsexperten über die Aktivitäten anderer Bedrohungsakteure trainieren und anschließend Kampagnen unter falscher Flagge durchführen. Das würde die ohnehin schon schwierige Zuordnung von Cyberangriffen zu bestimmten Gruppen noch komplizierter machen.

Auf der Jagd nach Schwachstellen

Bereits seit einigen Jahren sehen wir, dass Zero-Day-Schwachstellen ein lukratives Geschäft für Cyberkriminelle sind – sowohl für diejenigen, die auf das Abgreifen von Informationen aus sind als auch für jene, die schnelles Geld machen wollen. Werden KI-Modelle trainiert, um unbekannte, ausnutzbare Schwachstellen zu finden, könnte das die (Hinter-)Türen zu nahezu jeder IT-Umgebung auf dem Planeten öffnen. Und je mehr Smart Devices in einem Netzwerk sind, desto anfälliger ist die Infrastruktur, da diese Geräte häufig unsicher und schwer zu patchen sind.

Verbesserte Opferauswahl

KI könnte genutzt werden, um die interessantesten Ziele für einen Angriff ausfindig zu machen, indem sie die in der Aufklärungsphase eines Angriffs gesammelten Datensätze durchforstet. Sie könnte dabei helfen, leichtgläubige bzw. unvorsichtige Mitarbeiter mit weitreichenden Systemprivilegien oder aber ein Subunternehmen mit schlecht geschützten Systemen zu identifizieren.

Lernende Botnetze

Apropos smarte Geräte: Bedrohungsakteure könnten mithilfe von KI-Modellen neue Botnetze aufbauen, die kollektiv lernfähig sind. Das würde sie noch effizienter machen, sodass sie größere und komplexere Operationen durchführen könnten. Bislang werden Botnetze häufig für DDoS-Angriffe (Distributed Denial of Service) eingesetzt. Künftig wäre auch denkbar, dass sie für die Suche nach Schwachstellen oder das Sammeln von Informationen genutzt werden.

Schlussfolgerung

Künstliche Intelligenz ist für die Cybersicherheit eine überaus nützliche Technologie. In Sicherheitslösungen integriert, kann KI die Erkennungs- und Reaktionsmöglichkeiten verbessern, das Bewusstsein für Gefahren stärken und die Zugänglichkeit von Services wie Threat Intelligence und Threat Hunting optimieren. Zudem kann das „Grundrauschen" von Meldungen und damit potenzielle Alarm Fatigue verringert werden. Dadurch sind IT-Sicherheitsexperten in der Lage, schädliche Aktivitäten besser zu erkennen und schneller darauf zu reagieren.

KI kann und wird wahrscheinlich auch in weiteren Bereichen der Cybersicherheit eine transformative Wirkung haben, z. B. bei der Entwicklung neuer Erkennungsmethoden, der Suche nach unbekannten Schwachstellen und der richtigen Konfiguration von Sicherheitslösungen. Gleichzeitig hat die Technologie ihre Grenzen und Herausforderungen. Sie benötigt qualitativ hochwertige Trainingseinheiten, unter gewissen Umständen ist sie anfällig, hohe Fehlalarmraten zu generieren und sie bedarf immer einer menschlichen, fachkundigen Überprüfung sowie Optimierung.

Natürlich kann **KI auch missbräuchlich eingesetzt werden und Kriminellen ein nützliches Werkzeug sein**. Sie kann verwendet werden, um überzeugende Spam- und Betrugskampagnen zu erstellen, Social Engineering-Methoden zu verbessern, einer Erkennung zu entgehen und sogar Malware zu optimieren – und einige Bedrohungsakteure tun dies bereits. Obwohl diese Entwicklungen besorgniserregend sind, möchten wir betonen, dass **KI nicht in der Lage ist, einen intelligenten menschlichen Angreifer vollständig zu ersetzen**. Das gilt insbesondere bei komplexen Aktivitäten wie der Erstellung ganzer Angriffsketten oder neuer Schadsoftware.

Mit diesem Whitepaper möchten wir unterstreichen, wie wichtig es ist, die Chancen und Risiken von KI im Bereich der Cybersicherheit zu verstehen. Wir vertreten einen ausgewogenen Ansatz, der KI weder verteufelt noch als Allheilmittel verkauft. Unserer Meinung nach ist **die Kombination von KI-Technologien und menschlicher Expertise der richtige Weg, um wirksame und zuverlässige Cybersicherheitslösungen zu entwickeln.**

Anhang A:
Begrifflichkeiten

Künstliche allgemeine Intelligenz

Mit **Künstlicher Allgemeiner Intelligenz** (artificial general intelligence, AGI) wird das bislang unerreichte Ideal eines intelligenten, sich selbst erhaltenden künstlichen Agenten beschrieben, der ohne aktives menschliches Zutun lernt und Entscheidungen trifft. Ein solches System ist in der Lage, ein breites Spektrum von Aufgaben zu erfüllen – im Gegensatz zur „engen" künstlichen Intelligenz, die in der Regel nur begrenzte Aufgaben in einem konkreten Bereich lösen kann.

Künstliche Intelligenz

Die Begrifflichkeit **Künstliche Intelligenz** (KI) bezieht sich auf computerbasierte Agenten, die in Software oder Hardware implementiert sind und in einer vorgegebenen Umgebung intelligent handeln können. Die gezeigte Intelligenz umfasst die Fähigkeiten zu lernen, sich an Veränderungen in der Umgebung anzupassen, die Folgen von Entscheidungen abzuwägen und geeignete Vorgehensweisen auszuwählen, die die aktuellen Ziele, Kenntnisse und Einschränkungen berücksichtigen.

Machine Learning

Machine Learning (ML; auf Deutsch „Maschinelles Lernen") befasst sich hauptsächlich mit Algorithmen, die große Datensätze analysieren und anschließend Vorhersagen für neue Daten erstellen können. Modelle, die von der Funktionsweise der Neuronen im menschlichen Gehirn inspiriert sind, werden als neuronale Netze bezeichnet. Sie sind besonders hilfreich, um komplexe Probleme zu lösen, zu denen es eine Unmenge an Beispieldaten gibt.

Generative künstliche Intelligenz

Fortschritte bei der Verarbeitung natürlicher Sprache sowie Transformer-basierten neuronalen Netzen haben zu Generativer Künstlicher Intelligenz geführt. Solche Modelle werden in der Regel mit großen Mengen unbeschrifteter Daten trainiert. Über einfache Mensch-Maschine-Schnittstellen wie eine simple Eingabemaske sind solche generativen KI-Modelle in der Lage, natürliche Sprache zu verstehen und auf Abruf neue Inhalte zu erschaffen. Dazu gehören Texte, Bilder, Audiodateien, Videos und Quellcode.

Anhang B:
KI – Wo die Realität aufhört und Mythen beginnen

Löst ein bestimmtes Thema einen solchen Hype aus wie derzeit Künstliche Intelligenz, tauchen unweigerlich auch Mythen auf. Die Cybersicherheit ist gegen diesen Trend nicht immun. So gibt es eine ganze Reihe wilder Behauptungen, mit denen diverse Akteure versuchen, Kapital zu schlagen. Im Folgenden werden wir für alle, die sich für den tatsächlichen Stand der Dinge interessieren, einige der aktuellen KI-Behauptungen näher beleuchten.

Behauptung: **KI kann Code analysieren und schädliches Verhalten erkennen**

Realitätsprüfung: Die Behauptung ist zwar nicht ganz falsch, aber die Qualität der Analyse von Malware-Samples aktueller KI-Modelle ist bestenfalls fragwürdig. Ja, von einer generativen KI erstellte Bedrohungsauswertungen lassen sich gut lesen und weisen eine fehlerfreie Grammatik sowie einen einwandfreien Sprachstil auf. Bisweilen sind sie aber unvollständig, fehlerhaft oder aus dem Zusammenhang gerissen und nur Experten mit jahrelanger Erfahrung in der Malware-Analyse sind in der Lage, die Probleme zu erkennen. Nutzen Menschen mit weniger Fachwissen solche Informationen als Grundlage für ihre Entscheidungen, kann das fatale Folgen haben. Erschwerend kommt hinzu, dass Angreifer aktiv versuchen könnten (und wohl auch werden), ihren Code zu verschleiern bzw. so zu ändern, dass das Modell falsche oder unbrauchbare Ergebnisse liefert.

Behauptung: **KI kann eigenständig neue Malware schreiben**

Realitätsprüfung: Einige Online-Dienste nutzen generative KI, um neuen Code zu erstellen. Das ist nützlich und effektiv, wenn es sich um langweilige oder weniger komplexe Aufgaben handelt, die ansonsten die wertvolle Zeit erfahrener Entwickler in Anspruch nehmen würden. Tests haben jedoch gezeigt, dass das Schreiben von Software von Grund auf ein ganz anderes Thema ist und aktuelle KI-Modelle überfordert. Das gilt auch für Malware, insbesondere weil es sich hierbei um ein komplexeres Unterfangen handelt, zu dem auch die Verbreitung des endgültigen „Produkts", der Schutz vor Erkennung und Analyse sowie andere Schritte gehören. Für Angreifer mit mittelmäßigen Programmierkenntnissen ist es viel einfacher, mit Tutorials oder geleaktem Quellcode zu arbeiten, als ein generatives KI-Modell zu verwenden.

Behauptung: **Je größer das KI-Modell, desto besser**

Realitätsprüfung: Eines der Hauptmerkmale von LLMs ist ihre Größe. Einige IT-Sicherheitsanbieter preisen diese Eigenschaft gerne als einen der wichtigsten Vorteile ihrer Malware-Analyse an. Doch mit der Größe des Modells steigen auch die Kosten. Das beginnt bei der erforderlichen Hardware, der Datenmenge und Trainingszeit bis hin bis zum Stromverbrauch sowie Bedarf an anderen Ressourcen. Ein kleineres LLM mit einer spezifischen Aufgabenstellung ist einfacher zu trainieren, zu warten, zu verstehen und zu kontrollieren. In der Cybersicherheit können solche Modelle eingesetzt werden, um große Datenmengen zu verarbeiten und verständliche Ergebnisse wie die Einstufung von Samples in schädlich und gutartig zu liefern.

Behauptung: **KI ist die einzige Sicherheitsebene, die man braucht**

Realitätsprüfung: Wie auch schon andere Technologien zuvor wurde KI von einigen Unternehmen als das Allheilmittel zur Lösung aller Probleme gefeiert. Auch in der Cybersicherheit gab es einige wenige Anbieter, die die seit Jahren bewährten Erkennungstechnologien zugunsten von KI verwerfen wollten. Neuronale Netze, Deep Learning und generative KI sind zwar wichtige Werkzeuge, aber es gibt keinen magischen Algorithmus, der alleine jede erdenkliche Bedrohung erkennen kann. Die Kombination aus mehreren Schutzschichten – wie bei ESET LiveSense® – bietet eine viel bessere Chance, schädliches Verhalten rechtzeitig zu erkennen und zu stoppen.

3 VON ÜBER 400.00 ZUFRIEDENEN KUNDEN

BVB 09 CHAMPION PARTNER

Seit 2019 ein starkes Team auf dem Platz und digital

Allianz Suisse

Seit 2016 durch ESET geschützt
Mehr als 4.000 Postfächer

T..

ISP Security Partner seit 2008
2 Millionen Kunden

BEWÄHRT

Secur|Ty made in EU
Trust Seal www.teletrust.de/itsmie

ESET wurde das Vertrauenssiegel „IT Security made in EU" verliehen

CERTIFIED ISO 9001:2015 COMPANY
CERTIFIED ISO 27001:2013 COMPANY

Unsere Lösungen sind nach Qualitätsstandards zertifiziert

ESET IN ZAHLEN

110.000.000+
Geschützte Nutzer weltweit

195+
Länder & Regionen

400.000+
Geschützte Unternehmen

12
Forschungs- und Entwicklungszentren weltweit

ÜBER ESET

Als europäischer Hersteller mit mehr als 30 Jahren Erfahrung bietet ESET ein breites Portfolio an Sicherheitslösungen für jede Organisationsgröße. Wir schützen betriebssystemübergreifend sämtliche Endpoints und Server mit einer vielfach ausgezeichneten mehrschichtigen Technologie und halten Ihre Infrastruktur mithilfe von Cloud Sandboxing frei von Zero-Day-Bedrohungen. Mittels Multi-Faktor-Authentifizierung und zertifizierter Verschlüsselungslösungen unterstützen wir Sie bei der Umsetzung von Datenschutzbestimmungen sowie Compliance-Maßnahmen.

Unsere Endpoint Detection and Response-Lösung, dedizierte Services wie z.B. Managed Detection and Response und Frühwarnsysteme in Form von Threat Intelligence ergänzen das Angebot im Hinblick auf Incident Management sowie den Schutz vor gezielter Cyberkriminalität und APTs. Dabei setzt ESET nicht allein auf modernste KI-Technologie, sondern kombiniert Erkenntnisse aus der cloudbasierten Reputationsdatenbank ESET LiveGrid® mit Machine Learning und menschlicher Expertise, um Ihnen den besten Schutz zu gewährleisten.

welivesecurity BY ESET

ESET Digital Security Guide

Digital Security
Progress. Protected.

ESET.DE | ESET.AT | ESET.CH

WHITEPAPER

Cybersicherheit im Zeitalter Künstlicher Intelligenz

Ondrej Kubovič und René Holt

mit Unterstützung von:

Juraj Jánošík

Filip Mazán

Peter Košinár

Jakub Debski

Peter Stančík

Digital Security
Progress. Protected.

April 2024

Inhaltsverzeichnis

Digital Security
Progress. Protected.

Einführung

Künstliche Intelligenz (KI) ist aktuell eines der am meisten gehypten Themen – sowohl Presse als auch Verkaufs- und Marketingmaterialien sind voll davon und eine schier endlose Zahl an Online-Diensten setzt ebenfalls auf den Trend. Führte 2019 die Online-Suche nach dem englischen Äquivalent „AI" (Artificial Intelligence) noch zu über 2 Milliarden Treffern, sind es 5 Jahre später fast 18 Milliarden Ergebnisse. Offensichtlich ist also das öffentliche Interesse gewachsen.

Der größte Teil des Hypes lässt sich auf große Sprachmodelle (large language models, LLMs) zurückführen, wodurch die Kommunikationsfähigkeit und die visuelle Kreativität von KI-Systemen bedenklich nahe an die menschlichen Fähigkeiten gekommen sind. Einige sehen darin lukrative Geschäftsmöglichkeiten und eine blühende Zukunft voller KI-Lösungen, andere wiederum fürchten um Millionen von Arbeitsplätzen und sogar die Menschheit selbst.

In diesem Whitepaper werden wir uns von großen Visionen und Untergangsszenarien fernhalten. Wir konzentrieren uns stattdessen auf die **tatsächlichen Vorteile und Risiken dieser Technologie für die Cybersicherheit**.

Wir werden zeigen, wie wir **präzise ausgewählte KI-Algorithmen in unsere Erkennungssysteme integriert** und sie so zu hocheffektiven Schutzlösungen mit maximalen Erkennungsraten und minimalen Fehlalarmen entwickelt haben.

Darüber hinaus erklären wir, wie unsere **Threat Intelligence- und Threat Hunting-Services von KI profitieren**, indem sie unsere Experten u. a. auf interessante Angriffsszenarien und Malware-Merkmale aufmerksam macht.

Uns ist auch bewusst, welchen Schaden KI in den Händen von Cyberkriminellen und staatlich gelenkten Bedrohungsakteuren anrichten kann. Wir werden uns also auch **mit den Gefahren befassen, die mit dieser Technologie einhergehen**. Unser Schwerpunkt liegt dabei auf den generativen Fähigkeiten der KI, also der Möglichkeit, neue Malware zu entwickeln, die Finesse bei der Adressierung der Opfer bei Social Engineering-Kampagnen zu verbessern oder die Qualität und Quantität von Spam-Kampagnen zu steigern.

Im Kampf gegen solche Bedrohungen spielt KI eine zentrale Rolle. **Sie kann riesige Datensätze in Echtzeit analysieren und dabei schnell Muster und Anomalien erkennen**, die auf neue Bedrohungen oder Sicherheitslücken hindeuten. Das wiederum ermöglicht eine rasche Behebung. Für uns als Sicherheitsexperten ist hierbei besonders wichtig, dass **KI zur Gefahrenprävention beiträgt und gleichzeitig eine proaktive und adaptive Verteidigungsstrategie** unterstützt.

KI im Dienste des Guten

Angesichts komplexer Bedrohungsszenarien sowie einer immer größer werdenden Zahl an Cyberkriminellen und staatlich organisierten Bedrohungsakteuren einerseits und dem anhaltenden Fachkräftemangel in der IT Security-Branche andererseits können KI-basierte Innovationen in der Cybersicherheit einen wichtigen Beitrag leisten.

Bereits seit vielen Jahren wird KI bei der Erkennung von Bedrohungen, Filterung und Analyse entsprechender Daten sowie in verschiedenen Tools eingesetzt. Die Möglichkeiten sind aber bei Weitem noch nicht ausgeschöpft. Besonders deutlich wird das am Aufkommen generativer KI und großer Sprachmodelle (LLMs), die eine am Nutzer orientierte Kommunikation ermöglichen und auch im Bereich der IT-Sicherheit dessen Fragen angemessen und verständlich beantworten können. Davon profitieren Entwickler und Spezialisten genauso wie Anwender mit weniger Wissen über Cybersicherheit.

Aktueller Einsatz von KI bei der Bedrohungsabwehr:

- Verarbeitung riesiger Datenmengen zur Erkennung von Angriffen durch Korrelation verschiedener Indikatoren.
- Identifizierung und Analyse von Programmen mit verdächtigem oder schadhaftem Code und auffälligem Verhalten.
- Überwachung und Analyse des Netzwerkverkehrs auf bösartige oder auffällige Muster.
- Beschreibung und Erläuterung komplexer Bedrohungsinformationen, um sie leichter zugänglich zu machen.
- Priorisierung von Warnmeldungen, damit sich die Sicherheitsexperten auf die kritischsten Vorfälle konzentrieren können.
- Bereitstellung ergänzender Funktionen für andere Schutzschichten.

Einsatz von KI bei der Bedrohungsabwehr, der derzeit entwickelt wird oder geplant ist:

- Erstellung neuer Erkennungen anhand der Beschreibungen von Bedrohungen.
- Analyse, Kontextualisierung und Erklärung von vergangenen oder aktuellen Ereignissen in einer konkreten Umgebung.
- Prüfung der Unternehmensumgebung auf versteckte oder unbekannte Schwachstellen.

Was IT-Sicherheitsexperten über KI wissen müssen:

- Hohe Raten an Fehlalarmen einer allein auf KI gestützten IT-Sicherheit schränken ihre Nutzbarkeit ein.
- Ohne Updates und menschliche Aufsicht wird sich ein KI-basierter Schutz aller Wahrscheinlichkeit nach stetig verschlechtern.
- Qualitativ hochwertige Trainingssätze führen zu qualitativ hochwertigen KI-Modellen.

5

Ausgewählte Beispiele:

Erläuterungen zu Bedrohungsdaten

LLMs sind in der Lage, aus langen und komplizierten Inhalten die wichtigsten Informationen herauszufiltern und zusammenzufassen. Das kann insbesondere für Endnutzer, IT-Administratoren und Entscheidungsträger sehr nützlich sein. In der Regel fällt es diesen Gruppen schwer, die hochtechnischen Berichte von Sicherheitsexperten zu verstehen. Mithilfe generativer KI-Modelle können im Handumdrehen kurze, leicht verständliche Texte mit allen relevanten Bedrohungsinformationen und passenden Handlungsempfehlungen erstellt werden.

KI-Assistenten in Produkten

Immer wieder entstehen Schwachstellen durch falsch konfigurierte Systeme und Sicherheitsprodukte in Unternehmensumgebungen und bieten Kriminellen ein potenzielles Einfallstor. Integrierte KI-Assistenten können Organisationen bei der Einrichtung ihrer Umgebung und passenden Konfiguration der Schutzlösungen unterstützen. Das führt nicht nur zu einer verbesserten Benutzererfahrung, sondern steigert unter Umständen auch die Schutzfunktion. Darüber hinaus können KI-Assistenten definierte Aufgaben übernehmen und z. B. eine Ausnahme zur Firewall-Regel hinzufügen oder bestimmte Einstellungen ändern. Zudem können sie dem IT-Administrator bei der Behebung von Sicherheits- oder Systemmeldungen helfen.

KI für den richtigen Fokus

Auch IT-Sicherheitsexperten können von KI-Modellen enorm profitieren. Heutzutage sind solche Teams konfrontiert mit zunehmend komplexen Umgebungen und einer entsprechend großen potenziellen Angriffsfläche, einer wachsenden Anzahl an Sicherheitstools und vor allem einer Flut an Daten. Die Verwaltung aller Aufgaben ist mühsam, zeitaufwändig und erfordert umfassendes Fachwissen. KI kann einen Teil der Belastung abfangen. Sie kann erkannte Ereignisse nach Priorität ordnen und dringende Probleme hervorheben, den Kontext bereitstellen und weniger relevante Alarme selbst lösen. Darüber hinaus kann die KI verschiedene Ereignisse in einen Zusammenhang bringen und dynamische Karten erstellen, die einen umfassenden Überblick über Vorfälle bieten und deren Untersuchung erleichtern. Dadurch können sich Mitarbeitende auf die relevanten Dinge fokussieren und effektiver arbeiten.

Erstellung neuer Erkennungen

Täglich werden unzählige neue Bedrohungsakteure und Angriffstechniken aufgedeckt. Ein KI-Modell kann öffentliche Informationen und Indicators of compromise (IoCs) mit Data Feeds sowie dem Input der Experten kombinieren und Erkennungen generieren. So kann die Unternehmensumgebung auf diese neuesten Bedrohungen gescannt werden. Zur Vermeidung von Fehlalarmen und zur besseren Einordnung von Vorfällen müssen solche Erkennungen jedoch noch von Experten auf ihre Genauigkeit und Wirksamkeit überprüft werden.

Verbesserung des Sicherheitsbewusstseins

Technische Maßnahmen allein reichen für eine gute IT-Sicherheit nicht aus, geschulte Mitarbeiter bzw. Nutzer mit einem entsprechenden Bewusstsein tragen ebenfalls dazu bei. Auch hier kann KI in vielfältiger Weise unterstützen – angefangen bei der Erstellung von kurzen Zusammenfassungen oder Infografiken über Medienartikel zu den neuesten Bedrohungen. Sie können Phishing-Mails für interne Tests erstellen sowie Ergebnisse und Risiken darlegen. Für eine spielerische Sensibilisierung lassen sich Quizfragen und andere Schulungsmaterialien entwerfen, die die interne Kultur, Produkte, Prozesse und andere Besonderheiten des jeweiligen Unternehmens berücksichtigen.

Mit KI erweitertes Sandboxing

Generative KI kann bei der Untersuchung von potenziellen Bedrohungen in Sandbox-Umgebungen nützlich sein. Sie kann beispielsweise Texte in Screenshots oder Low-Level-Ereignisse wie Änderungen in der Windows Registry analysieren und leicht verständliche Beschreibungen des Verhaltens und der Fähigkeiten einer Bedrohung verfassen.

Erweiterter Schutz vor Phishing und Spam

Spam- und Phishing-Mails sind jedem Nutzer und Sicherheitsverantwortlichen ein Dorn im Auge. KI-Modelle könnten genutzt werden, um anhand der bisherigen E-Mail-Kommunikation eines bestimmten Benutzers trainiert zu werden und anschließend auffälliges Verhalten zu erkennen. Ist die KI in der Lage, eine plötzliche Änderung der Schreibgewohnheiten oder des Nachrichteninhalts zu erkennen, können z. B. Angriffe via E-Mail-Antwortketten (reply-chain attack) verhindert werden. Bei solchen Angriffen lassen Kriminelle ihre schädlichen E-Mails vertrauenswürdig aussehen, indem sie auf eine bereits bestehende E-Mail-Kommunikation ihres Opfers antworten.

Eine derartige Schutzlösung ist für die breite Masse aufgrund des hohen Trainingsaufwands des KI-Modells zwar zu kostspielig, aber für ausgewählte Personen könnte solch eine Lösung sinnvoll sein. Insbesondere Mitarbeitende, die mit besonders sensiblen Daten arbeiten oder spezielle Befugnisse haben und entsprechend häufiger potenzielle Ziele von Spearphishing- und Whaling-Angriffen sind, könnten davon profitieren.

ESET setzt seit über 25 Jahren auf KI

Die meisten der zuvor erwähnten Anwendungsmöglichkeiten von KI werden bereits in den Sicherheitsprodukten vieler Anbieter eingesetzt oder sind in Planung. Im Folgenden erfahren Sie, wie wir im Laufe der Jahre auf verschiedene Weisen KI-Modelle in unsere Technologien implementiert haben. Und natürlich ist diese Entwicklung noch lange nicht am Ende.

1997: Erste Experimente – Erkennung von Makroviren

Bereits in den frühen Jahren unseres Unternehmens experimentierten unsere Entwickler mit Künstlicher Intelligenz und Machine Learning (ML) Modellen. 1997 setzten wir erstmals neuronale Netze in unseren Produkten ein, um die Erkennung von Makroviren zu verbessern. Das war nur der Startschuss, um zu testen, ob die Nutzung von KI und ML beim Schutz vor digitalen Bedrohungen überhaupt sinnvoll ist.

2005: DNA-Erkennungen

Der nächste Meilenstein war die Entwicklung sogenannter DNA-Erkennungen. Angelehnt an die Biologie werden verschiedene Eigenschaften eines Samples im Sinne von Genen herausgearbeitet und zu einem DNA-Profil bzw. einer DNA-Erkennung zusammengefasst. Je nach Eigenschaften und Verhaltensweisen werden diese Profile in gutartig und bösartig unterteilt und in einer Datenbank gespeichert. Alle weiteren geprüften Samples werden anschließend mit den vorhandenen DNA-Erkennungen abgeglichen und entsprechend als gutartig oder schädlich eingestuft. Sofern es keine Übereinstimmung gibt, wird eine neue Erkennung angelegt. Die Aktualisierung des Modells erfolgt sowohl automatisiert als auch mithilfe unserer Malware-Experten.

2006: Backend-Systeme zur Massenverarbeitung

Kurz nach der Einführung der DNA-Erkennungen haben wir KI-gestützte Systeme im Backend eingeführt, die in der Lage sind, Hunderttausende von Samples täglich zu verarbeiten. Auch heute noch bilden diese Systeme das Rückgrat unserer Technologien und unterstützen die Malware-Experten in den meisten Fällen bei der Sichtung, Filterung und Kennzeichnung des eingehenden Materials.

2010: ESET LiveGrid®

Der nächste Schritt bei der Weiterentwicklung unserer Erkennungsmethoden bestand darin, die KI-Systeme mit Cloud-Technologien zu verbinden. So entstand unser cloudbasiertes Reputationssystem ESET LiveGrid®, das Daten in Echtzeit erhältund verarbeitet, um Nutzer innerhalb von wenigen Minuten mit Updates zu versorgen.

Weiterführende Informationen:

Meilensteine unserer KI-basierten Entwicklung

1997

Erster Einsatz von neuronalen Netzen in ESET Produkten zur Erkennung von Makroviren.

2005

DNA-Erkennungen ermöglichen die Identifizierung von aktuellen und neuartigen Bedrohungen anhand ihres Verhaltens.

2010

Unser Reputationssystem ESET LiveGrid® verbindet KI mit Cloud-Technologien und beschleunigt die Bereitstellung von Updates.

2017

Erweitertes Machine Learning in der Cloud nutzt KI-Technologien für unsere automatisierten Erkennungssysteme.

2018

Unsere KI-gestützte Cloud Sandbox ESET LiveGuard® ermöglicht automatisierte und On-Demand-Analysen von unbekannten Samples innerhalb weniger Minuten.

2019

Erweitertes Machine Learning auf den Endpoints verwendet ebenfalls KI-Technologien für unsere automatisierten Erkennungssysteme.

2020-2021

Transformer-basierte Modelle werden in unseren Cloud- und Endpoint-Lösungen eingesetzt.

2023

Der Incident Creator in ESET Inspect nutzt KI, um verschiedene Ereignisse miteinander zu korrelieren und zu Vorfällen zusammenzufassen.

2024

Der generative KI-Assistent ESET AI Advisor erstellt bei Bedarf detaillierte Beschreibungen von Vorfällen und beantwortet alle Fragen rund um erkannte Ereignisse und Vorfälle.

2017–2019: ESET LIVESENSE®

In dieser Zeit eroberten Deep Learning-Algorithmen die digitale Welt im Sturm. Viele aufstrebende IT-Sicherheitsanbieter nutzten den Hype und sahen in dieser Technologie das Allheilmittel zur Lösung aller Probleme in der Cybersecurity. Doch schon bald erkannte man, dass diese Deep Learning-Systeme ohne menschliche Übersicht zwar alle möglichen Angriffsszenarien identifizieren und stoppen konnten, aber mit der Zeit immer mehr Fehlalarme generierten, bis die IT-Teams damit überflutet wurden.

Natürlich haben auch wir mit diesem neuen Zweig innerhalb der KI experimentiert. Wir haben neuronale Netze mit Long short-term memory (LSTM) in Kombination mit Entscheidungsbäumen und anderen Algorithmen getestet und so neue Schutzschichten für unsere Erkennungs-Engine entwickelt – das erweiterte Machine Learning. 2017 haben wir es in unsere Cloud implementiert, 2019 in die Produkte auf den Endpoints. Diese Technologie sorgt für hohe Erkennungsraten bei minimalen Fehlalarmen und ist Teil unserer Kerntechnologie namens ESET LiveSense®.

2018: ESET LIVEGUARD®

Unsere bis dato gesammelten Erfahrungen ebneten den Weg für eine neue, hochleistungsfähige Cloud Sandbox – ESET LiveGuard® (ehemals ESET Dynamic Threat Defense). Sie kombiniert vier Analysestufen: intelligentes Entpacken und Scannen, erweitertes Machine Learning, experimentelles Erkennungssystem und tiefgehende Verhaltensanalyse. Mit dieser KI-gestützten Technologie werden bis dahin unbekannte Bedrohungen innerhalb weniger Minuten oder gar Sekunden identifiziert und gestoppt.

2020–2021: Einzug von transformer-basierten Modellen

Unsere Experten verfolgten das Aufkommen der Transformer-Modelle und nahmen sie unter die Lupe. Insbesondere testeten sie den Nutzen dieser Systeme für die Erkennung schädlicher Objekte. Aufgrund der Ergebnisse wurde die Technologie 2020 bzw. 2021 in unsere Cloud- und Endpoint-Lösungen aufgenommen.

2023: Der Incident Creator in ESET INSPECT

Auch in unserer Detection and Response-Lösung ESET Inspect kommt KI zum Einsatz – und zwar in Form des sogenannten Incident Creators. Dieses Tool unterstützt IT-Sicherheitsverantwortliche bei der täglichen Arbeit, indem es die Ereignisse auf verschiedenen Endpoints miteinander korreliert und nach Priorität ordnet. Außerdem werden sie in einer visuellen Darstellung zu umfassenden Vorfällen zusammengefasst. Hierdurch lässt sich die Zeit zur Bearbeitung eines Vorfalls enorm verkürzen und die „Alarm Fatigue" (also die Desensibilisierung aufgrund einer Vielzahl von Alarmen) verringern.

Transformer-basierte generative KI

Die Idee, mithilfe von KI-basierten Modellen neue Inhalte zu generieren, gibt es schon seit Jahren und wurde auch in bestimmten Bereichen bereits angewandt, z. B. der computergestützten Chemie.

2017 veröffentlichte Google einen Artikel mit dem Titel „Attention Is All You Need". Darin wurde eine neue Architektur für ML-Modelle vorgestellt, die auf Aufmerksamkeitsmechanismen basiert. Diese Architektur mit dem Namen „Transformer" erwies sich als sehr effektiv bei der Verarbeitung natürlicher Sprache und der Erstellung einer Vielzahl von für Menschen verständlichen Inhalten.

Im Jahr 2022 zogen Modelle wie ChatGPT, Midjourney und DALL-E die Aufmerksamkeit der Öffentlichkeit auf sich, indem sie zeigten, dass Transformer-basierte Modelle mit einer einfachen Benutzereingabe in Form einer Textaufforderung einen kompletten Artikel schreiben, ein realistisches Foto erzeugen und neue Videos produzieren können. Natürlich ist dies nur die berühmte Spitze des Eisbergs, die es bisher in die Schlagzeilen der Medien geschafft hat.

2024: ESET AI Advisor

Mit dem ESET AI Advisor haben wir eine einzigartige Schnittstelle zwischen unseren Sicherheitslösungen und Nutzern geschaffen. Hierbei handelt es sich um einen generativen KI-Assistenten, der dem IT-Sicherheitspersonal für verschiedene Anliegen zur Verfügung steht. Über einfache Texteingaben erhalten Nutzer über das Tool leicht verständliche Zusammenfassungen und Erklärungen zu Bedrohungen, wie z. B. Kontextinformationen und den während eines Vorfalls erkannten Taktiken, Techniken und Verfahren (tactics, techniques and procedures, TTPs).

Mitarbeiter ohne spezifische Fachkenntnisse können sich von ESET AI Advisor sicherheitsrelevante Fragen beantworten lassen. Erfahrene Mitarbeiter können verständliche Übersichten erstellen lassen, um mit anderen Teams und Personen mit unterschiedlichem technischen Verständnis in den Austausch zu treten. ESET AI Advisor kann sogar Anleitungen verfassen, damit bestimmte Mitarbeiter(-gruppen) bei der Vorbeugung oder Behebung von Sicherheitsvorfällen mitwirken können.

Auch für unseren Threat Intelligence Service wird der ESET AI Advisor erfolgreich eingesetzt. Mit diesem Service erhalten Kunden umfangreiche Informationen über Bedrohungsakteure, genutzte TTPs, gängige Angriffsszenarien sowie deren Kontext und Indicators of compromise (IoCs). ESET AI Advisor erleichtert die Suche nach der berühmten Nadel im Heuhaufen, indem Sicherheitsexperten im Handumdrehen einen Überblick über alle wichtigen Informationen erhalten und spezifische Daten schnell finden.

ESET AI Advisor durchsucht IoCs, TTPs und spezifische Daten zu Zeit, Ort sowie Branche und verknüpft sie mit bestimmten Bedrohungsakteuren, um umfassende und gleichzeitig leicht verständliche Zusammenfassungen zu erstellen. Das Tool kann zudem Berichte für bestimmte Zielgruppen wie IT-Mitarbeiter, CISOs oder Mitarbeiter auf Vorstandsebene verfassen. Um möglichen Halluzinationen vorzubeugen (siehe Abschnitt „Halluzinationen in generativen Modellen“), verweist ESET AI Advisor stets auf die Quelldokumente.

Mithilfe von Retrieval-Augmented Generation (RAG) greift ESET AI Advisor auf die Leistung unserer internen Tools und Daten zu, um umfassende Bedrohungs- und Vorfallberichte zu erstellen.

Mit dieser umfassenden Nutzung von KI wird nicht nur der proaktive Schutz vor Bedrohungen gestärkt, sondern auch die Handhabung der Sicherheitslösungen vereinfacht.

Retrieval-Augmented Generation (RAG) ist eine Methode zur Verbesserung der Genauigkeit von Ergebnissen großer Sprachmodelle (LLMs). Hierbei erhält das zugrundeliegende LLM Zugang zu einer Auswahl von Tools, die auf externe Informationsquellen zugreifen können. Das Modell hat damit eine umfangreichere Informationsbasis und kann entsprechend bessere, aktuellere und zuverlässigere Antworten auf bestimmte Aufforderungen und Fragen formulieren.

ChatGPT könnte zum Beispiel mithilfe einer Suchmaschine die neuesten Informationen sammeln, die zum Zeitpunkt einer Anfrage zur Verfügung stehen, anstatt die Antwort nur auf Informationen aufzubauen, die während des Trainings in der Vergangenheit verfügbar waren.

Showstopper für KI

Neuronale Netze, Deep Learning, Verarbeitung natürlicher Sprache, Entscheidungsbäume, Transformer-basierte Modelle, LLMs und im Grunde alle anderen KI-Technologien können in bestimmten Rahmen zur Verbesserung der Cybersicherheit beitragen. Aufgrund unserer langjährigen Erfahrung wissen wir, dass der nutzbringende Einsatz von KI einer Menge Fachwissen bedarf und seine Grenzen hat. Hier sind ein paar Beispiele dafür, die einen erheblichen Einfluss auf den Schutz haben können.

Die Grenzen der KI:

- Fehlalarme und Warnungen mit niedriger Priorität können zu „Alarm Fatigue" führen und Fehlkonfigurationen von Sicherheitsprodukten zur Folgen haben.
- Ohne Überwachung und Optimierung durch Experten können sich ML-Modelle verschlechtern.
- Langfristige Zuverlässigkeit ist für Sicherheitslösungen unerlässlich, aber bei reinen KI-Modellen nicht garantiert.
- Bei generativen KI-Modellen besteht die Gefahr sogenannter Halluzinationen.
- Angemessene IT-Sicherheit erfordert neben KI weitere Schutzebenen und Werkzeuge.

Fehlalarme sind nach wie vor relevant

Stufen Sicherheitsexperten oder KI-gestützte Tools gutartige Dateien oder Ereignisse fälschlicherweise als bösartig ein – „False Positive" genannt – kann das fatale Folgen haben. Unter Umständen kann das für ein Unternehmen sogar schlimmer sein als das Übersehen einer schädlichen Malware – ein „False Negative". Im produzierenden Gewerbe könnte es zum Beispiel zu Produktionsunterbrechungen und Verzögerungen, Schäden am Produkt oder an der Produktionslinie und damit zu finanziellen Verlusten kommen.

Viele Fehlalarme können zudem zu „Alarm Fatigue" beim IT-Sicherheitspersonal führen. Das wiederum kann zur Folge haben, dass die Mitarbeiter entweder sehr viel Zeit mit der Problembehebung verbringen oder aber die Schutzvorkehrungen lockern und damit die Erkennungsraten verringern. Beides hat negative Auswirkungen auf die gesamte Sicherheit einer Organisation und öffnet Bedrohungsakteuren neue Angriffsmöglichkeiten.

Trainingsbedarf und Aktualität von ML- und LLM-Modellen

Als Machine Learning in den 2010er Jahren zum Standardrepertoire der meisten Sicherheitslösungen avancierte, behaupteten einige aufstrebende Anbieter, ihre Modelle könnten aktuelle und künftige Bedrohungen ohne jegliches Update erkennen. In der Praxis zeigte sich jedoch schnell, dass dieser Ansatz zu einer sehr hohen Zahl an Fehlalarmen führte und die Leistungsfähigkeit dieser Lösungen mit der Zeit abnahm. Unserer Erfahrung nach bedarf es einer kontinuierlichen Überwachung und Optimierung solcher ML-Modelle. Die Trainingsdaten sind hierbei der Schlüssel für den positiven Effekt, den sie für die IT-Sicherheit haben.

Bei LLMs sieht das ein bisschen anders aus. Die Sprache als Basis entwickelt sich nicht so schnell weiter, sodass auch die Modelle nicht ständig neu trainiert werden müssen wie ML-Systeme, die für die Erkennung von Schadsoftware genutzt werden. Um allerdings in der Lage zu sein, aktuelle und detaillierte Antworten auf Benutzerfragen zu geben, sollte ein LLM RAG-Methoden nutzen, um die benötigten Informationen online oder aus proprietären Quellen zu

beziehen. Ist diese Schnittstelle fehlkonfiguriert oder manipuliert worden, kann das Modell mit falschen Daten gefüttert werden und verzerrte oder problematische Ergebnisse liefern. Der Verzicht auf solche Methoden hingegen hätte zur Folge, dass bestimmte Antworten oder Details nicht bereitgestellt werden können.

Qualität und langfristige Zuverlässigkeit

Für Cybersicherheit sind konstante Leistung und Zuverlässigkeit entscheidend. Eine KI-gestützte Lösung, die in der einen Woche hervorragende Erkennungsergebnisse und nur wenige Fehlalarme erzielt, in der nächsten Woche aber keine Malware erkennt oder eine Flut von Fehlalarmen verursacht, erhöht die Belastung für das IT-Sicherheitsteam. Eine fachkundige Überprüfung der Sicherheitslösung durch die Entwickler ist daher von entscheidender Bedeutung, um langfristig hohe Erkennungs- und niedrige Fehlalarmraten aufrechtzuerhalten. Auch wenn es mehr Aufwand bedeutet, das Modell vor dem Einsatz anhand der Besonderheiten einer Organisation richtig zu trainieren, ist das einer Flut von Fehlalarmen oder übersehenen Bedrohungen vorzuziehen.

Halluzinationen bei generativen KI-Modellen

Glauben Sie nicht alles, was Sie (online) sehen – diese Regel gilt insbesondere für Inhalte, die eine generative KI erstellt hat. Viele der heutigen KI-Modelle sind in der Lage, das perfekte Wort oder Pixel für eine bestimmte Anfrage zu berechnen, um ein logisch klingendes und glaubwürdiges Ergebnis zu erzeugen. In manchen Fällen kann das aber dazu führen, dass ein Nutzer plausibel erscheinende Resultate erhält, die falsch sind und auf halluzinierten – also ausgedachten – Referenzen, Quellen, Daten, Autoren, Aussagen oder URLs beruhen. Das ist ein weiterer Grund, weshalb eine kontinuierliche Überprüfung durch Experten wichtig ist.

Halluzinationen von generativen KI-Modellen stellen in vielen Bereichen eine Herausforderung dar, insbesondere aber bei der Cybersicherheit. Schließlich können die Ergebnisse von Sample-Analysen, die auf erfundenen Daten beruhen, zu einer falschen Bewertung führen. Ebenso kann eine auf Halluzinationen basierende Interpretation von Bedrohungsdaten schlechte oder gar gefährliche Ratschläge und Entscheidungen zur Folge haben, die möglicherweise die Sicherheit ganzer Organisationen gefährden.

HINWEIS: Für bestimmte Anwendungsfälle sind Halluzinationen bei generativen KI-Modellen von Vorteil. Wenn das Ziel darin besteht, komplett neue Audio-, Video- oder grafische Inhalte zu generieren, braucht der Algorithmus die „kreative Freiheit“, neue Ideen zu produzieren, die über die Trainingsdaten hinausgehen.

(Generative) KI allein wird nicht ausreichen

Der Einsatz generativer KI – oder anderer Modelle – kann in bestimmten Fällen sehr aufwändig sein. Das liegt in der Regel am Trainingsaufbau und den hierfür genutzten Daten, die genau ausgewählt und markiert werden müssen, um die gewünschten Ergebnisse zu erzielen. Es gibt viele Beispiele, wo fehlende Markierungen und Rahmenbedingungen zu schlechten und verzerrten Ergebnissen geführt haben. Auch bei der Cybersicherheit ist das ein wichtiger Knackpunkt: Werden Trainingsdaten nicht sorgfältig ausgewählt und markiert, kann das Modell entweder überempfindlich werden und eine Flut an Fehlalarmen generieren oder aber wichtige Aspekte missachten und tatsächliche Malware übersehen.

Erschwerend kommt hinzu, dass Bedrohungsakteure in der Regel versuchen, ihre Schädlinge unsichtbar zu machen oder harmlos aussehen zu lassen, indem sie sie z. B. verpacken, verschleiern oder verschlüsseln. Ohne geeignete zusätzliche Werkzeuge, Trainings und die Aufsicht von Experten können KI-Modelle mit diesen Schwierigkeiten nicht umgehen. Sie sind nicht ohne weiteres in der Lage, Tarnschichten zu entfernen, um den schädlichen Kern eines Samples freizulegen.

Eine weitere beliebte Methode zur Verschleierung von Malware ist die Aufteilung in mehrere Module. Jedes Modul für sich genommen scheint sauber und erst wenn sich die einzelnen Teile zusammensetzen, tritt die schädliche Wirkung zutage. In diesen Fällen gibt es vor der Ausführung keinerlei Warnsignale und selbst eine gut aufgesetzte KI-Lösung wird diese Dateien aller Wahrscheinlichkeit nach als harmlos einstufen.

Intelligente und anpassungsfähige Angreifer

Moderne Computer können Menschen beim Schach und Go besiegen und werden auch bei der Lösung anderer Aufgaben immer effektiver. Häufig handelt es sich aber um Aufgaben in definierten Umgebungen mit festen Regeln. Bedrohungsakteure hingegen interessieren sich nicht für Vorgaben oder Grenzen und werden ohne Vorwarnung betrügen, manipulieren und das Spielfeld neu definieren.

Aufgrund dieser sich ständig wandelnden Bedrohungslage ist es unmöglich, eine universelle Sicherheitslösung zu entwickeln, die alle aktuellen und künftigen Bedrohungen abwehrt. Daran ändern auch die neuesten KI-Modelle nichts.

Ein gutes Beispiel sind selbstfahrende Autos. Trotz massiver Investitionen in ihre Entwicklung sind diese Fahrzeuge angewiesen auf markierte Objekte wie Verkehrsschilder und Ampeln. Ein Krimineller könnte diese fahrerlosen Fahrzeuge angreifen, indem er Verkehrsschilder verdeckt oder Ampeln in einer für das menschliche Auge nicht erkennbaren Geschwindigkeit blinken lässt. Durch diese Manipulation wären die Autos nicht mehr in der Lage, die richtigen Entscheidungen zu treffen und könnten gar schlimme Unfälle verursachen.

KI im Dienste des Bösen

Aktuelle KI-gestützte Bedrohungen

Unterschätzt man die Möglichkeiten, die KI-Technologien Cyberkriminellen und anderen Bedrohungsakteuren bieten, kann das für Unternehmen und IT-Sicherheitsspezialisten fatale Folgen haben. Deshalb haben wir bereits 2018 die erste Version dieses Whitepapers veröffentlicht und einige potenzielle Angriffsszenarien aufgezeigt, von denen manche mittlerweile zur alltäglichen Realität gehören.

Spam und Betrug

Die Erstellung neuer schädlicher Spam- oder Betrugs-Mails nimmt in dieser Liste einen wichtigen Platz ein. 2018 waren es vor allem KI-gestützte Übersetzungen, die hier einen entscheidenden Beitrag leisteten. Mittlerweile nutzen Angreifer LLMs, um den Schreibstil einer beliebigen Person zu imitieren oder anspruchsvolle Spam- und Betrugskampagnen zu entwerfen, die sich nur schwer allein anhand des Nachrichteninhalts als solche erkennen lassen.

Desinformationskampagnen

Gleiches gilt für Desinformationskampagnen. Früher waren sie ein mühsames Unterfangen, für das man ganze „Troll-Armeen" mit Dutzenden, wenn nicht Hunderten von Menschen benötigte. Mithilfe generativer KI-Modelle ist es heutzutage sehr viel einfacher, einen Online-Artikel mit falschen Informationen, manipulierten Fotos oder Deepfake-Videos anzureichern und zu verbreiten. Dafür werden nunmehr eine Handvoll ausgebildeter Personen benötigt. Nicht zuletzt über Soziale Netzwerke, wo die Menschen oft nur Schlagzeilen und die dazugehörigen Bilder überfliegen, lassen sich solche Kampagnen schnell und leicht verbreiten.

Tarnung krimineller Aktivitäten

Bedrohungsakteure können mithilfe von KI und ML schädliche Infrastrukturen schützen bzw. tarnen. So geschehen ist dies bereits bei Emotet – einem berüchtigten Botnetz. Um eine Erkennung zu vermeiden, wurden alle potenziellen Opfer dahingehend überprüft, ob eine Sicherheitslösung vorhanden ist. Man kann davon ausgehen, dass die Angreifer dafür von ML-Modellen Gebrauch machten, denn ansonsten wäre diese Herangehensweise enorm aufwändig gewesen.

Gefälschte E-Mail-Antworten

Auch das Spearphishing wird deutlich lukrativer, wenn man ein LLM zu Hilfe nimmt. Füttert man ein solches Tool mit den E-Mails und anderen Informationen eines potenziellen Opfers, kann es im Handumdrehen eine täuschend echt aussehende Nachricht verfassen. Wird diese Nachricht dann in eine bestehende Konversation des Opfers eingeschleust – diese Technik nennt man reply-chain attack – ist die Wahrscheinlichkeit für den Erfolg des Angriffs relativ hoch.

KI-gestützte Bedrohungen, die 2018 erwartet wurden:

- Erstellung von Social Engineering-Kampagnen und Spear-Phishing
- Optimierung von Malware, einschließlich ihrer Anpassung an spezifische Umgebungen
- Implementierung und Verbreitung falscher Hinweise
- Optimierung der Opferauswahl und des Targetings
- Suche nach neuen Schwachstellen in Software und Smart Devices
- Erstellung neuer Malware oder Übertragung in verschiedene Programmiersprachen
- Auslösen selbstzerstörerischer Mechanismen in der Malware, um Untersuchungen und Analysen zu vereiteln
- Verkürzung der Angriffszeit, um die Reaktionszeit der Opfer zu verkürzen
- Kollektives Lernen von (IoT-)Botnets

Weitere KI-gestützte Bedrohungen, die heute und künftig erwartet werden:

- Erstellung einer großen Menge an hochwertigen Spam-, Betrugs- und Phishing-Kampagnen
- Generierung einer großen Menge an Falsch- und Desinformationen, einschließlich Bildern und Deepfake-Videos, um Opfer zu beeinflussen, zu betrügen oder zu erpressen
- Analyse des Netzwerkverkehrs und der Tastatureingaben von kompromittierten Geräten, um anschließend schädliche Infrastrukturen, Codes und Operationen zu verschleiern
- Extrahieren rechtlich geschützter oder anderweitig sensibler Informationen aus generativen KI-Modellen mithilfe spezieller Eingabeaufforderungen
- Weitere Optimierungen von Social Engineering-Kampagnen durch Ausnutzung der menschenähnlichen Kommunikationsfähigkeit von LLMs

Erstellung neuer Malware

Nicht alle Bedrohungsszenarien sind tatsächlich so real, wie es in manchen Schlagzeilen anmutet – ein Beispiel dafür ist das Erstellen von Malware durch eine KI von Grund auf. Bislang sind die Programmierfähigkeiten von KI-Modellen begrenzt. Einige der aktuellen generativen KI-Modelle können für spezifische Aufgaben wie die Übersetzung von Bibliotheken in andere Sprachen, Debugging, Code-Optimierung und vielleicht sogar die Erstellung einer einfachen, konkret spezifizierten Funktion nützlich sein. Bei der Erstellung komplexer Tools oder Software – auch jene, die für destruktive Zwecke genutzt werden könnten – sind die Ergebnisse der KI jedoch nicht optimal.

Selbst wenn es künftig KI-Modelle geben wird, die hochwertige Malware schreiben, bedeutet das nicht gleich unseren Untergang. Schließlich handelt es sich hierbei nur um einen von vielen Schritten auf dem Weg zu einer effektiven und für die Kriminellen profitablen Bedrohung. Angreifer müssen Strategien für die Verbreitung und die Vermeidung einer Erkennung durch Sicherheitslösungen ausarbeiten. Sie müssen sich überlegen, wie sie die Malware nutzen können, um an Geld zu gelangen. Unter Umständen bedarf es einer weiteren Kommunikation mit dem Opfer. KI kann zwar bei einigen dieser Schritte nützlich sein, aber sie kann den intelligenten menschlichen Angreifer nicht vollständig ersetzen – zumindest bislang.

Science-Fiction oder nahe Zukunft?

Wir möchten betonen, dass bei der derzeitigen Entwicklungsgeschwindigkeit im Bereich der KI die Modelle in den kommenden Jahren oder sogar Monaten vermutlich in allen oben genannten Bereichen besser werden. Das führt uns zu den Science-Fiction-Szenarien, die bislang noch nicht eingetreten sind, aber in absehbarer Zeit Realität werden könnten.

Platzierung falscher Indizien

Cyberkriminelle könnten ihre generativen KI-Modelle mit den veröffentlichten Informationen von Sicherheitsexperten über die Aktivitäten anderer Bedrohungsakteure trainieren und anschließend Kampagnen unter falscher Flagge durchführen. Das würde die ohnehin schon schwierige Zuordnung von Cyberangriffen zu bestimmten Gruppen noch komplizierter machen.

Auf der Jagd nach Schwachstellen

Bereits seit einigen Jahren sehen wir, dass Zero-Day-Schwachstellen ein lukratives Geschäft für Cyberkriminelle sind – sowohl für diejenigen, die auf das Abgreifen von Informationen aus sind als auch für jene, die schnelles Geld machen wollen. Werden KI-Modelle trainiert, um unbekannte, ausnutzbare Schwachstellen zu finden, könnte das die (Hinter-)Türen zu nahezu jeder IT-Umgebung auf dem Planeten öffnen. Und je mehr Smart Devices in einem Netzwerk sind, desto anfälliger ist die Infrastruktur, da diese Geräte häufig unsicher und schwer zu patchen sind.

Verbesserte Opferauswahl

KI könnte genutzt werden, um die interessantesten Ziele für einen Angriff ausfindig zu machen, indem sie die in der Aufklärungsphase eines Angriffs gesammelten Datensätze durchforstet. Sie könnte dabei helfen, leichtgläubige bzw. unvorsichtige Mitarbeiter mit weitreichenden Systemprivilegien oder aber ein Subunternehmen mit schlecht geschützten Systemen zu identifizieren.

Lernende Botnetze

Apropos smarte Geräte: Bedrohungsakteure könnten mithilfe von KI-Modellen neue Botnetze aufbauen, die kollektiv lernfähig sind. Das würde sie noch effizienter machen, sodass sie größere und komplexere Operationen durchführen könnten. Bislang werden Botnetze häufig für DDoS-Angriffe (Distributed Denial of Service) eingesetzt. Künftig wäre auch denkbar, dass sie für die Suche nach Schwachstellen oder das Sammeln von Informationen genutzt werden.

Schlussfolgerung

Künstliche Intelligenz ist für die Cybersicherheit eine überaus nützliche Technologie. In Sicherheitslösungen integriert, kann KI die Erkennungs- und Reaktionsmöglichkeiten verbessern, das Bewusstsein für Gefahren stärken und die Zugänglichkeit von Services wie Threat Intelligence und Threat Hunting optimieren. Zudem kann das „Grundrauschen" von Meldungen und damit potenzielle Alarm Fatigue verringert werden. Dadurch sind IT-Sicherheitsexperten in der Lage, schädliche Aktivitäten besser zu erkennen und schneller darauf zu reagieren.

KI kann und wird wahrscheinlich auch in weiteren Bereichen der Cybersicherheit eine transformative Wirkung haben, z. B. bei der Entwicklung neuer Erkennungsmethoden, der Suche nach unbekannten Schwachstellen und der richtigen Konfiguration von Sicherheitslösungen. Gleichzeitig hat die Technologie ihre Grenzen und Herausforderungen. Sie benötigt qualitativ hochwertige Trainingseinheiten, unter gewissen Umständen ist sie anfällig, hohe Fehlalarmraten zu generieren und sie bedarf immer einer menschlichen, fachkundigen Überprüfung sowie Optimierung.

Natürlich kann **KI auch missbräuchlich eingesetzt werden und Kriminellen ein nützliches Werkzeug sein**. Sie kann verwendet werden, um überzeugende Spam- und Betrugskampagnen zu erstellen, Social Engineering-Methoden zu verbessern, einer Erkennung zu entgehen und sogar Malware zu optimieren – und einige Bedrohungsakteure tun dies bereits. Obwohl diese Entwicklungen besorgniserregend sind, möchten wir betonen, dass **KI nicht in der Lage ist, einen intelligenten menschlichen Angreifer vollständig zu ersetzen**. Das gilt insbesondere bei komplexen Aktivitäten wie der Erstellung ganzer Angriffsketten oder neuer Schadsoftware.

Mit diesem Whitepaper möchten wir unterstreichen, wie wichtig es ist, die Chancen und Risiken von KI im Bereich der Cybersicherheit zu verstehen. Wir vertreten einen ausgewogenen Ansatz, der KI weder verteufelt noch als Allheilmittel verkauft. Unserer Meinung nach ist **die Kombination von KI-Technologien und menschlicher Expertise der richtige Weg, um wirksame und zuverlässige Cybersicherheitslösungen zu entwickeln.**

Anhang A: Begrifflichkeiten

Künstliche allgemeine Intelligenz

Mit **Künstlicher Allgemeiner Intelligenz** (artificial general intelligence, AGI) wird das bislang unerreichte Ideal eines intelligenten, sich selbst erhaltenden künstlichen Agenten beschrieben, der ohne aktives menschliches Zutun lernt und Entscheidungen trifft. Ein solches System ist in der Lage, ein breites Spektrum von Aufgaben zu erfüllen – im Gegensatz zur „engen" künstlichen Intelligenz, die in der Regel nur begrenzte Aufgaben in einem konkreten Bereich lösen kann.

Künstliche Intelligenz

Die Begrifflichkeit **Künstliche Intelligenz** (KI) bezieht sich auf computerbasierte Agenten, die in Software oder Hardware implementiert sind und in einer vorgegebenen Umgebung intelligent handeln können. Die gezeigte Intelligenz umfasst die Fähigkeiten zu lernen, sich an Veränderungen in der Umgebung anzupassen, die Folgen von Entscheidungen abzuwägen und geeignete Vorgehensweisen auszuwählen, die die aktuellen Ziele, Kenntnisse und Einschränkungen berücksichtigen.

Machine Learning

Machine Learning (ML; auf Deutsch „Maschinelles Lernen") befasst sich hauptsächlich mit Algorithmen, die große Datensätze analysieren und anschließend Vorhersagen für neue Daten erstellen können. Modelle, die von der Funktionsweise der Neuronen im menschlichen Gehirn inspiriert sind, werden als neuronale Netze bezeichnet. Sie sind besonders hilfreich, um komplexe Probleme zu lösen, zu denen es eine Unmenge an Beispieldaten gibt.

Generative künstliche Intelligenz

Fortschritte bei der Verarbeitung natürlicher Sprache sowie Transformer-basierten neuronalen Netzen haben zu Generativer Künstlicher Intelligenz geführt. Solche Modelle werden in der Regel mit großen Mengen unbeschrifteter Daten trainiert. Über einfache Mensch-Maschine-Schnittstellen wie eine simple Eingabemaske sind solche generativen KI-Modelle in der Lage, natürliche Sprache zu verstehen und auf Abruf neue Inhalte zu erschaffen. Dazu gehören Texte, Bilder, Audiodateien, Videos und Quellcode.

Anhang B:
KI – Wo die Realität aufhört und Mythen beginnen

Löst ein bestimmtes Thema einen solchen Hype aus wie derzeit Künstliche Intelligenz, tauchen unweigerlich auch Mythen auf. Die Cybersicherheit ist gegen diesen Trend nicht immun. So gibt es eine ganze Reihe wilder Behauptungen, mit denen diverse Akteure versuchen, Kapital zu schlagen. Im Folgenden werden wir für alle, die sich für den tatsächlichen Stand der Dinge interessieren, einige der aktuellen KI-Behauptungen näher beleuchten.

Behauptung: **KI kann Code analysieren und schädliches Verhalten erkennen**

Realitätsprüfung: Die Behauptung ist zwar nicht ganz falsch, aber die Qualität der Analyse von Malware-Samples aktueller KI-Modelle ist bestenfalls fragwürdig. Ja, von einer generativen KI erstellte Bedrohungsauswertungen lassen sich gut lesen und weisen eine fehlerfreie Grammatik sowie einen einwandfreien Sprachstil auf. Bisweilen sind sie aber unvollständig, fehlerhaft oder aus dem Zusammenhang gerissen und nur Experten mit jahrelanger Erfahrung in der Malware-Analyse sind in der Lage, die Probleme zu erkennen. Nutzen Menschen mit weniger Fachwissen solche Informationen als Grundlage für ihre Entscheidungen, kann das fatale Folgen haben. Erschwerend kommt hinzu, dass Angreifer aktiv versuchen könnten (und wohl auch werden), ihren Code zu verschleiern bzw. so zu ändern, dass das Modell falsche oder unbrauchbare Ergebnisse liefert.

Behauptung: **KI kann eigenständig neue Malware schreiben**

Realitätsprüfung: Einige Online-Dienste nutzen generative KI, um neuen Code zu erstellen. Das ist nützlich und effektiv, wenn es sich um langweilige oder weniger komplexe Aufgaben handelt, die ansonsten die wertvolle Zeit erfahrener Entwickler in Anspruch nehmen würden. Tests haben jedoch gezeigt, dass das Schreiben von Software von Grund auf ein ganz anderes Thema ist und aktuelle KI-Modelle überfordert. Das gilt auch für Malware, insbesondere weil es sich hierbei um ein komplexeres Unterfangen handelt, zu dem auch die Verbreitung des endgültigen „Produkts", der Schutz vor Erkennung und Analyse sowie andere Schritte gehören. Für Angreifer mit mittelmäßigen Programmierkenntnissen ist es viel einfacher, mit Tutorials oder geleaktem Quellcode zu arbeiten, als ein generatives KI-Modell zu verwenden.

Behauptung: **Je größer das KI-Modell, desto besser**

Realitätsprüfung: Eines der Hauptmerkmale von LLMs ist ihre Größe. Einige IT-Sicherheitsanbieter preisen diese Eigenschaft gerne als einen der wichtigsten Vorteile ihrer Malware-Analyse an. Doch mit der Größe des Modells steigen auch die Kosten. Das beginnt bei der erforderlichen Hardware, der Datenmenge und Trainingszeit bis hin bis zum Stromverbrauch sowie Bedarf an anderen Ressourcen. Ein kleineres LLM mit einer spezifischen Aufgabenstellung ist einfacher zu trainieren, zu warten, zu verstehen und zu kontrollieren. In der Cybersicherheit können solche Modelle eingesetzt werden, um große Datenmengen zu verarbeiten und verständliche Ergebnisse wie die Einstufung von Samples in schädlich und gutartig zu liefern.

Behauptung: **KI ist die einzige Sicherheitsebene, die man braucht**

Realitätsprüfung: Wie auch schon andere Technologien zuvor wurde KI von einigen Unternehmen als das Allheilmittel zur Lösung aller Probleme gefeiert. Auch in der Cybersicherheit gab es einige wenige Anbieter, die die seit Jahren bewährten Erkennungstechnologien zugunsten von KI verwerfen wollten. Neuronale Netze, Deep Learning und generative KI sind zwar wichtige Werkzeuge, aber es gibt keinen magischen Algorithmus, der alleine jede erdenkliche Bedrohung erkennen kann. Die Kombination aus mehreren Schutzschichten – wie bei ESET LiveSense® – bietet eine viel bessere Chance, schädliches Verhalten rechtzeitig zu erkennen und zu stoppen.

3 VON ÜBER 400.00 ZUFRIEDENEN KUNDEN

Seit 2019 ein starkes Team auf dem Platz und digital

Seit 2016 durch ESET geschützt
Mehr als 4.000 Postfächer

ISP Security Partner seit 2008
2 Millionen Kunden

BEWÄHRT

ESET wurde das Vertrauenssiegel „IT Security made in EU" verliehen

Unsere Lösungen sind nach Qualitätsstandards zertifiziert

ÜBER ESET

Als europäischer Hersteller mit mehr als 30 Jahren Erfahrung bietet ESET ein breites Portfolio an Sicherheitslösungen für jede Organisationsgröße. Wir schützen betriebssystemübergreifend sämtliche Endpoints und Server mit einer vielfach ausgezeichneten mehrschichtigen Technologie und halten Ihre Infrastruktur mithilfe von Cloud Sandboxing frei von Zero-Day-Bedrohungen. Mittels Multi-Faktor-Authentifizierung und zertifizierter Verschlüsselungslösungen unterstützen wir Sie bei der Umsetzung von Datenschutzbestimmungen sowie Compliance-Maßnahmen.

Unsere Endpoint Detection and Response-Lösung, dedizierte Services wie z.B. Managed Detection and Response und Frühwarnsysteme in Form von Threat Intelligence ergänzen das Angebot im Hinblick auf Incident Management sowie den Schutz vor gezielter Cyberkriminalität und APTs. Dabei setzt ESET nicht allein auf modernste KI-Technologie, sondern kombiniert Erkenntnisse aus der cloudbasierten Reputationsdatenbank ESET LiveGrid® mit Machine Learning und menschlicher Expertise, um Ihnen den besten Schutz zu gewährleisten.

ESET IN ZAHLEN

110.000.000+
Geschützte Nutzer weltweit

195+
Länder & Regionen

400.000+
Geschützte Unternehmen

12
Forschungs- und Entwicklungs-zentren weltweit

welivesecurity by ESET

ESET Digital Security Guide

Digital Security
Progress. Protected.

ESET.DE | ESET.AT | ESET.CH

Stand: 10/2024 | Druck: 10/2024 | Artikelnummer: M_PRINT2024_22