MENLO SECURITY

Wie KI den modernen Arbeitsplatz gestaltet

Erkenntnisse über die Nutzung generativer KI basierend auf Menlo Telemetry

Inhaltsverzeichnis

Wie KI den modernen Arbeitsplatz prägt

Erkenntnisse über den Einsatz von generativer KI auf der Grundlage von Menlo Telemetry

Künstliche Intelligenz:
Von Science-Fiction zur alltäglichen Realität

Nur wenige Technologien haben eine höhere Produktivität und weniger Aufwand versprochen – und sind diesem Hype so gerecht geworden wie die generative KI (GenAI). Die tatsächlichen Kosteneinsparungen, die sich aus dem Einsatz von GenAI ergeben, lassen sich nicht berechnen, da es inzwischen in Browsern und Anwendungen sowie in einer Fülle von Tools und Diensten integriert ist. Aber da zwischen der Veröffentlichung von ChatGPT Ende 2022 und heute viel Zeit verstrichen ist, sind sich die meisten einig, dass die Gesamtauswirkungen wahrscheinlich in die Hunderte von Milliarden US-Dollar gehen.

Als schnell wachsendes Feld gibt es viele Fragen zur KI, aber eines ist sicher: Es gibt kein Zurück.
Wie Stan Lee einmal sagte: „Aus großer Kraft folgt große Verantwortung." Und obwohl das Sprichwort damals speziell auf die von radioaktiven Spinnen gewonnenen Kräfte angewandt wurde, könnte es nicht wahrer sein als heute in unserer neuen Welt der KI. KI beweist bereits ihre Leistungsfähigkeit im modernen Arbeitsumfeld, indem sie Geschäftsabläufe rationalisiert, Aufgaben automatisiert, bessere Geschäftsentscheidungen ermöglicht und die Kundenerfahrung verbessert. Die Unternehmen haben nun die Verantwortung, sicherzustellen, dass der Einsatz von KI sicher und angemessen ist.

In diesem Beitrag betrachten wir sowohl die Vorteile als auch die Gefahren der KI, basierend auf den von Menlo Security erfassten Telemetriedaten. Wir geben Ihnen einen Einblick in die Möglichkeiten, wie Sie die vielen Vorteile der KI optimieren und gleichzeitig die Fallstricke umgehen können.

Ein einzigartiger Blickwinkel auf die KI-Landschaft

Im Zuge der digitalen Transformation sind viele wichtige Unternehmensanwendungen und -dienste von einem Client/Server-Modell zu einem cloudbasierten und SaaS-fähigen Modell übergegangen. Mit anderen Worten, sie sind über den Webbrowser zugänglich.
Menlo Security konzentriert sich seit über einem Jahrzehnt auf das Thema Browser-Sicherheit. Wir haben miterlebt, wie sich der Browser von einer einfachen Anwendung, die HTML gerendert hat, zu einem eigenständigen Betriebssystem entwickelt hat. Er ist zum Ausgangspunkt für die meisten Anwendungen, Tools und Dienste geworden, einschließlich KI.

> Die starke Abhängigkeit vom Internet hat Menlo Security einen ungehinderten Blick auf die laufenden Veränderungen im Bereich der KI ermöglicht.

Der Einsatz von KI ist so weit fortgeschritten, dass er das allgemeine Muster des Datenverkehrs im Internet verändert hat. Tatsächlich wuchs der Datenverkehr auf Websites von KI-Tools von sieben Milliarden Besuchen im Februar 2024 auf 10,53 Milliarden Besuche im Januar 2025, ein Anstieg um 50 Prozent.[1] Während sich einige KI-Einsätze auf Desktop- oder private Anwendungen verlagern, wird geschätzt, dass bis zu 80 Prozent der GenAI immer noch über den Browser abgerufen werden, sei es über webbasierte KI-Seiten oder die Integration in bestehende Webdienste. Das ist naheliegend, denn:

- Der Browser ist die beliebteste und am weitesten verbreitete Anwendung auf Desktops und mobilen Geräten und gewährleistet eine breite Zugänglichkeit auf verschiedenen Betriebssystemen und Geräten.
- Der Browser interagiert bereits mit anderen Online-Tools und -Diensten, was die Integration in die verschiedenen KI-Dienste erleichtert.
- Die Aufgabe, webbasierte Anwendungen zu entwickeln und bereitzustellen, ist wohl bekannt, sodass ein webbasierter Ansatz die Markteinführung von KI-Diensten beschleunigt.

Das Erste, was Sie über KI wissen müssen, ist, dass KI nicht immer dasselbe bedeutet

Die KI, mit der die meisten von uns vertraut sind, ist von Natur aus reaktiv – sie sucht nach Mustern und Regeln und erfordert ein umfangreiches Management durch den Menschen. Die Antworten auf Anfragen werden aus dem ursprünglichen großen Sprachmodell (LLM) in der KI-Plattform generiert. Es kann neue Informationen, die in Form von Anfragen eingereicht werden, einbeziehen, wenn dies erlaubt ist.

Viele Unternehmen beginnen damit, ihre eigene KI zu entwickeln, indem sie kommerziell verfügbare LLMs nutzen, die auf Open-Source-Modellen basieren, die dann angepasst werden können. Eine solche Implementierung erfordert spezielle Fähigkeiten innerhalb des Unternehmens, bietet aber volle Transparenz – mit anderen Worten, es wird klar sein, woher das KI-Tool seine Informationen hat. Ein weiterer Ansatz besteht darin, ein Closed-Source-LLM von einem Drittanbieter zu erwerben. Diese Methode ist einfacher und schneller zu implementieren und wird vom Anbieter geschützt, aber die Methodik, mit der die Schlussfolgerungen gezogen werden, kann etwas undurchsichtig sein.

Agentic AI basiert jedoch auf einer anderen Prämisse. Agentic AI ist so konzipiert, dass sie proaktiv ist, die Ziele des Benutzers im Kontext versteht und eigenständig Schritte unternimmt, um diese Ziele zu erreichen. Sie kann aus Interaktionen lernen, Daten interpretieren und sich an neue Informationen anpassen, was sie ideal für die Automatisierung komplexer Prozesse macht.

Obwohl agentische KI über den Rahmen dieses Papiers hinausgeht, dient sie dazu, zu veranschaulichen, wie schnell sich die Branche verändert und wie wichtig eine solide Grundlage ist.

Bis 2028 werden 33 % der Unternehmenssoftwareanwendungen agentenbasierte KI enthalten, gegenüber weniger als 1 % im Jahr 2024, wodurch 15 % der täglichen Arbeitsentscheidungen autonom getroffen werden können.[2]

[1] Generative AI's Continued Growth: Traffic to Top 50 AI Tools Up 50% in 11 Months, Similarweb, April 2025
[2] Intelligent Agents in AI Really Can Work Alone. Here's How. Gartner, Oktober 2024

Beobachtungen und Erkenntnisse von Menlo

Während eines 30-tägigen Zeitraums im Mai–Juni 2025 analysierte Menlo Security generative KI-Interaktionen von Hunderten globaler Unternehmen. Wie in der Vergangenheit beobachtet, wird GenAI in Unternehmen jeder Größe und Branche eingesetzt.

Datenverkehrsbeobachtungen nach Branche über einen Zeitraum von 30 Tagen

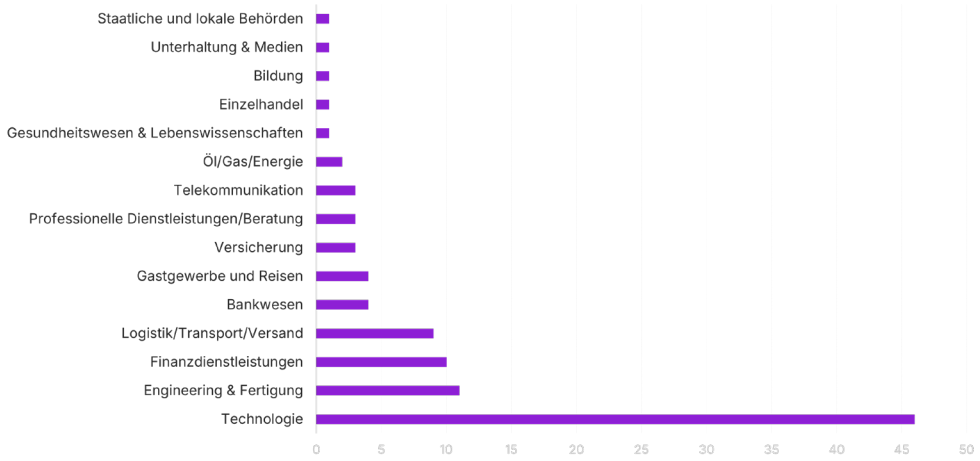

Abbildung 1: Prozentualer Anteil des Datenverkehrs auf Websites, die als GenAI klassifiziert wurden, nach vertikaler Branche, wie von Menlo beobachtet

Der Browser-Datenverkehr zu GenAI-Websites wächst weiter

Der Webverkehr zu Websites, die als GenAI klassifiziert sind, ist weiter gewachsen, wenn auch langsamer als im Jahr 2024. Dies steht im Einklang mit den branchenübergreifenden Beobachtungen. Menlo beobachtete:

> Über 5,6 Millionen Website-Besuche in einem einzigen 30-Tage-Zeitraum.

Die geografische Verbreitung von GenAI könnte ein Vorbote künftiger Probleme sein

Regional gesehen haben wir den stärksten Datenverkehr auf Websites, die als GenAI klassifiziert wurden, in der Region Nord- und Südamerika und insbesondere in den Vereinigten Staaten beobachtet. Der Einsatz von KI hat in den USA stetig zugenommen. Dies liegt möglicherweise daran, dass einige der zugrunde liegenden Technologien, darunter neuronale Netzwerke und Deep Learning, an amerikanischen Universitäten entwickelt wurden und viele Fortschritte, die zu öffentlich verfügbaren GenAI-Tools führten, von US-Unternehmen gemacht wurden. Es ist erwähnenswert, dass die rasche Einführung von GenAI in dieser Region mit einzelnen Nutzern begann, eine Tatsache, die zu eigenen Problemen geführt hat, wie wir später in diesem Bericht erläutern werden.

Während Nord- und Südamerika beim Datenverkehr führend waren, verzeichnete der asiatisch-pazifische Raum das größte Wachstum bei der Einführung von GenAI. China und Indien stehen an der Spitze dieses Trends, wobei 75 % bzw. 73 % der befragten Unternehmen angeben, GenAI implementiert zu haben.[3]

Das Wachstum in der Region Europa, Naher Osten und Afrika (EMEA) hat sich stetig entwickelt, wenn auch in einem deutlich langsameren Tempo als in anderen Regionen. Dieses Ergebnis steht im Einklang mit den Beobachtungen von Branchenanalysten. Das langsamere Wachstum der Region kann teilweise auf die vergleichsweise strengen Vorschriften der EU in Bezug auf Datenschutz und Transparenz zurückgeführt werden.

Erkenntnis 1: Die Nutzung von GenAI beschleunigt sich, und Sie müssen das Steuer übernehmen

Jeder Annahme, dass der Einsatz von KI in Unternehmen bald abflauen würde, hat sich eindeutig als falsch erwiesen. Angesichts des Wachstums und der Diversifizierung der KI-Tools wird es immer dringlicher, eine umfassende Unternehmensstrategie für KI zu entwickeln.

Dies gilt insbesondere, wenn man bedenkt, wie sich das globale Wachstum der KI auf die Compliance mit Datensicherheits- und Datenschutzbestimmungen auswirken könnte. Das KI-Gesetz der EU zum Beispiel ist das erste umfassende KI-Gesetz der Welt, und andere Länder werden folgen. Unternehmen, insbesondere solche mit globaler Präsenz, müssen ernsthaft darüber nachdenken, wie sich der Einsatz von KI auf die Einhaltung bestehender Vorschriften zu Datensicherheit und Datenschutz auswirken wird.

Die Anzahl der KI-Websites und -Anwendungen nimmt zu

Menlo hat mehr als doppelt so viele Domains wie Apps beobachtet

Über 6 500 Domains	Über 3 000 Apps

Die Unterscheidung zwischen Domains und Apps ist wichtig, da eine einzelne App mit mehreren Domains verknüpft sein kann. Angreifer haben sich diese Tatsache zunutze gemacht und eigene „KI“-Seiten erstellt, die Kombinationen aus bekannten App-Namen verwenden, um Legitimität zu verleihen.

Dieser Betrug zeigt sich oft mit Erweiterungen. Solche Bedrohungen wurden kürzlich von MalwareBytes gemeldet, die eine Analyse von Forschern bei Extension Total zitierten, die feststellten, dass es Cyberkriminellen gelungen war, die Konten von mindestens 36 Google Chrome-Erweiterungen zu übernehmen, die KI- und VPN-Dienste anbieten.[4] Zu den kompromittierten Erweiterungen gehören „Bard AI Chat“, „ChatGPT für Google Meet“, „ChatGPT App“, „ChatGPT Quick Access“ und andere.

[3] https://fintechnews.sg/110219/ai/apac-emerges-as-a-leading-genai-adopter/
[4] https://www.malwarebytes.com/blog/news/2025/01/google-chrome-ai-extensions-deliver-info-stealing-malware-in-broad-atta

Top GenAI Apps nach Datenverkehrsaufkommen über einen Zeitraum von 30 Tagen

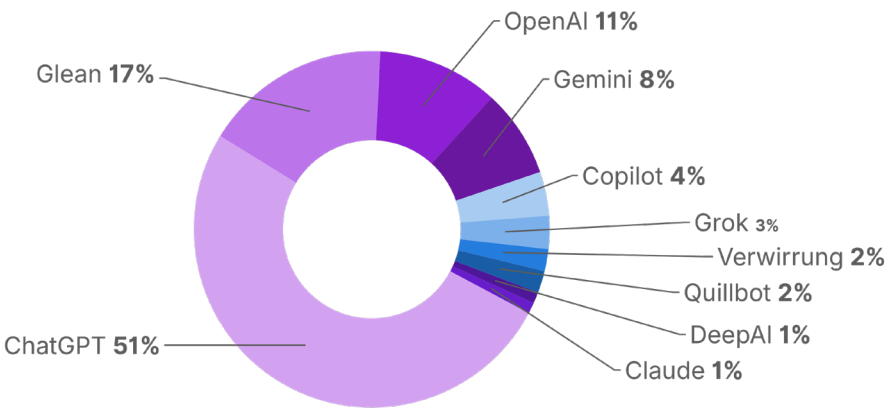

Abbildung 2: Prozentsatz des als GenAI klassifizierten Datenverkehrs nach App, wie von Menlo beobachtet

Bei der Betrachtung von Apps verzeichnete ChatGPT mit Abstand den meisten Datenverkehr. OpenAI steht an dritter Stelle der Beliebtheitsskala, was daran erinnert, dass es auch DALL-E, ein Modell zur Generierung von Text-zu-Bild, und Sora, ein Modell zur Generierung von Text-zu-Video, anbietet. Beide sind als Teil von ChatGPT für zahlende Abonnenten verfügbar.

ChatGPT hat sich als eines der ersten KI-Tools etabliert, die auf breiter Basis verfügbar waren, daher ist seine Dominanz in diesem Bereich nicht überraschend. Bemerkenswert ist jedoch, dass ChatGPT zwar verschiedene Stufen für Unternehmen anbietet, darunter Plus-Abonnenten und Unternehmenskunden, die überwiegende Mehrheit der Nutzer (schätzungsweise über 95 Prozent) jedoch mit dem kostenlosen Angebot arbeitet, das Anfragen und Antworten mit dem Basismodell teilt.

Stand Mai 2025 hat ChatGPT 400 Millionen aktive wöchentliche Nutzer, einschließlich 15,5 Millionen Plus-Abonnenten und 1,5 Millionen Unternehmenskunden.[5]

Bemerkenswert ist auch die hohe Nutzung von Glean, einem Tool zur Aggregation von Unternehmensdaten. Zusätzlich zum allgemeineren Assistententool von Glean bietet das Unternehmen Agenten an, die mit Unternehmensdaten synchronisiert werden und Workflows autonom planen und ausführen können. Das Glean-System verfügt außerdem über Sicherheitsmaßnahmen und eine detaillierte Berechtigungsstruktur, die von entscheidender Bedeutung sind.

[5] https://nerdynav.com/chatgpt-statistics

Auch Angreifer springen auf den KI-Zug auf

Später in diesem Bericht werden wir unsere Beobachtungen über die Verwendung von gefälschten KI-Websites als Lockmittel für Nutzer diskutieren. Fälschungen sind zwar von großer Bedeutung, aber sie sind bei weitem nicht die einzige Methode, mit der Angreifer KI für ihre Zwecke einsetzen. Laut dem Verizon 2025 Data Breach Investigations Report hat sich der Anteil der KI-gestützten bösartigen E-Mails in den letzten zwei Jahren verdoppelt (von etwa fünf Prozent auf 10 Prozent).[6]

Dieser Anstieg steht im Einklang mit den Beobachtungen des Menlo Threat Intelligence Teams. Anfang dieses Jahres fand und analysierte das Team eine Phishing-Kampagne, in der ein Phishing-Kit namens „Greatness“ verwendet wurde. Dieses Kit ist Teil einer Phishing-as-a-Service-Infrastruktur (PhaaS), die über Telegram erworben werden kann. Das Greatness PhaaS umfasst E-Mail-Vorlagen, Zugriff auf das Command-and-Control-Panel (C2), das Kit, Telegram-Support und weitere Telegram-Optionen. Die Möglichkeit, Kampagnen aus einem Bausatz zu erstellen, macht den Prozess noch einfacher.

KI kann Angreifern nicht nur dabei helfen, „bessere“ Phishing-Inhalte zu erstellen, indem sie Rechtschreibung und Aussehen verbessert. Sie kann auch für die Erstellung hyperzielgerichteter „Spear-Phishing“-Angriffe gegen ein bestimmtes Ziel verwendet werden. Menlo Threat Labs fand heraus, dass Bedrohungsakteure zunehmend KI-gestützte Angriffe einsetzen, um Browser-Schwachstellen auszunutzen. Dies führte zu einem Anstieg der Zero-Hour-Phishing-Angriffe um 130 Prozent im Vergleich zum Vorjahr und zu einem Anstieg der GenAI-Imposterseiten,[7] auf die wir später in diesem Bericht näher eingehen. Mit KI können Hacker Phishing-Angriffe Mit KI können Hacker Phishing-Angriffe automatisieren, Nachrichten personalisieren, menschliches Verhalten imitieren und Sicherheitsfilter umgehen, so dass sie schwerer zu entdecken sind.[8]

Eine Taktik, die mit großer Wirkung eingesetzt wird, ist die Kombination von KI mit Open Source Intelligence (OSINT). Traditionell von Verteidigern genutzt, wird OSINT von Sicherheitsteams eingesetzt, um öffentlich zugängliche Informationen über ihre Organisation zu sammeln und Angreifern zuvorzukommen. Die Kombination von OSINT mit KI kann jedoch genutzt werden, um effektive Spear-Phishing-Angriffe zu erstellen, den bisher zeitaufwändigen Prozess des Social Engineering zu automatisieren, überzeugende Deepfakes zu erzeugen und mehr. Wenn sie mit KI-Web-Scraping-Bots kombiniert werden, um Daten zu sammeln, können die Ergebnisse verheerend sein.

Erkenntnis 2: GenAI breitet sich aus und erfährt möglicherweise mehr über Ihr Unternehmen, als es sollte

Mit der zunehmenden Verbreitung und Spezialisierung von GenAI-Websites könnte die Sensibilität der Daten und Ressourcen, die zur Erfüllung ihrer Aufgaben erforderlich sind, zunehmen. Dedizierte KI-Tools, die auf interne Unternehmensanwendungen und -daten zugreifen, sowie solche, die Code korrigieren und schreiben, haben erhebliche Auswirkungen. Zwar deutet alles darauf hin, dass die Anbieter die Datensicherheit und den Datenschutz ernst nehmen, doch ihre Verwendung birgt immer noch mehr Risiken für die Unternehmen, die sie verwalten müssen. Die Verwendung kostenloser Tools, die Abfragen und Antworten mit ihren Basismodellen zu Trainingszwecken austauschen, muss von jedem Unternehmen, das sich mit Datenverlusten befasst, sorgfältig geprüft werden.

[6] Verizon 2025 Data Breach Investigations Report
[7] https://www.menlosecurity.com/resources/state-of-browser-security-report
[8] https://www.webasha.com/blog/how-hackers-use-ai-for-creating-spear-phishing-attacks-the-next-gen-cyber-threat

Um sicherzustellen, dass alle KI-Tools reibungslos zusammenarbeiten, ist es für Unternehmen unerlässlich, eine übergreifende Strategie für Datensicherheit und Datenschutz zu haben. Es ist auch wichtig, mit der Festlegung allgemeiner DLP-Regeln zu beginnen, insbesondere für Werkzeuge, die Zugriff auf sensible Inhalte haben. Die Protokollierung ist sicherlich ein guter erster Schritt, um Klarheit darüber zu schaffen, welche Inhalte wohin gehen, aber die Protokollierung kann nicht die endgültige Maßnahme sein.

Zusätzlich zu dem rasanten Anstieg von Apps und Diensten gibt es auch eine bemerkenswerte Zunahme von Domains. Es ist üblich, dass einer Domäne viele verschiedene URLs zugeordnet sind, aber der Anstieg in der Kategorie „GenAI" erfordert Vorsicht, da URLs, die „nahe" an echten URLs sind, auf die Verwendung einer Angriffstechnik namens „Typosquatting" hinweisen können. Diese Ähnlichkeit erleichtert es böswilligen Akteuren, den Erfolg legitimer Websites auszunutzen, wie Sie später in diesem Bericht sehen werden.

Eine weitere wichtige Beobachtung ist der Einsatz von KI, um „größere, bessere" Phishing- und Malware-Angriffe zu entwickeln. GenAI macht es Angreifern relativ einfach, Details über ihre Ziele zu sammeln. Das Ergebnis sind sehr gezielte Spear-Phishing-Angriffe, die auf einer KI-Analyse der Online-Präsenz des Ziels beruhen, Data Harvesting mit Hilfe von Bots, die Informationen über das Ziel ausspähen und sammeln, Deepfake-Audio- und Video-Kommunikation und sogar gefälschte Chatbots. Weniger glamourös, aber ebenso effektiv nutzen Angreifer GenAI, um JavaScript-Elemente zu erstellen oder umzugestalten, die zum Aufbau dynamischer Webseiten verwendet werden, und entgehen so oft der Entdeckung.

Wie überzeugend sind Deepfakes geworden?

Letztes Jahr wurde ein Finanzangestellter eines internationalen Unternehmens dazu gebracht, 25 Millionen Dollar an Betrüger zu zahlen, die sich mithilfe der Deepfake-Technologie als Mitarbeiter des Unternehmens ausgegeben hatten, darunter auch der CFO.

Die Anfrage wurde nicht in einer E-Mail mit einem schlechten Link oder über eine glaubwürdige Sprachnachricht gestellt. Der Mitarbeiter saß in einer Videokonferenz mit mehreren Personen, die genauso aussahen und klangen wie Kollegen, die er wiedererkannte. Sie alle waren gefälscht.

Die Betrüger hatten die Ausweise dieser Angestellten gestohlen und benutzten KI-Deepfakes, um Gesichtserkennungsprogramme auszutricksen, indem sie die auf den Ausweisen abgebildeten Personen nachahmten.[9]

DLP im Kontext von KI

Laut einer aktuellen Studie glauben 75 % der Kunden, dass generative KI neue Datensicherheitsrisiken einführt,[10] und die Sorge vor Datenverlust oder Datenlecks dominiert. Der Begriff „Datenverlust" bezieht sich auf sensible Daten, die außerhalb einer kontrollierten Umgebung übertragen werden. Data Loss Prevention (DLP) kann mit DLP-Software oder einem Sicherheits-Framework erreicht werden, das den Fluss von sensiblen Daten zwischen Endnutzern und internen Ressourcen kontrolliert.

Während Datenverluste in Unternehmen eine berechtigte Sorge sind, können Datenlecks, bei denen sensible Informationen versehentlich preisgegeben werden, ein größeres Problem darstellen, insbesondere wenn GenAI im Spiel ist. Das liegt daran, dass viele Benutzer in Unternehmen über direkte Texteingabe mit KI-Modellen arbeiten.

[9] https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk
[10] https://www.amplifai.com/blog/generative-ai-statistics#generative-ai-statistic-56

Menlo-Kunden können eine Vielzahl von DLP-Regeln anwenden, z. B. Beschränkungen für Kopier- und Einfügefunktionen, einschließlich der Anzahl der Zeichen, sowie Kontrollen, die eine Ereignisprotokollierung auslösen können, und die Möglichkeit, bestimmte Arten von Inhalten für das Hochladen zu sperren. Dadurch kann Menlo jedes Mal beobachten, wenn diese Kundenrichtlinien eingesehen werden.

IN EINEM EINZIGEN MONAT BEOBACHTETE MENLO:

- **155.005** Kopierversuche
- **313.120** Einfügeversuche
- **327 Besuche** auf mit Watermarking versehenen GenAI-Websites

Einschränkungen beim Kopieren und Einfügen können immer dann angezeigt sein, wenn Unternehmen sich Sorgen über Datenverluste oder -übertragungen machen; im Zusammenhang mit GenAI wird diese Notwendigkeit noch dringlicher. Es kann gut sein, dass die Benutzer nicht beabsichtigen, sensible Informationen zu übertragen, wenn sie versuchen, Inhalte zusammenzufassen oder umzuformulieren, aber es passiert trotzdem. Ein weiteres potenzielles Problem sind die Datenschutzrichtlinien, die solche Beschränkungen vorschreiben können.

Watermarking ist eine Möglichkeit, Benutzer subtil daran zu erinnern, dass sie es mit sensiblen Inhalten zu tun haben könnten. Eine weitere Methode besteht darin, die Anzahl der Zeichen zu begrenzen, die in ein KI-Tool eingegeben werden können.

Natürlich ist Kopieren und Einfügen bei weitem nicht die einzige Art und Weise, wie Nutzerinnen und Nutzer mit GenAI interagieren; einige Kunden von Menlo entscheiden sich auch dafür, Inhalte mit bestimmten DLP-Regeln zu kennzeichnen. Als Menlo den Datenverkehr zu Websites untersuchte, die als GenAI eingestuft wurden, fanden wir Inhalts-Uploads, die DLP-Regeln auslösten.

KUNDENDEFINIERTE REGEL	PROZENTSATZ
SENSIBEL	**70 %**
BESCHRÄNKT	**25 %**
PII	**5 %**

Die Menge der Daten, die ausdrücklich als personenbezogene Daten (Personally Identifiable Information, PII) bezeichnet werden, mag gering erscheinen, aber es ist wichtig zu wissen, dass PII-Daten auch als Teilmenge der viel umfassenderen Klassifizierung „sensible Daten“ auftauchen.

Inhalte, die als „sensibel“ eingestuft sind, können eine Vielzahl von Daten enthalten, darunter:

- PII über direkte Identifikatoren wie vollständige Namen, Sozialversicherungs- und Nationalversicherungsnummern, Passnummern und mehr sowie indirekte Identifikatoren, einschließlich biometrischer Daten wie Fingerabdrücke, Gesichtserkennungsinformationen und Netzhautscans
- Geschützte Gesundheitsinformationen (Protected Health Information, PHI), wie z. B. Krankengeschichten und Laborergebnisse
- Finanzinformationen, einschließlich Kreditkarten-, Bankkonto- und Bankleitzahlen
- Geistiges Eigentum (IP), einschließlich Geschäftsgeheimnissen, proprietärem Code, Kundenlisten und Datenbanken sowie unveröffentlichten Finanz-, Produkt- und Marketingplänen
- Zugangsdaten und Systeminformationen, wie z. B. Benutzernamen und Passwörter, PINs, private Schlüssel und Zertifikate, Netzwerkspezifika und Verwaltungsdetails
- Daten zur Einhaltung gesetzlicher und behördlicher Vorschriften, die sich auf Vorschriften wie DS-GVO, HIPAA, PCIDSS, CCPA und andere beziehen können

„Eingeschränkte" Informationen unterscheiden sich von „sensiblen" Informationen und können allgemein als Informationen definiert werden, die aus einer Vielzahl von Gründen nicht weitergegeben werden sollten.

Die wichtigsten Dateitypen, die mit DLP-Ereignissen im Zusammenhang mit GenAI-Datenverkehr verbunden sind

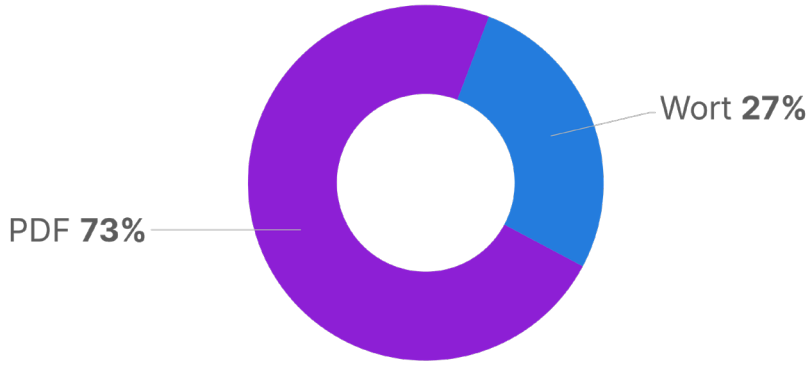

Abbildung 3: DLP-Ereignisse (Zulassen/Blockieren/Protokollieren) nach Dateityp, die von Menlo über einen Zeitraum von 30 Tagen beobachtet wurden

Downloads aus KI-Quellen sollten nicht automatisch als vertrauenswürdig gelten

Inhalte, die auf GenAI-Seiten hochgeladen werden, sind nicht der einzige Bereich, der für das Unternehmen ein Problem darstellen kann. Es ist auch wichtig, die Inhalte zu berücksichtigen, die von solchen Websites heruntergeladen werden können.

Die häufigsten Dateitypen, die von AI-Tools heruntergeladen werden

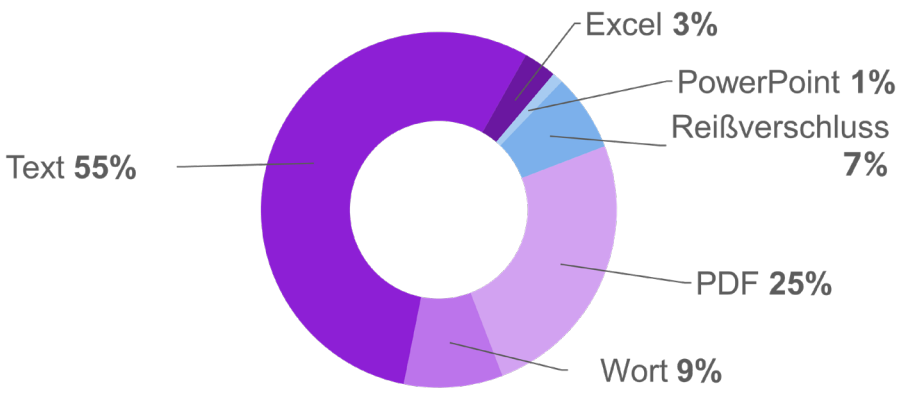

Abbildung 4: Von Menlo über einen Zeitraum von 30 Tagen beobachtete

Dateitypen, die von als GenAI klassifizierten Websites heruntergeladen wurden Dateien, die von KI-Tools in das Unternehmen gelangen, dürfen als Bedrohungsvektor nicht übersehen werden. PDFs werden seit langem genutzt, um eine Vielzahl von Malware zu verbreiten, darunter JavaScript-Exploits, eingebettete ausführbare Dateien oder Phishing-Links.[11] Die Telemetrie von Menlo zeigte, dass die Mehrheit der Downloads von Websites, die als GenAI klassifiziert wurden, entweder PDF- oder Textdateien waren.

[11] https://nordlayer.com/blog/can-pdf-have-virus

Dies ist besorgniserregend, da PDFs die perfekte Umgebung für eingebettete Malware bieten können. Laut dem IBM X-Force 2025 Threat Intelligence Index „... tarnt PDF-Malware bösartige URLs, indem sie diese verschlüsselt, in komprimierten Streams versteckt oder hexadezimale Darstellungen verwendet, was auch die automatisierte Analyse von E-Mail-Sicherheitslösungen behindern kann.“[12] Auch die Häufigkeit von Text- und Word-Dateien steht ganz oben auf der Liste, und die Kombination macht die Inhaltskontrolle zu einem Muss.

Shadow AI hebt DLP auf ein neues Niveau

Die DLP-Probleme werden noch dadurch verschlimmert, dass viele Beschäftigte der „BYOAI“ (Bring Your Own AI) – auch bekannt als Schatten-KI – huldigen und mit der kostenlosen Version der von ihnen gewählten Tools arbeiten.

BYOAI wirft zwei Probleme auf. Erstens verliert das Unternehmen in einer solchen Situation die Kontrolle, und zweitens geben die meisten kostenlosen Dienste an, dass sie die übermittelten Daten zum Trainieren ihrer Modelle verwenden. Eine kürzlich durchgeführte Umfrage ergab, dass fast sieben von zehn (68 Prozent) Unternehmensmitarbeitern, die GenAI bei der Arbeit nutzen, angeben, dass sie über persönliche Konten auf öffentlich verfügbare GenAI-Assistenten wie ChatGPT, Microsoft Copilot oder Google Gemini zugreifen, und mehr als die Hälfte (57 Prozent) hat zugegeben, dass sie sensible Informationen in diese eingegeben haben.[13]

Im Verizon Data Breach Investigations Report 2025 heißt es: „Einige der häufigsten Anwendungsfälle von GenAI-Tools – wie z.B. Zusammenfassungen oder Codierungshilfen – fordern den Nutzer auf, vertrauliche Dokumente und Codebases hochzuladen, um diese zu erreichen.“[14] Im Verizon-Bericht heißt es weiter, dass eine große Anzahl von Nutzern, die GenAI einsetzten, „entweder unternehmensfremde E-Mails als Identifikatoren für ihre Konten verwendeten (72 %) oder ihre Unternehmens-E-Mails ohne integrierte Authentifizierungssysteme nutzten (17 %), was höchstwahrscheinlich auf eine Nutzung außerhalb der Unternehmensrichtlinien hindeutet.“

Das ist einer der Gründe, warum die Nutzung von BYOAI-Tools, die in der Regel kostenlos sind und in der Regel Daten austauschen, so bedenklich sein kann. Die Benutzer versuchen einfach, produktiver zu sein, wenn sie die KI-Tools, die sie zu Hause benutzen, bitten, den Inhalt eines sensiblen Dokuments zusammenzufassen oder Fehler im Code zu finden. Es kann jedoch auch sein, dass sie dabei unwissentlich Informationen weitergeben.

Erkenntnis 3: Sie müssen GenAI so steuern, als ob die Sicherheit Ihres Unternehmens davon abhängt ... denn das tut sie

Die Implementierung von DLP ist ein wesentlicher erster Schritt für die sichere Nutzung von KI. Um wirklich effektiv zu sein, ist es wichtig, dass DLP die beiden häufigsten Arten von Anwendungsfällen abdeckt. Dazu gehören die Fälle, in denen ein Benutzer direkt über Texteingaben mit GenAI interagiert, sowie die Fälle, in denen ein Benutzer ein Dokument, ein Bild oder einen anderen Inhaltstyp hochlädt.

[12] https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/2025-threat-intelligence-index
[13] TELUS Digital Survey
[14] Verizon 2025 Data Breach Investigations Report

Menlo war in der Lage, sowohl die Kategorieklassifizierung als auch die hochgeladenen Dateitypen anzuzeigen, da viele Menlo-Kunden DLP-Regeln für Inhalte hinzugefügt haben, einschließlich Zulassen, Sperren und Protokollieren. Für Unternehmen ist es jedoch unerlässlich geworden, über die einfache Protokollierung von DLP-Verstößen hinauszugehen. Wenn Unternehmen die Arten von Inhalten kennen, die geteilt werden, und wissen, wo sie geteilt werden, können sie Regeln für die Arbeit mit KI aufstellen und kommunizieren und die Benutzer aufklären, die möglicherweise nicht wissen, was ein Datenleck oder eine Datenschutzverletzung für das Unternehmen insgesamt bedeuten könnte. Wenn Sie jetzt mit diesem Prozess beginnen und die Disziplin beibehalten, ist dies für Benutzer und Unternehmen gleichermaßen von Nutzen.

Kontrollen müssen über Upload und Download hinausgehen

In Fällen, in denen eine direkte Texteingabe erlaubt ist, sind einige Schritte erforderlich. Zuerst muss das Unternehmen Einblick in das verwendete KI-Modell haben. Wenn es sich bei dem Modell um ein kostenloses Angebot handelt, auf das über ein persönliches Login zugegriffen wird, sollte es verboten werden. Nicht nur, dass Sicherheits- und IT-Teams in solchen Situationen den Überblick verlieren, diese Modelle trainieren auch oft auf Daten, die von Benutzern übermittelt werden. Als nächstes müssen Unternehmen ein genehmigtes KI-Modell einrichten und Benutzer ausschließlich dorthin schicken.

In Fällen, in denen das KI-Tool direkte Eingaben ermöglicht, müssen detaillierte Steuerungen integriert sein, einschließlich Kopier- und Einfüge-Steuerungen sowie Regelungen für Uploads und Downloads. Kurz gesagt, das Unternehmen muss KI-Tools wie sensible Anwendungen behandeln.

Schatten-KI muss gesteuert oder beseitigt werden.

Jedes DLP-Problem ist besorgniserregend, aber da der zunehmende Einsatz von KI die Kontrolle durch das Unternehmen immer weiter erschwert, wird alles Schlechte nur noch schlimmer werden. Eine einfache Benachrichtigung der Benutzer über die Unternehmensrichtlinien bietet nicht den dringend benötigten Schutz. Um Schatten-KI zu eliminieren, müssen Unternehmen genehmigte KI-Systeme oder -Tools auswählen, denen sie vertrauen – und deren ausschließliche Verwendung anordnen.

Während Unternehmen steuern können (und wohl auch sollten), welche KI-Tools in ihrem Netzwerk verwendet werden, gibt es keine Möglichkeit zu kontrollieren, welche KI-Tools von Mitarbeitern, die BYOD nutzen, oder von Drittanbietern und Partnern verwendet werden. Wenn Unternehmen nicht steuern können, welche KI-Tools außerhalb des Netzwerks verwendet werden, müssen sie steuern, was innerhalb erlaubt ist. Malware muss schon an der Schwelle gestoppt werden.

Nicht alle KI-Websites sind das, was sie zu sein scheinen

Viele der schwerwiegendsten Bedrohungen heutzutage betreffen bösartige Websites, die sich als legitime KI-Websites ausgeben. In seinem Bericht 2025 State of Browser Sicherheit beobachtete Menlo fast 600 Fälle von Phishing-Websites, die URLs verwenden, die eine Verbindung zu legitimen GenAI-Namen implizieren – und dieser Trend verlangsamt sich nicht. Allein im März blockierte Menlo Protect mit HEAT Shield AI mehr als 40 neue Websites, die vorgaben, GenAI-Websites zu sein, aber tatsächlich als Phishing-Websites identifiziert wurden. Darüber hinaus beobachtete Menlo bösartige Websites, die „ChatGPT“ oder „Copilot“ im Domainnamen führten. Im Jahr 2025 begannen Websites, die sich als GenAI ausgaben, in den meisten erfassten Fällen „Gemini“ im Domainnamen zu verwenden.

Wie SecurityWeek berichtet, haben Forscher zwischen 2024 und 2025 Tausende von ähnlich aussehenden Domain-Namen und Imitations-Websites aufgespürt (etwa 2.600 Websites tauchten zwischen dem 1. Dezember 2024 und dem 3. Februar 2025 auf), die DeepSeek imitieren.[15] Ständig tauchen neue Imitations-Domains auf, und während einige offensichtliche Fälschungen sind, sind andere Berichten zufolge schwerer zu erkennen. Dies ist nur einer der Gründe, warum es für Unternehmen unerlässlich ist, den Benutzern nicht mehr zu erlauben, Entscheidungen über KI-Tools selbst zu treffen.

> Im letzten Jahr hat Menlo fast 600 gefälschte Websites mit GenAI-bezogenen Namen identifiziert, die tatsächlich als Phishing-Sites identifiziert wurden.

Einige bösartige Akteure machen sich die KI zunutze, indem sie spezielle „KI"-Tools entwickeln. Ein Beispiel ist eine gefälschte KI-Tool-Website, die sich als App von Novaleads, einer Affiliate-Online-Marketing-Plattform, ausgibt. Die Opfer werden mit dem Versprechen eines kostenlosen 12-monatigen Abonnements verleitet, das Tool herunterzuladen. Sobald es heruntergeladen wurde, verschlüsselt die CyberLock-Ransomware Dateien auf mehreren Festplattenpartitionen.[16]

Umgekehrt sind nicht alle KI-Websites, die Schaden anrichten, automatisch gefälscht. Im Jahr 2025 haben Millionen Menschen DeepSeek heruntergeladen. Mitarbeiter, die dies getan haben, haben die Daten ihres Unternehmens in Gefahr gebracht - eine ungesicherte DeepSeek-Datenbank hat eine Million Zeilen von Log-Streams mit Chat-Verläufen, geheimen Schlüsseln, Backend-Details und anderen hochsensiblen Informationen offengelegt. Das zeigt, dass Unternehmen alle KI-Apps genau prüfen und sichern müssen.[17]

Erweiterungen können ebenso problematisch sein.

Seit der Veröffentlichung von ChatGPT tauchen fast ständig bösartige Browsererweiterungen auf, die angeblich KI-Funktionen bieten. Eines der ersten, „Chat GPT" (beachten Sie das Leerzeichen), behauptete, ChatGPT-Ergebnisse in die Google-Suche zu integrieren. Die Erweiterung, die auf demselben Open-Source-Code wie die echte Erweiterung (ChatGPT for Google) basierte, wurde online beworben und im Chrome Web Store angeboten, bevor sie entdeckt wurde. Ein Aspekt, der die bösartige Erweiterung so überzeugend machte, war, dass sie tatsächlich wie vorgesehen funktionierte; leider sammelte sie jedoch im Hintergrund auch Facebook- und OpenAI-Token, was böswilligen Akteuren den Zugriff auf Benutzerdaten ermöglichte.

Obwohl dies eine der ersten bösartigen Erweiterungen war, die sich als KI tarnten, war es bei weitem nicht die letzte. Im Jahr 2025 wurde eine Reihe von Bedrohungen aufgedeckt, darunter Erweiterungen, die mit bekannten Namen ausgestattet sind, darunter „ChatGPT for Google Meet", „Bard AI Chat", „AI Assistant – ChatGPT and Gemini for Chrome", „Search Copilot AI Assistant for Chrome", „GPT4 Summary with OpenAI" und andere. Diese Erweiterungen haben Datendiebstahlcode eingeschleust, vertrauliche Benutzerdaten gesammelt, Zugangsdaten kompromittiert und mehr.

[15] https://www.securityweek.com/beware-of-deepseek-hype-its-a-breeding-ground-for-scammers/
[16] https://www.bleepingcomputer.com/news/security/cybercriminals-exploit-ai-hype-to-spread-ransomware-malware
[17] https://info.varonis.com/en/state-of-data-security-report-2025

Erkenntnis 4: Gefälschte KI-Tools stellen ein ernsthaftes Risiko dar

Die Präsenz von Phishing-Websites, die vorgeben, KI-Websites zu sein, nimmt zu, teilweise angetrieben durch die große Anzahl bösartiger URLs, die um bekannte Domains herum erstellt wurden. Erweiterungen sind ein weiterer Bereich, in dem Benutzer versuchen können, die Produktivität durch KI zu steigern, indem sie eine Erweiterung installieren, von der sie glauben, dass sie Suchergebnisse verbessert, Inhalte zusammenfasst oder beim Schreiben hilft, neben anderen Vorteilen. Es ist wichtig, dass Unternehmen Anleitungen bereitstellen und Maßnahmen ergreifen, um die Nutzung illegaler oder bösartiger Websites und Erweiterungen zu verhindern.

Zusammenfassung

Wenn KI-Tools den heutigen Unternehmen viel zu bieten haben, sind diese Vorteile jedoch von der sicheren Nutzung der Technologie abhängig. Glücklicherweise können Sie durch die Sicherung von Unternehmensbrowsern und deren Datenverkehr Ihre Sicherheitslage verbessern und eine solide Grundlage für den Umgang mit neuen und aufkommenden Bedrohungen schaffen. Nicht nur die Sicherung des Browsers und des Browser-Datenverkehrs schützt Sie vor den potenziellen Fallstricken der KI, sondern sie erhöht auch die Sicherheit Ihres Unternehmens insgesamt.

Es gibt mehrere wichtige Schritte, die Sie jetzt unternehmen können, um sichere Praxis im Zusammenhang mit der Verwendung von GenAI zu etablieren:

- Eliminieren Sie Schatten-KI
- Regeln für die sichere Nutzung von KI-Tools festlegen und durchsetzen
- Schützen Sie Ihre Daten
- Inhaltsinspektion aktivieren
- Schützen Sie das Unternehmen vor gefälschten KI-Websites und verteidigen Sie sich gegen die zunehmende Nutzung von KI bei Phishing- und anderen Angriffen.
- Schützen Sie den Zugriff auf interne und SaaS-Anwendungen.
- Bieten Sie eine Defense-in-Depth, indem Sie den lokalen Browser absichern.

Für weitere Informationen über diese Schritte und den Ansatz von Menlo, das Beste aus GenAI herauszuholen, kontaktieren Sie uns bitte.

Über Menlo Security

Menlo Security beseitigt ausweichende Bedrohungen und schützt die Produktivität mit dem Menlo Secure Cloud Browser. Menlo erfüllt das Versprechen der cloudbasierten Sicherheit und ermöglicht einen Zero-Trust-Zugriff, der einfach zu implementieren ist. Der Menlo Secure Cloud Browser verhindert Angriffe und macht Cyber-Abwehrmaßnahmen für Endbenutzer unsichtbar, während sie online arbeiten, wodurch die operative Belastung der Sicherheitsteams verringert wird.

Menlo schützt Ihre Benutzer und sichert den Zugriff auf Anwendungen, indem es eine vollständige Browserlösung für Unternehmen bietet. Mit Menlo können Sie Browser-Sicherheitsrichtlinien mit einem einzigen Klick einrichten, den Zugriff auf SaaS und private Anwendungen sichern und Unternehmensdaten „bis zur letzten Meile“ schützen. Sichern Sie Ihre digitale Transformation mit zuverlässigem und bewährtem Cyber-Schutz, auf jedem Browser.

MENLO SECURITY

Mehr erfahren: https://www.menlosecurity.com
Kontakt: ask@menlosecurity.com