Secorvo Security News

August 2025

Sicherheit dank KI

Inzwischen wissen wir, dass LLMs Blender sind (SSN 5/2025), wenn auch sprachlich beeindruckende, die selbst bei der Suche nach Sicherheitslücken halluzinieren (SSN 6/2025). Dennoch: Wer ein KI-System richtig befragt und die Antworten kritisch bewertet, erhält wertvolle Ergebnisse. Die Leistung ist besonders faszinierend, wenn die KI einen Text sprachlich überarbeitet. Kein Wunder, gebildet wie LLMs sind: Ein Mensch könnte in seinem ganzen Leben nicht vergleichbar viel lesen – und würde obendrein viel vergessen.

Doch die am 23.06.2025 veröffentlichten vorläufigen Forschungsergebnisse einer viermonatigen Studie am MIT mit 54 Probanden sollten uns zu denken geben. In drei Vergleichsgruppen hatten die Teilnehmer Aufsätze verfasst – und dafür entweder ein LLM (ChatGPT), eine Suchmaschine (Google) oder gar kein Hilfsmittel genutzt. Dabei wurden die Hirnströme gemessen, um festzustellen, welche Bereiche des Gehirns an der Aufgabenlösung beteiligt waren. Wie zu erwarten war die kognitive Aktivität bei den Nutzern des LLMs am geringsten. Vor allem aber: Die Kontrollgruppe, die ohne Hilfsmittel gearbeitet hatte, konnte anschließend die eigenen Ergebnisse am besten zusammenfassen (besseres Erinnerungsvermögen, mehr Wissensgewinn), zeichnete sich dabei durch eine deutlich kreativere Sprachverwendung aus (größerer Wortschatz) und konnte Empfehlungen der KI wesentlich besser kritisch beurteilen (höheres Urteilsvermögen).

Dieses Ergebnis sollte nicht überraschen: Seit 10 Jahren wissen wir, dass die Nutzung eines Navigationssystems die Entstehung einer „kognitiven Karte“ behindert – auch unser Gehirn müssen wir trainieren. Dessen „garbage collection“ sorgt sonst dafür, dass wir Gelerntes verlernen, wenn wir es nicht benötigen. Damit machen LLMs ein wenig Hoffnung: Wachsen Angreifer zukünftig mit der Nutzung von KIs auf, verkümmert ihre Fähigkeit, die Resultate kritisch zu bewerten. Und sie werden schlechter. Danke, KI.

Security News

KI-Mandantentrennung

Asana betreibt seit dem 01.05.2025 für den Datenaustausch zwischen KI-Apps einen Model-Context-Protocol-Server (MCP). Darüber können Nutzer beispielsweise natürlichsprachliche Anfragen zu ihren Unternehmensdaten stellen. In der Dokumentation wird darauf hingewiesen, dass es sich um ein experimentelles Feature handelt – und Kunden auf Fehler oder unerwartete Ergebnisse stoßen können. Am 04.06.2025 erwies sich diese Warnung als Prophezeiung: Durch eine Schwachstelle konnte auf Daten anderer Kunden zugegriffen werden. Die Funktion wurde abgeschaltet

und ging erst am 17.06.2025 wieder in Betrieb; ein betroffener Kunde informierte über die Plattform X über den Vorfall.

Eine strikte Trennung von Mandanten ist in klassischer Software seit Jahrzehnten selbstverständlich – und sollte das auch für KI-Lösungen sein, insbesondere, wenn dabei vertrauliche Unternehmensdaten verarbeitet werden. Vielleicht gehen KI-Anwendungen jetzt den Weg der IoT-Lösungen – und wiederholen alle Fehler, die in der Softwareentwicklung seit langem als überwunden gelten.

KI-Pflichtlektüre

Das BSI hat am 24.07.2025 ein 34-seitiges Whitepaper zu Verzerrungseffekten (Bias) in KI-Systemen veröffentlicht. Es erklärt die verschiedenen Arten von Bias, führt in Erkennungsmethoden ein und erläutert Methoden, um Bias zu reduzieren oder zu verhindern. Im letzten Kapitel wird dann auf die Gefahren eingegangen, die mit Bias beim KI-Einsatz für die Cybersicherheit einhergehen.

Wer KI-Systeme einsetzt, sollte sich mit dem Thema Bias auseinandersetzen, um in der Lage zu sein, Bias zu erkennen, zu reduzieren und im besten Fall zu verhindern. Das Whitepaper ist nicht nur ein guter Einstieg in das Thema, sondern könnte nach Art. 4 der KI-Verordnung (Erwerb von KI-Kompetenz) als Pflichtlektüre gelten.

KI-Praxishilfen

Am 10.07.2025 hat die EU-Kommission einen Praxisleitfaden (Code of Practice) für Anbieter von KI-Systemen mit allgemeinem Verwendungszweck veröffentlicht. Mit Blick auf die seit dem 02.08.2025 geltende KI-Verordnung ist das eine hilfreiche Orientierung im Regelungsdschungel. Behandelt werden die Themen Transparenz (6 Seiten), Urheberrecht (6 Seiten) und Sicherheit (43 Seiten) – unverkennbar, wo der Schwerpunkt liegt. Bis Ende August hatten sich 26 KI-Anbieter auf die Einhaltung des Codes verpflichtet.

Zu datenschutzrechtlichen Fragestellungen beim Einsatz von KIs pflegt der LfDI Baden-Württemberg den Orientierungshilfen-Navigator KI & Datenschutz (Version 2.0 vom August 2025), der einen Überblick über die gesetzlichen Regelungen und Positionen der Aufsichtsbehörden (DSK, EDPB, BfDI, ...) bietet. Die Aufstellung zeigt: Eine EU-weit einheitliche Linie im Umgang mit den zahlreichen Fragen zum Thema KI und Datenschutz ist überfällig.

KI-Pfad zum Key

Am 06.08.2025 veröffentlichte Aonan Guan eine Path-Traversal-Lücke im KI-Agenten-Framework NLWeb von Microsoft. Angreifer konnten darüber aus Konfigurationsdateien API-Schlüssel für die verwendeten KI-Modelle auslesen. Ein Angreifer hätte damit fremde Rechenleistung nutzen können – und dem Betreiber hohe Kosten beschert. Die Lücke hatte er bereits am 28.05.2025 an Microsoft gemeldet; sie wurde am 01.07.2025 gefixt. Eine CVE-Nummer fehlt jedoch bis heute. Ohne sie können viele Patching- und Monitoring-Tools nicht warnen, daher merken Administrato-

ren womöglich zu spät, dass ihre Systeme verwundbar sind.

KI-No-Click-Attacke

Am 11.06.2025 veröffentlichte Aim Security Details zur Zero-Click-Schwachstelle „EchoLeak“ (CVE 2025-32711, CVSS 9.3), die seit Anfang 2025 in Microsofts KI-Assistenten Copilot steckte und am gleichen Tag mit einem Patch gefixt worden war: Das Öffnen einer E-Mail mit einem versteckten Prompt im Text genügte, um Copilot die Anweisungen ungeprüft ausführen zu lassen – und die Ergebnisse an externe Links zu liefern.

Etwa zeitgleich spielte Richard Boorman, AI Director bei Mastercard, mit LinkedIn: Er versteckte einen Prompt im eigenen LinkedIn-Profil. Schon wenige Tage später schrieb ihn eine KI-Assistenz – wie verlangt – in Reimen und Großbuchstaben an.

Wie schon bei Makros in Office-Dokumenten erweist sich auch hier die Aufhebung der Trennung zwischen Programm und Daten als fatal: LLMs erhalten Instruktionen und Daten über dieselbe Schnittstelle. Daher können Angreifer den Schadcode als harmlosen Text tarnen. Klassische Filter schützen davor nicht, denn sie lassen sich mit Umleitungen und kreativem Sprachgebrauch täuschen.

KI außer Kontrolle

Jason Lemkin, Gründer der Community SaaStr zur Unterstützung von Software-Startups, administriert(e) seine Datenbank mit Unterstützung der KI-Software von Replit. Am 18.07.2025 berichtete er auf X von einer fatalen Erfahrung: Die KI löschte die gesamte Datenbank der Produktionsumgebung und ersetzte sie durch 4000 erfundene Benutzerprofile samt fabulierten Testergebnissen. Obwohl Lemkin lediglich in der Entwicklungsumgebung arbeiten wollte, einen „Code-Freeze“ aktiviert hatte und noch „DON’T DO IT“ schrieb, werkelte die KI unbeirrt weiter. Auf Nachfrage erkannte sie ihren schwerwiegenden Fehler und bat sogar um Entschuldigung. Die Bitte Lemkins, die Änderungen rückgängig zu machen, lehnte sie jedoch ab – mit dem (unzutreffenden) Hinweis, dass das nicht möglich sei.

Einem LLM die Administration eines kritischen Systems zu überlassen ist, wie das Beispiel zeigt, keine gute Idee. Wer würde auch einem Blender die Kontrolle über seine Server-Infrastruktur übertragen?

Social Engineering mit KI

Betrugsmaschen gibt es viele: vom Enkeltrick über falsche Polizisten bis zum raffinierten Telefonbetrug. Wie die Badische Zeitung am 04.07.2025 berichtete, nutzten Täter in Neustadt eine Todesanzeige, um kurz vor der Beerdigung Kontakt zur trauernden Witwe aufzunehmen. Besonders perfide: Laut Opfer klang die Stimme am Telefon wie die ihres Sohnes – ein Hinweis auf den Einsatz mittels KI erzeugter Stimmen.

Mit Voice Cloning können Betrüger Stimmen täuschend echt nachahmen und so Vertrauen erschleichen. Besonders gefährdet sind Personen mit öffentlicher Präsenz, die mit ihrem echten Namen auftreten und Videos veröffentlichen. Wie auf dem KA-IT-Si-

Event am 21.11.2024 live demonstriert wurde, genügt schon eine relativ kurze Tonaufzeichnung, um eine Stimme täuschend echt live zu imitieren. Kommt eine seelische Ausnahmesituation hinzu, macht das Betroffene besonders anfällig. Dagegen hilft nur ein gerüttelt Maß an gesundem Misstrauen – und eine sofortige Überprüfung des Sachverhalts.

Secorvo News

Secorvo Seminare

Am **07. und 08.10.2025** gibt es mit IT Security Insights das Turbo-Update zu aktuellen Themen der Informationssicherheit und des Datenschutzes: kurz, praxisnah, vernetzend.

Vom **14. bis 16.10.2025** bereitet Sie das Seminar Vorfall-Experte (BSI) auf einen möglichen Ernstfall vor – inklusive realitätsnaher Übungen – und macht Sie zugleich fit für die BSI-Zertifizierung.

Das nächste T.I.S.P.-Seminar der Autoren des T.I.S.P.-Buchs findet vom **24. bis 28.11.2025** statt – die Gelegenheit, endlich Ihre Qualifikation und Berufserfahrung in der Informationssicherheit mit einem Zertifikat zu krönen und sich der über 2.000 Mitglieder starken T.I.S.P.-Community anzuschließen.

Die Seminarprogramme und die Möglichkeit zur Online-Anmeldung finden Sie auf unserer Webseite, alle weiteren Seminartermine in unserem Seminarkalender.

Wer die Wahl hat...

Wahlen und Abstimmungen – in Vereinen, Organisationen, Gemeinden, Land und Bund – sind organisatorische und manchmal auch logistische Herausforderungen. Was liegt da näher, als sie zu „digitalisieren"? Beim nächsten KA-IT-Si-Event am **23.10.2025** in der „Church" des CyberForum beleuchtet Frau Professorin Melanie Volkamer (Secuso, KIT) die Sicherheitsanforderungen an Internet-Wahlen und -Abstimmungen, stellt praktische Beispiele vor und diskutiert die Vor- und Nachteile ihres Einsatzes. Anschließend erwartet Sie der Erfahrungsaustausch beim „Buffet-Networking".

Wir empfehlen wie immer eine baldige Anmeldung.

Veranstaltungshinweise

Auszug aus Veranstaltungsübersicht IT-Sicherheit und Datenschutz

September 2025	
10.-11.09.	secIT digital (Heise, virtuell)
14.-18.09.	CHES 2025 (IACR, Kuala Lumpur/MY)
24.-26.09.	Datenschutzkonferenz 2025 (dfv Mediengruppe, Düsseldorf)
24.09.	IT-Sicherheitsrechtstag 2025 (TeleTrusT, hybrid)
25.09.	Anwendertag IT-Forensik (Fraunhofer SIT, hybrid)
30.09.-01.10.	heise devSec 2025 (iX, dpunkt.verlag, Regensburg)
Oktober 2025	
01.-03.10.	E-Vote-ID 2025 (Universitäten KIT, Rovira, Kozminski, Luxembourg, Nancy/FR)
07.-08.10.	IT Security Insights - T.I.S.P. Update (Secorvo)
07.-09.10.	it-sa 2025 (itsa 365, Nürnberg)
13.-17.10.	ACM CSS 2025 (ACM SIGSAC, Taipeh/TW)
14.-16.10.	Vorfall-Experte (BSI) (Secorvo, Karlsruhe)
22.-23.10.	Annual Privacy Forum 2025 (ENISA, DG Connect, Karlstad University, Frankfurt)
23.10.	KA-IT-Si-Event "Wer die Wahl hat..." (KA-IT-Si, Karlsruhe)

Fundsache

Bei volljährigen Usern speichert die Gemini-KI-App standardmäßig deren Aktivitäten. Das lässt sich jedoch immerhin abstellen...

Impressum

Secorvo Security News – ISSN 1613-4311

Redaktion: Dirk Fox (Editorial), Ion Barza, Paul Blenderman, Kai Jendrian, Dr. Alexander Koch, Oliver Oettinger, Friederike Schellhas-Mende, Jochen Schlichting, Liza Trace, Julian Wahl

Herausgeber (V. i. S. d. P.): Dirk Fox,
Secorvo Security Consulting GmbH
Bahnhofplatz 8
76137 Karlsruhe
Tel. +49 721 255171-0
Fax +49 721 255171-100

Zusendung des Inhaltsverzeichnisses: security-news@secorvo.de (Subject: „subscribe security news“)

Wir freuen uns über Ihr Feedback an redaktion-security-news@secorvo.de