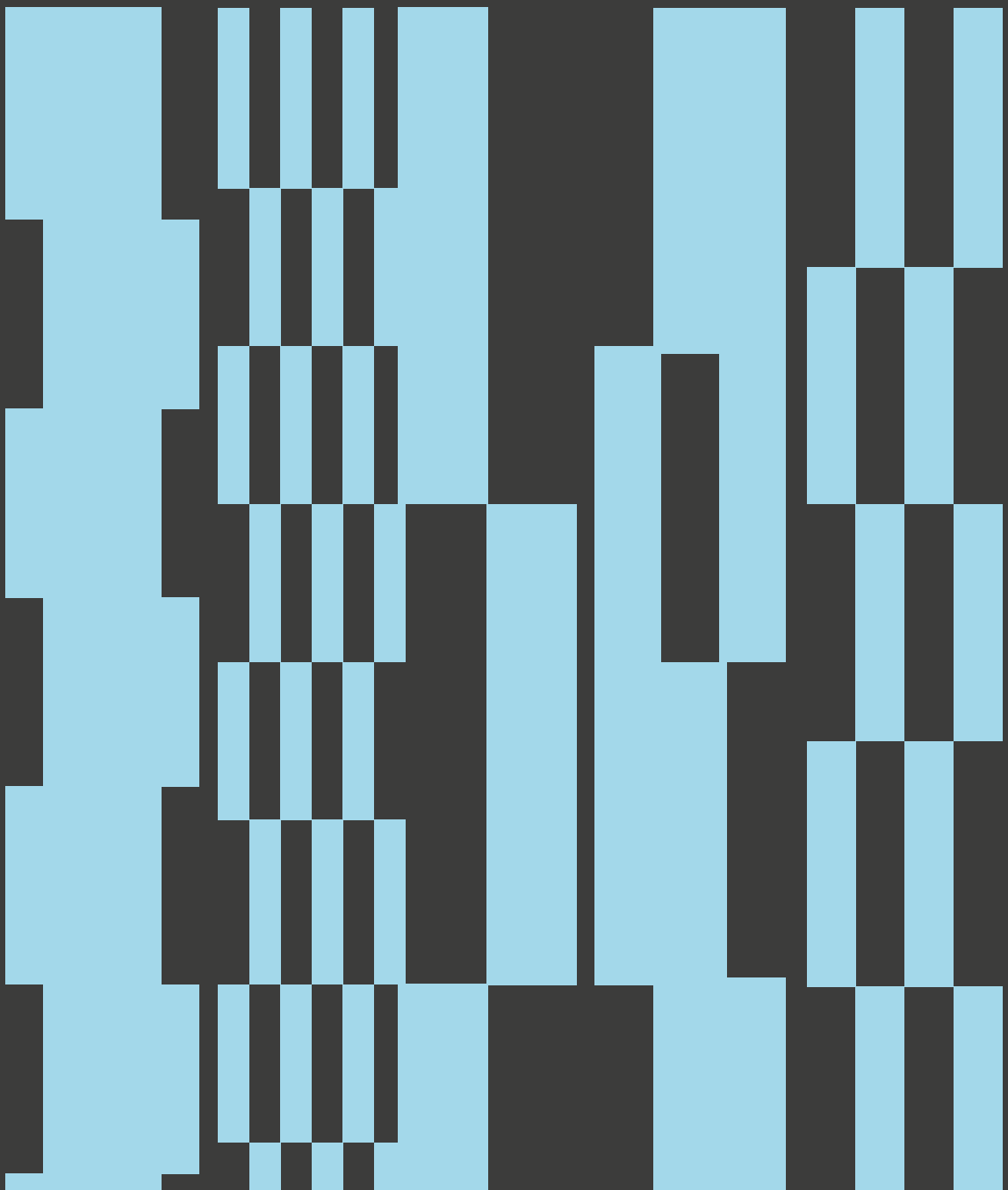
CISPA DISPLAY

DE

EDITION 2026



VORWORT

Die CISPA-Forschung sichtbar und zugänglich zu machen, ist eine unserer zentralen Aufgaben. Dies geschieht etwa über die Verbreitung von Forschungsnachrichten, Veranstaltungen des CISPA Cysec Labs, Messebesuche, Besuche von Stakeholdern und Politiker:innen – wie im Jahr 2025 Bundeskanzler Friedrich Merz – bis hin zu großen Forschungsfestivals wie CISPA loves IGB. Unsere Formate sind vielfältig und reichen von vermittelnden Texten und Gesprächen bis hin zu Workshops. In dieser Vielfalt ist das nun im dritten Jahr erscheinende Forschungsjahrbuch CISPA DISPLAY zu einem festen Bestandteil des Wissenstransfers geworden. Es bringt die im Vorjahr publizierten Texte über ausgewählte wissenschaftliche Paper unserer Forschenden zusammen, die bei renommierten internationalen Konferenzen veröffentlicht wurden. Dies ermöglicht einen kompakten Einblick in die Forschungsthemen am CISPA.

Rahmenbedingungen für exzellente Forschung

Zwei tragende Pfeiler exzellenter Forschung sind eine demokratische Gesellschaftsordnung und eine verlässliche öffentliche Finanzierung. Beides ist in Deutschland gegeben. Mit der Helmholtz-Gemeinschaft verfügt die Forschungslandschaft darüber hinaus über ein starkes Netzwerk, das sich wissenschaftlicher Exzellenz verpflichtet hat. Gleichzeitig zeigt ein Blick in andere Länder, wie fragil diese politisch getragenen Rahmenbedingungen sein können: Wissenschaft und Forschung stehen dort oft vor Herausforderungen wie populistisch geprägten Debattenkulturen oder politischen Eingriffen in die Autonomie und die Förderstrukturen von Hochschulen und Forschungseinrichtungen. Hinzu kommen wachsende technologische Abhängigkeiten und sich verändernde digitale Ökosysteme, die die Diskussion darüber prägen, wie Gesellschaften ihre technologische Zukunft gestalten wollen.

Digitale europäische Souveränität

Angesichts dieser Entwicklungen ist die Bedeutung digitaler Souveränität in Deutschland und Europa zuletzt verstärkt ins Zentrum gesellschaftlicher und politischer Debatten gerückt. Dabei geht es längst nicht mehr nur um wirtschaftliche Wettbewerbsfähigkeit, sondern auch um strategische Handlungsfähigkeit, technologische Unabhängigkeit und die Frage, wie demokratische Gesellschaften Schlüsseltechnologien eigenständig gestalten können. Informations- und Datensicherheit, vertrauenswürdige künstliche Intelligenz, robuste Infrastrukturen

In einer Zeit, in der „Digitalisierung“ allgegenwärtig ist und weder Forschung zu KI und Cybersicherheit noch der Wissens-transfer ohne digitale Werkzeuge denkbar sind, wirkt ein gedrucktes Forschungsjahrbuch fast aus der Zeit gefallen. Das CISPA DISPLAY ist jedoch ein bewusster Medienwechsel.

und klare regulatorische Leitplanken zählen zu den zentralen Bausteinen einer europäischen Antwort auf die globale Technologiedynamik.

Für Forschungszentren wie das CISPA bedeuten diese Entwicklungen eine doppelte Verantwortung: einerseits technologische Innovationen entscheidend mitzugestalten und andererseits einen Beitrag zur Stärkung europäischer Werte wie Demokratie, Freiheit und Sicherheit zu leisten. Die herausragende wissenschaftliche Evaluation des CISPA im Jahr 2025, sein erfolgreicher Aufwuchsprozess sowie seine internationale Sichtbarkeit unterstreichen das große Potenzial des Zentrums. Forschung aus Deutschland und Europa liefert nicht nur exzellente wissenschaftliche Beiträge, sondern schafft Grundlagen für eine digitale Zukunft, die unabhängig, sicher und gesellschaftlich verantwortungsvoll gestaltet ist.

Künstliche Intelligenz im Fokus

Kaum ein anderes Technologiethema hat die öffentliche Debatte im vergangenen Jahr so stark geprägt wie die künstliche Intelligenz. Diese Aufmerksamkeit spiegelt sich auch im vorliegenden CISPA DISPLAY. Während KI-Anwendungen längst in den privaten und beruflichen Alltag eingezogen sind und zunehmend als beratende Instanzen genutzt werden, intensiviert sich auch die Diskussion über ihre Bedeutung für unsere demokratische Gesellschaft. CISPA-Forschende arbeiten täglich daran, KI nicht nur vertrauenswürdig zu gestalten, sondern auch sicherzustellen, dass ihr Potenzial eher zum gesellschaftlichen Segen als zum Fluch wird – insbesondere dort, wo KI und Cybersicherheit untrennbar miteinander verflochten sind.

CISPA DISPLAY: Brückenschlag zwischen analoger und digitaler Welt

In einer Zeit, in der „Digitalisierung“ allgegenwärtig ist und weder Forschung zu KI und Cybersicherheit noch der Wissenstransfer ohne digitale Werkzeuge denkbar sind, wirkt ein gedrucktes Forschungsjahrbuch fast aus der Zeit gefallen. Das CISPA DISPLAY ist jedoch ein bewusster Medienwechsel: Wir verstehen es als einen produktiven Brückenschlag zwischen der analogen und der digitalen Welt. Das Überreichen einer gedruckten Broschüre, das bewusste Blättern, das Verweilen an spannenden Inhalten und Visualisierungen schaffen Momente des konzentrierten Innehaltens. Es sind Momente, die in unserem beschleunigten Alltag immer seltener werden. Umso mehr wünschen wir unseren Leser:innen eine anregende und inspirierende Lektüre der Edition 2026 von CISPA DISPLAY.

INDEX

INDEX

© Chiara Schwarz

Prozessortyp, IP-Adresse, genutzter Browser, installierte Schriftarten – durch das Sammeln solcher und weiterer Merkmale der Browsereinstellungen und des zugrundeliegenden Betriebssystems lässt sich ein sehr genaues und in einigen Fällen sogar einzigartiges Profil von Nutzer:innen erstellen. Bekannt geworden ist dieses Phänomen als Browser Fingerprinting. Eine Untersuchung von CISPA-Forscher Leon Trampert und Kollegen legt jetzt nahe, dass dieses Trackingverfahren nicht nur beim Agieren im Web, sondern auch in Mails anwendbar ist, und zwar auf einem bislang eher unterbeleuchteten Umweg: über den Einsatz von CSS (Cascading Style Sheets), einer Sprache zur Gestaltung von Websites. Das Paper „Cascading Spy Sheets: Exploiting the Complexity of Modern CSS for Email and Browser Fingerprinting" wird auf dem Network and Distributed System Security Symposium (NDSS) 2025 vorgestellt.

Digitaler Fingerabdruck: CSS eröffnet neue Möglichkeiten zum Nutzer:innen-Tracking

Leon Trampert

Auch in einer großen Gruppe von Websitebesucher:innen sind Sie wahrscheinlich eindeutig identifizierbar. Warum? Überall dort, wo die Programmiersprache JavaScript zum Einsatz kommt – und das ist so ziemlich im gesamten Web – können auch spezifische Attribute zu den von Ihnen genutzten Geräten und deren Einstellungen gesammelt werden. Diese Infos sollen eigentlich Webentwickler:innen helfen, bessere Nutzererlebnisse und Funktionalitäten zu schaffen. Aber auch hier gilt: Wissen ist Macht, und nicht jede:r will, dass dieses Wissen über ihn in der Welt ist. „Mittlerweile ist das Fingerprinting über JavaScript ziemlich bekannt. Menschen, denen Privatsphäre besonders wichtig ist, können sich schützen, indem sie JavaScript blockieren. Das geht entweder mithilfe von Plugins oder durch Nutzung des Tor-Browsers. Das kann zum Beispiel für Journalist:innen, die Angst vor Verfolgung haben, hilfreich sein“, erklärt Leon Trampert.

Modernes CSS lässt Daten durchsickern

Wo sich eine Tür schließt, geht eine andere auf, heißt es, und so scheint es auch beim Fingerprinting zu sein. „Forschende haben kürzlich herausgefunden, dass auch durch den Einsatz von CSS Infos über Nutzende durchsickern können“, so Trampert. CSS (kurz für Cascading Style Sheets) sorgt dafür, dass Texte, Bilder und Menüs an der richtigen Stelle stehen: es bestimmt Schriftarten und Farben sowie die Größe von Elementen auf Websites. Zudem hilft es, dass sich deren Ansicht an verschiedene Bildschirmgrößen anpassen kann. „CSS wird immer beliebter und hat in den vergangenen Jahren immer neue Funktionen hinzugewonnen. Einige davon wurden von Forschungskolleg:innen bereits auf ihr Potenzial für Privatsphäre-Verletzungen untersucht. Eine ganzheitliche Betrachtung stand allerdings noch aus.“ Und so hatte sich Trampert vor einigen Monaten entschieden, systematisch moderne CSS-Funktionen zu untersuchen. „Wir wollten sehen, wie viel wir damit herausfinden können und ob CSS das Tracking auch jenseits des Webs ermöglicht.“

Verräterische Schriften

Trampert hat mehrere Fingerprinting-Ansätze untersucht und mithilfe verschiedener Techniken drei Wege aufgezeigt, mit denen basierend auf CSS Fingerprints von Nutzenden erstellt werden können. „Wir haben zunächst 1176 Kombinationen aus Browser und Betriebssystemen mit verschiedenen Einstellungen untersucht und konnten in 97,95 Prozent Rückschlüsse auf das System der Nutzenden ziehen. Verräterisch sind zum Beispiel installierte Schriftarten. Sie können Hinweise auf den genutzten Browser, das Betriebssystem und installierte Programme geben“, erklärt Trampert. Welche Schriftarten genutzt werden, haben die Forschenden mit ein paar Tricks herausgefunden: „Wir sehen das nicht im Klartext, können aber zum Beispiel durch das Ausnutzen bestimmter an sich sinnvoller CSS-Funktionen Höhen und Breiten von Wörtern messen und daraus nicht nur auf die Schriftart, sondern zum Beispiel auch auf die Systemsprache schließen“, sagt Trampert.

CSS erlaubt Tracking auch jenseits des Webs

Noch spannender war für ihn aber das Testen von Mailanwendungen. Denn während JavaScript von vielen Mailclients standardmäßig blockiert wird, ist der Einsatz von CSS bislang nicht begrenzt. „Wir haben 21 Mailclients untersucht, darunter sowohl Android- und iOS- als auch Desktop- und Web-Clients. In neun Fällen konnten wir alle unsere Techniken erfolgreich einsetzen und so Informationen über die Nutzenden sammeln. 18 der 21 Mailclients davon waren für mindestens eine bestimmte Technik anfällig“, erklärt Trampert. Seiner Einschätzung nach könnte das ganz neue Bedrohungsszenarien eröffnen. „Angriffe könnten zum Beispiel darauf abzielen, die Web-Sitzungen von Besucher:innen mit deren E-Mail-Konto zu verknüpfen oder alle E-Mail-Adressen bestimmter Nutzer:innen zu identifizieren“, erklärt Trampert.

Und nun?

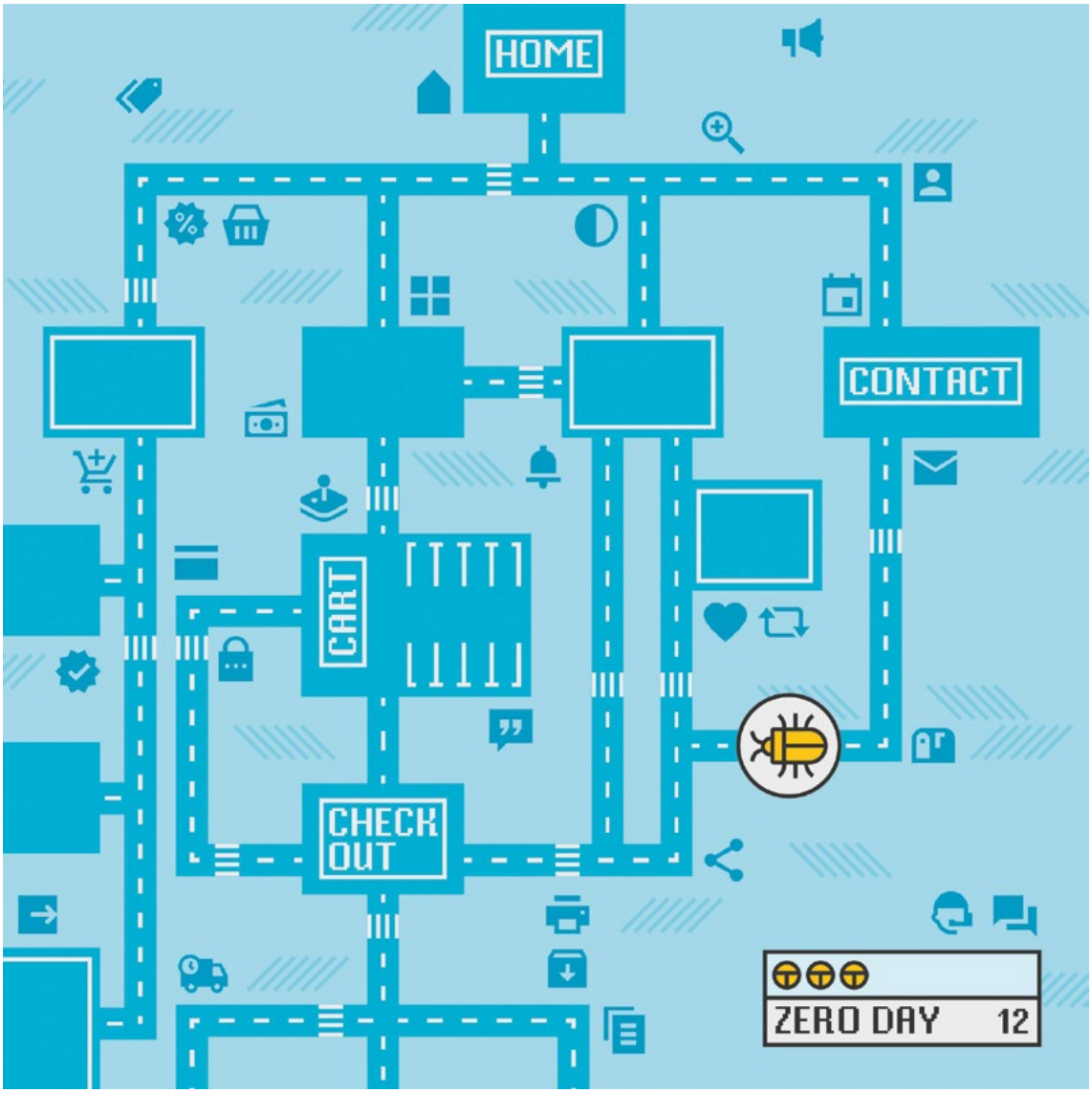
Wer sich im Web bewegt, ist aufgrund des Einsatzes von Tracking-Cockies und JavaScript längst ungewollt vermessen. „Trotzdem ist es wichtig aufzuzeigen, welche technischen Möglichkeiten es gibt und wo sich neue Missbrauchsmöglichkeiten eröffnen – wie hier gesehen plötzlich auch in Mailprogrammen. Nur so können wir auch robuste Verteidigungsmaßnahmen entwickeln“, sagt Trampert. Der PhD-Student forscht am CISPA betreut von den CISPA-Faculty Dr. Michael Schwarz und Prof. Dr. Christian Rossow und will sich auch in Zukunft weiter mit Mailsicherheitsfragen beschäftigen.

»CSS wird immer beliebter und hat in den vergangenen Jahren immer neue Funktionen hinzugewonnen. Einige davon wurden von Forschungskolleg:innen bereits auf ihr Potenzial für Privatsphäre Verletzungen untersucht.«

Trampert, Leon; Weber, Daniel; Gerlach, Lukas; Rossow, Christian; Schwarz, Michael (2025): Cascading Spy Sheets: Exploiting the Complexity of Modern CSS for Email and Browser Fingerprinting. In: NDSS 2025, 24–28 Febr, 2025, San Diego CA, USA, Conference: Network and Distributed System Security Symposium (NDSS)

Forscher: ***Leon Trampert***
Autorin: ***Annabelle Theobald***

Veröffentlichung
03.01.2025



© Chiara Schwarz

Ein neuer automatisierter Scanner für Webanwendungen kann Tasks und Workflows in Webanwendungen selbstständig erkennen und ausführen. YuraScanner nutzt das Weltwissen in Large Language Models (LLMs), um wie menschliche Nutzende durch Webanwendungen zu navigieren. Er kann Tasks auf kohärente Weise bearbeiten und dabei die korrekte Abfolge von Arbeitsschritten ausführen, wie sie etwa ein Online-Shop erfordert. YuraScanner wurde auf 20 Webanwendungen getestet und hat dabei zwölf bislang unbekannte Cross-Site-Scripting-Schwachstellen aufgedeckt. Die Methode hinter YuraScanner sowie das Tool selbst hat der CISPA-Forscher Aleksei Stafeev zusammen mit seinen Kollegen entwickelt. Das Paper „YuraScanner: Leveraging LLMs for Task-driven Web App Scanning“ wird auf dem Network and Distributed System Security Symposium (NDSS) 2025 vorgestellt.

LLM-basierter Scanner für Webanwendungen erkennt Tasks und Workflows

Aleksei Stafeev

Automatisierte Webanwendungsscanner werden eingesetzt, um die Sicherheit von Online-Anwendungen wie zum Beispiel Online-Shops, Lernplattformen oder Projekt-management-Tools zu testen. In der Regel bestehen diese Scanner aus zwei Teilen: der Crawler-Komponente, die Webanwendungen auf der Suche nach Nutzerschnittstellen durchforstet, und dem Angriffsmodul, das die vom Crawler identifizierten Schnittstellen testet. Aleksei Stafeev aus der Forschungsgruppe von CISPA-Faculty Dr. Giancarlo Pellegrino unterstreicht die Bedeutung der Crawler-Komponente für den Erfolg automatisierter Testung: „Eine der größten Herausforderungen bei der Sicherheitstestung besteht darin, den Umfang von Webanwendungen zu bestimmen und ihre Funktionalitäten und Workflows zu identifizieren. Wir wissen recht gut, wie wir Sicherheitsprobleme erkennen können, aber wie finden wir alle Eingangspunkte?“ Stafeev und seine CISPA-Kollegen haben YuraScanner mit dem Ziel entwickelt, so viel wie möglich von der Angriffsfläche identifizieren zu können.

Mithilfe von LLMs navigiert YuraScanner durch Webanwendungen

Die wichtigste Neuerung des YuraScanners ist die Anbindung der Crawler-Komponente an ein LLM, um die Reichweite und Leistung des Crawlers zu erhöhen. „LLMs sind mit den Daten aus dem Internet trainiert worden, die umfangreiche Dokumentationen über den Umgang mit Websites beinhalten. Wir nutzen dieses Wissen, indem wir einen Crawler mit einem LLM kombinieren, um die Erkundung von Webanwendungen anzuleiten”, erklärt Stafeev. Für ihre Studie haben die CISPA-Forscher die OpenAI-API genutzt, um die Verbindung zwischen ihrer Crawler-Komponente und dem OpenAI-Modell GPT-4 herzustellen. Das Angriffsmodul des YuraScanners ist identisch mit Black Widow, einem etablierten Cross-Site-Scripting-Scanner auf dem neuesten Stand der Technik. Dieses parallele Setup hat es den Forschern ermöglicht, die Leistung der Crawler-Komponenten von YuraScanner und Black Widow miteinander zu vergleichen. Sie haben YuraScanner auf 20 Webanwendungen getestet und dabei zwölf bisher unbekannte XSS-Schwachstellen entdeckt, während Black Widow nur drei aufgespürt hat.

Automatisiertes Scannen auf tieferliegenden Ebenen der Webanwendung

Unter der Anleitung eines LLM arbeitet YuraScanner task-orientiert, wodurch er in die tieferen Ebenen der zu testenden Webanwendung vordringt. Er kann die in der Webanwendung vorgesehenen Tasks nicht nur erkennen, sondern sie auch gezielt ausführen. Dabei beachtet er die Abfolge von Schritten, die zum Erledigen des jeweiligen Tasks erforderlich ist. Stafeev erläutert: „Normalerweise unterscheiden Testing-Tools nicht zwischen verschiedenen Arten von Buttons, sondern klicken einfach auf alles, was verfügbar ist. Der größte Nachteil dabei ist, dass bei einem sehr spezifischen mehrschrittigen Workflow, wie zum Beispiel in einem Online-Shop, bei dem man einen Artikel in den Warenkorb legen, zur Kasse gehen und ein Formular ausfüllen muss, die Wahrscheinlichkeit sehr gering ist, dass ein einfacher Webcrawler das erfolgreich erledigen kann.“ Mit YuraScanner haben Stafeev und seine Kollegen gezeigt, dass LLMs für Web-Sicherheitsscans eingesetzt werden können und damit den Weg für weitere Forschung auf diesem Gebiet geebnet.

Förderhinweise auf Seite 78

»LLMs sind mit den Daten aus dem Internet trainiert worden, die umfangreiche Dokumentationen über den Umgang mit Websites beinhalten. Wir nutzen dieses Wissen, indem wir einen Crawler mit einem LLM kombinieren, um die Erkundung von Webanwendungen anzuleiten.«

Stafeev, Aleksei; Recktenwald, Tim; De Stefano, Gianluca; Khodayari, Soheil; Pellegrino, Giancarlo (2024): YuraScanner: Leveraging LLMs for Task-driven Web App Scanning. In: NDSS 2025, 24–28 Febr, 2025, San Diego CA, USA, Conference: Network and Distributed System Security Symposium (NDSS)

© Chiara Schwarz

Millionen Websites weltweit basieren auf Content-Management-Systemen (CMS) wie Wordpress, die es auch Personen ohne Programmierkenntnisse ermöglichen, eigene Seiten zu erstellen und digitale Inhalte zu verwalten. Gerade ihre weite Verbreitung macht sie zu einem attraktiven Ziel für Cyberangriffe. Regelmäßige Sicherheitsupdates sind das beste Mittel diese abzuwehren, werden aber in mehr als der Hälfte der Systeme nicht durchgeführt. Die Gründe dafür hat CISPA-Forscherin und Psychologin Dr. Maria Hellenthal zusammen mit Kolleg:innen in einer qualitativen Studie untersucht. Ihr Paper „The (Un)usual Suspects – Studying Reasons for Lacking Updates in WordPress" stellt sie auf dem Network and Distributed System Security Symposium (NDSS) 2025 vor.

Das unterschätzte Risiko: warum viele WordPress-Websites zu selten aktualisiert werden

Maria Hellenthal

Cyberkriminelle kapern unsichere oder veraltete Websysteme, stehlen Daten, missbrauchen WordPress-Server für Spam und DDoS und betreiben darauf sogar Fake-Shops. Um Schwachstellen zu schließen und so die Risiken zu minimieren, stellen die Anbieter:innen von Content-Management-Systemen (CMS) ihren Kund:innen regelmäßig neue Sicherheitsupdates bereit. „Leider werden diese Updates von vielen Websitebetreiber:innen nicht oder nicht regelmäßig gemacht“, erklärt Maria Hellenthal. Warum dieses vermeidbare Risiko eingegangen wird, hat das Forscherteam anhand veralteter WordPress-Seiten und Interviews mit deren Betreiber:innen untersucht. Zudem hat es mit Webentwickler:innen und Hosting-Providern gesprochen, um deren professionelle Perspektiven einzubeziehen. „Wir haben Wordpress gewählt, weil es mit mehr als 60 Prozent Marktanteil weltweit derzeit das verbreitetste CMS ist“, sagt Hellenthal.

Fehlende Updates: Ursachen und Hindernisse

Dass Sicherheitsupdates nicht regelmäßig gemacht werden, ist nicht nur im Fall von Wordpress ein Problem: „Wir sehen dieses Phänomen im gesamten Online-Ökosystem“, so Hellenthal. Ein häufig genannter Grund, der auch in Hellenthals Studie auftaucht, ist mangelndes Risikobewusstsein: „Viele Website-Betreiber:innen erkennen nicht, dass Cyberangriffe nicht nur ihnen schaden, sondern der gesamten Netz-Gemeinschaft. Wenn eine Seite gehackt wird, können nicht nur deren Besucher:innen geschädigt werden, sondern auch andere Website-Betreiber:innen und Hosting-Provider – sprich die gesamte Online-Community“, so die Forscherin. Weitere Hindernisse für regelmäßige Updates seien die Angst, dass diese zu Problemen führen könnten, oder dass Zusatzkosten durch Updates anfallen, etwa weil danach nicht mehr alle Plug-Ins funktionieren.

Zwei der in der Studie von Hellenthal identifizierten Gründe wurden in der vorherigen Fachliteratur nicht explizit fürs Nicht-Updaten erwähnt: „Für das Updateverhalten scheint auch ein wichtiger Faktor zu sein, was die Website den Betreiber:innen bedeutet. Wer einen Online-Shop betreibt und für den die Website die Haupteinkommens-

quelle ist, misst ihr einen anderen Stellenwert bei, als sagen wir mal, ein Kleinunternehmer, der über die Seite nur Infos weitergibt und sein Geschäft eher auf Mund-zu-Mund-Propaganda stützt. In beiden Fällen können Updates vernachlässigt werden, allerdings lassen sich Betreiber:innen, denen ihre Seite noch etwas bedeutet, vermutlich eher durch gezielte und verständlich formulierte Hinweise auf potenzielle Schwachstellen zum Aktualisieren der Seiten bewegen“, so die Forscherin.

Ein weiteres Problem kann sich laut der Studie ergeben, wenn die Betreuung von Websites an Externe ausgelagert wird. „Die Übergabe der Websitebetreuung an eine erfahrenere Person sollte Vorteile bringen – kann aber auch Nachteile haben. Wir haben zum Beispiel in mehreren Fällen gesehen, dass es zu einer Verantwortungsdiffusion kommen kann, wenn die Betreuungsaufgaben nicht ganz klar geregelt, vergütet und vertraglich beschrieben sind. Niemand fühlt sich wirklich zuständig.“ Zudem sagte einer der Befragten, dass er oftmals überfordert war, wenn ein Externer ihm Plug-Ins eingebaut hat, mit denen er nicht umgehen konnte, und die Systeme so plötzlich zu komplex wurden. „Und natürlich spielt Geld eine Rolle. Viele können sich keine Agentur leisten, die sich um die Aufgaben kümmert. Technik-affine Freund:innen werden in manchen Fällen nur gefragt, wenn es sich nicht vermeiden lässt“, erklärt Hellenthal.

Sicherheitswarnungen werden häufig ignoriert

Seit Langem versuchen Sicherheitsexpert:innen, Betreiber:innen mit Schwachstellen-Benachrichtigungen zu Updates zu bewegen – oft mit mäßigem Erfolg. „Es gibt Studien, die sich damit beschäftigen, warum die Benachrichtigungen so oft ignoriert werden und wie sie gestaltet werden müssen, um eine größere Wirkung zu erzielen. Unsere Studie zur Frage, warum die Sicherheit der Systeme überhaupt erst vernachlässigt wird, kann uns helfen, besser zu verstehen, wen wir mit den Benachrichtigungen überhaupt noch erreichen können“, sagt Hellenthal. Um das Sicherheitsniveau von Wordpress-basierten Websites nachhaltig zu verbessern, reicht es laut der Forscherin aber nicht aus, nur auf Sicherheitswarnungen zu setzen.

Wie lassen sich Risiken mindern?

Die Forscherin sieht auch Verantwortung bei den CMS-Anbieter:innen: „Sie könnten zum Beispiel Sicherheitslösungen wie Static-Site-Generatoren, in denen keine unnötigen sicherheitsrelevanten Komponenten enthalten sind, deutlich benutzerfreundlicher gestalten. Zudem sollten sie ihre Kund:innen besser und in einer für Laien verständlicheren Art über die Risiken aufklären, die sie eingehen, wenn sie automatische Sicherheitsupdates deaktivieren“, sagt Hellenthal. Ebenfalls hilfreich könnten laut der Forscherin auch öffentliche Anerkennungsprogramme für sichere Websites sein.

Qualitative Forschung liefert neue Einblicke

Für die Studie wurden 19 Personen interviewt. Wie aussagekräftig ist das? „Bei qualitativer Forschung geht es nicht um Generalisierbarkeit, sondern darum, Handlungsmuster zu identifizieren", erklärt Hellenthal. „Auf dieser Basis können wir Theorien entwickeln und sie in weiteren quantitativen Studien testen oder – wie in diesem Fall – unter Berücksichtigung der Gründe für ausbleibende Updates schon erste Verbesserungsstrategien erarbeiten.“ Das interdisziplinäre Projekt ist aus einer gemeinsamen Forschungsidee von IT-Sicherheitsforscher und CISPA-Faculty Dr. Ben Stock und dem Leiter der Abteilung Empirical Research Support, dem Psychologen Dr. Michael Schilling, entstanden. Die Forschungsarbeit stützt sich auf die Masterarbeiten von Lena Gotsche und Sarah Kugel, beide Psychologinnen. Soziologe Dr. Rafael Mrowczynski brachte seine Expertise in der qualitativen Forschungsmethodik ein. „Wir haben uns methodisch und technisch perfekt ergänzt", resümiert Hellenthal, die am CISPA im Team Empirical Research Support arbeitet und IT-Sicherheitsforschende in Methodikfragen und beim Design von Studien unterstützt. „Ich komme aus der experimentellen kognitiven Psychologie und mache schon immer eher angewandte Forschung. Damit war ich lange eher eine Außenseiterin an meiner ehemaligen Hochschule. Am CISPA kann ich meine Fähigkeiten in einem hochinteressanten Bereich einbringen.“

Hellenthal, Maria; Gotsche, Lena; Mrowczynski, Rafael; Kugel, Sarah; Schilling, Michael; Stock, Ben (2025): The (Un)usual Suspects – Studying Reasons for Lacking Updates in WordPress. In: NDSS 2025, 24–28 Febr, 2025, San Diego CA, USA, Conference: Network and Distributed System Security Symposium (NDSS)

Forscherin: Maria Hellenthal
Autorin: Annabelle Theobald

Veröffentlichung 28.02.2025

© Chiara Schwarz

Die Videospielbranche ist ein sich ständig verändernder, milliardenschwerer Markt. Welche Herausforderungen die Einbeziehung von Sicherheitsüberlegungen in die Spielentwicklung mit sich bringt, untersuchte CISPA-Forscher Philip Klostermeyer aus dem Team von CISPA-Faculty Prof. Dr. Sascha Fahl in einer qualitativen Interviewstudie mit Expert:innen aus der Branche. Die Ergebnisse publizierte er im Paper „Skipping the Security Side Quests: A Qualitative Study on Security Practices and Challenges in Game Development", das auf der Conference on Computer and Communications Security (CCS) 2024 vorgestellt wurde.

Sicherheit läuft nur nebenher mit: Erkenntnisse aus der Videospielbranche

Philip Klostermeyer

Videospiele faszinieren Philip Klostermeyer, CISPA-Forscher und Doktorand am CISPA-Standort in Hannover, schon lange. Und dies nicht nur aus der Perspektive des Spielers. „Im Bachelorstudium musste ich ein Spiel schreiben. Da habe ich zum ersten Mal gemerkt, wie viele verschiedene Elemente selbst ein simples Spiel hat“, erzählt er im Gespräch. „Ich verstand plötzlich, wie die verschiedenen Software-Arten da zusammenspielen.“ Denn im Prinzip ist ein Videospiel nichts anderes als eine sehr komplexe Software, erklärt Klostermeyer: „Im Hintergrund haben wir Quellcode und Daten. Auf der Benutzeroberfläche kommen dann eine komplexe grafische Gestaltung und Elemente wie Audio hinzu. Ergänzt wird dies durch die jeweilige Spiellogik. Bei Onlinespielen kommt dann noch die Anbindung an Server hinzu, die die Spielelogik steuern, klassische Sicherheitsthemen wie Login und Authentifizierung managen, aber auch Werbeeinspielungen ermöglichen. Das zeigt, dass fast alle Themen, die in der Computersicherheit wichtig sind, für Videospiele Relevanz haben.“

Studienziel: Überblick verschaffen

Die Komplexität des Prozesses der Videospielentwicklung und die Bedeutung des Sicherheitsthemas machten diesen für Klostermeyer und seine Kolleg:innen zu einem interessanten Forschungsobjekt. „Wir haben uns entschieden, das Thema Sicherheit in der Videospielentwicklung mit Hilfe einer qualitativen Interviewstudie zu untersuchen“, erklärt der CISPA-Forscher. „Die Methode eignet sich gut, um sich einen Überblick über ein Themenfeld zu verschaffen. Denn was es bisher schon relativ viel gibt, sind Paper, die einzelne Themen aus der Spieleindustrie gut beschreiben. Gefehlt hat bisher eine zusammenhängende Übersicht über das ganze Feld.“ Aber noch ein weiterer Aspekt war Klostermeyer wichtig: „Unser Ziel war, Erkenntnisse in die Industrie zu übertragen. Deswegen war uns wichtig, dass wir die Probleme dieser Zielgruppe in den Fokus unserer Studie stellen.“

Konkret befragt wurden für die Studie 20 Personen aus 15 Ländern, die in unterschiedlichen Positionen in der Spieleindustrie tätig sind. „Wir haben geschaut, wer die Stakeholder sind, die bei einer Spieleproduktion dabei sind

und dann systematisch unsere Gesprächspartner:innen gesucht“, erklärt Klostermeyer. „Dazu gehörten etwa Spieleentwickler:innen, Manager:innen, Publisher von Spieleplattformen aber auch Sicherheitsexpert:innen. Damit wollten wir verschiedene Perspektiven auf das Thema Sicherheit bekommen. Ziel war, über die Interviews Erfahrungen aus erster Hand über das Bewusstsein, die Prioritäten, das Wissen und die Praktiken in Bezug auf die Sicherheit in der Branche zu bekommen.“

Sicherheit als nachgelagerter Faktor, der von vielen Aspekten abhängt

Über die Analyse der Interviews destillierten die CISPA-Forschenden zwei zentrale Bereiche heraus, die für das Thema Sicherheit in der Videospielentwicklung von zentraler Bedeutung sind. Das sind zum einen die besonderen Umstände in der Spieleindustrie, die die Spielentwicklung und damit die Sicherheit beeinflussen. „Hier sind Faktoren wie die Schnelllebigkeit der Industrie, unterschiedliche Sicherheitsstandards, Zeit- und Budgetrestriktionen sowie fehlende Beratung zu Sicherheitsthemen zu nennen“, erläutert Klostermeyer. Zum anderen konnten die Forscher:innen fünf sicherheitsrelevante Bereiche im Prozess der Spieleentwicklung identifizieren. „Konkret sind dies Maßnahmen zur Verhinderung von Betrug im Spiel, die Sicherheit von sogenannten Assets wie Sourcecode oder Grafiken, die Netzwerksicherheit, die Softwarestabilität sowie der Schutz von Benutzer:innendaten“, fährt er fort. Die Bedeutung der einzelnen Bereiche hängt dabei immer vom jeweiligen Spieltyp ab. „Dies bedeutet etwa, dass das Thema Netzwerksicherheit für Spiele, die nicht online gespielt werden, nur wenig Relevanz hat“, so der CISPA-Forscher.

Bezogen auf die Frage, ob und wie die Studios das Thema Sicherheit in den Prozess der Videospielentwicklung integrieren, konnte die Studie Zeit, Budget und Teamgröße als die wichtigsten Faktoren identifizieren. Externe Akteure wie Publisher liefern zwar sicherheitsrelevanten Input, priorisieren die Sicherheit aber hauptsächlich, um die Einnahmen ihrer Unternehmen oder deren öffentliches Image zu schützen. Während große Unternehmen eigene Sicherheitsspezialist:innen rekrutieren, fehlt kleinen Studios dafür meist das Budget. Und selbst wenn es bei Entwickler:innen ein Bewusstsein über Sicherheitsprobleme gibt, kann es sein, dass dies von Managementseite als weniger prioritär eingestuft wird als etwa die Spielbarkeit eines Produkts. „Ganz grundsätzlich lässt sich sagen, dass die Spieleindustrie beim Thema Sicherheit sehr sprunghaft agiert“, so Klostermeyer. „So verhindert die Schnelllebigkeit der Industrie, dass Entwickler:innen tiefgehende Sicherheitsmaßnahmen ergreifen und Bedrohungsmodelle für Videospiele entwickeln, die von Beginn der Spielentwicklung an implementiert werden.“

Für Klostermeyer und seine Kolleg:innen aus der Usable-Security-Forschungsgruppe in Hannover war die aktuelle Interviewstudie nur der Aufschlag, um tiefer in die Materie einzusteigen. „Das Schöne an der Studie ist, dass wir diese fünf sicherheitsrelevanten Bereiche im Prozess der Spieleentwicklung identifizieren konnten. Mit dem Wissen können wir anfangen, Vorschläge für Guidelines zu entwickeln.“ Aber bereits jetzt gibt es einige konkrete Empfehlungen für die Branche, die Klostermeyer aus den Ergebnissen der Studie ableitet. Kernpunkt ist, den Aspekt der Sicherheit möglichst früh in die Spielentwicklung zu integrieren und auf allen Ebenen mitzudenken. Hilfreich sind dabei Richtlinien, die jedes Entwicklerstudio ausgehend von den jeweiligen Anforderungen angepasst auf die eigenen Produkte selbst entwickeln sollte. „Das ist eine wichtige Querschnittsaufgabe, die jedes Studio ernst nehmen sollte“, zeigt sich Klostermeyer überzeugt.

»Ganz grundsätzlich lässt sich sagen, dass die Spieleindustrie beim Thema Sicherheit sehr sprunghaft agiert.«

Klostermeyer, Philip; Klivan, Sabrina; Hölter-vennhoff, Sandra; Krause, Alexander; Busch, Niklas; Fahl, Sascha (2024): Skipping the Security Side Quests: A Qualitative Study on Security Practices and Challenges in Game Development. In: CCS 2024, 14–18 Oct, 2024, Salt Like City, USA, Conference: ACM Conference on Computer and Communications Security (CCS)

© Chiara Schwarz

Ein Klick – und schon hat die App weitreichende Zugriffsrechte, etwa auf Kamera, Mikrofon oder Kontakte. Was viele nicht wissen: Ob wir auf „Erlauben“ oder „Ablehnen“ tippen, hängt oft maßgeblich von der Formulierung dieser Anfrage ab. Das zeigt eine aktuelle Studie der CISPA-Forscherin Yusra Elbitar. Ihr Paper „The Power of Words: A Comprehensive Analysis of Rationales and Their Effects on Users’ Permission Decisions” stellte sie auf dem Network and Distributed System Security Symposium (NDSS) 2025 vor. Wenn eine App auf eine sensible Funktion wie Kamera oder Standort zugreifen möchte, zeigt das Betriebssystem eine sogenannte Berechtigungsanfrage, im Englischen „Permission Request“ genannt. Entwickler:innen können diese durch einen zusätzlichen, erklärenden Text ergänzen – eine sogenannte Rationale. Diese Rationale soll Nutzer:innen begründen, warum die App eine bestimmte Erlaubnis benötigt.

Die Macht der Worte: wie Formulierungen das Zustimmungsverhalten bei App-Berechtigungs-anfragen beeinflussen

Yusra Elbitar

„Die App benötigt Speicherzugriff“ oder „Bitte erlaube den Zugriff auf deinen Speicher, um Fotos an deine Beiträge anzuhängen“ – welche dieser Formulierungen würde Sie eher dazu motivieren, der App die Erlaubnis zu erteilen? App-Entwickler:innen können bei der zweiten, konkreteren Variante mit einer höheren Zustimmungsrate rechnen. Zumindest legt dies Elbitars Studie nahe: „Wer versteht, warum Zugriff auf Kamera oder Speicher benötigt wird, fühlt sich besser informiert und hat ein stärkeres Gefühl von Kontrolle. Beides erhöht laut unserer Untersuchung die Zustimmung zu App-Berechtigungen“, erklärt die Forscherin. Gemeinsam mit Kolleg:innen analysierte sie mehr als 9.500 häufig genutzte Android-Apps, um herauszufinden, wie Berechtigungsanfragen formuliert und gestaltet werden.

Zentrale Bausteine für Erklärtexte in App-Berechtigungsanfragen

In der Praxis sahen die Erklärtexte in den untersuchten Apps höchst unterschiedlich aus. „Es gibt zwar – etwa von Apple oder Android – Vorgaben für Entwickler:innen, wie solche Anfragen gestaltet werden sollten. Bindend sind diese aber nicht“, so Elbitar. Und so konnten die Forschenden ganz unterschiedliche Bausteine herausarbeiten, die gemeinsam bestimmen, wie überzeugend und verständlich eine solche Erklärung auf Nutzende wirkt.

„Ein zentrales Element ist die Funktionalitätserklärung: Gute Anfragen benennen klar, für welche Funktion die Berechtigung benötigt wird – zum Beispiel, ‚um Fotos zu deiner Nachricht hinzufügen zu können‘. Fehlt diese Erklärung, bleibt die Begründung vage und verweist lediglich darauf, dass die App ‚sonst nicht richtig funktioniert‘“, so Elbitar.

Auch die Formulierung der Konsequenz spielt eine wichtige Rolle: Manche Anfragen heben positiv hervor, was Nutzer:innen durch die Zustimmung gewinnen. Andere machen deutlich, welche Funktion nicht verfügbar ist, wenn man die Berechtigung verweigert. Letzteres empfinden viele Nutzer:innen als hilfreicher und nachvollziehbarer.

Zudem variiert die Perspektive, aus der die App spricht. Manche wenden sich direkt an die Nutzenden – fordernd („Du musst erlauben…“) oder höflich („Bitte erlaube uns…“). Andere formulieren neutral („Zugriff auf die Kamera ist erforderlich“) oder aus der Sicht der App selbst („Diese App benötigt…“). Ergänzend enthalten manche Anfragen zusätzliche Informationen, die Vertrauen schaffen oder Kontrolle signalisieren: etwa Sicherheitsversprechen wie „Wir speichern keine persönlichen Daten“, Hinweise wie „Du kannst das jederzeit in den Geräteeinstellungen ändern“ oder Links zur Datenschutzerklärung. Je nachdem, wie diese Elemente kombiniert werden, wirkt eine Anfrage eher vertrauenswürdig – oder erzeugt Skepsis.

»Gute Anfragen benennen klar, für welche Funktion die Berechtigung benötigt wird – zum Beispiel, ‚um Fotos zu deiner Nachricht hinzufügen zu können‘.«

Rausfiltern, prüfen, sortieren – die aufwändige Analyse der App-Texte

Die Studie besteht aus zwei Teilen. „Zum einen haben wir über 9.500 beliebte Android-Apps analysiert, um die Formulierungen der Rationals zu erfassen. Zum anderen haben wir in einer Online-Umfrage 960 Personen befragt, wie sie auf unterschiedliche Formulierungen reagieren würden“, so Elbitar.

Vor allem der erste Teil erforderte viel Detailarbeit: „Wir haben ein Machine-Learning-Modell genutzt, um aus Tausenden App-Texten potenzielle Anfragen herauszufiltern – und konnten so über 35.000 solcher Texte identifizieren. Allerdings erkennt das Modell diese nicht immer eindeutig als Berechtigungsanfragen. Es extrahiert Sätze, die potenziell passen – oft aber kontextlos.“ In einigen Fällen stellte sich heraus, dass es sich nur um allgemein gehaltene Sätze handelte, die in anderen Bereichen der App vorkamen. Deshalb war viel manuelle Nacharbeit nötig: Am Ende wertete das Forschungsteam 1054 eindeutige Anfragen aus 709 Apps per Screenshot manuell aus.

Vom Experiment zum Alltag

Die Ergebnisse liefern erste Hinweise für Best-Practice-Empfehlungen an App-Entwickler:innen und UX-Designer:innen. „Wenn wir allerdings belastbare Vorhersagen darüber treffen wollen, wie Menschen auf bestimmte Formulierungen reagieren, brauchen wir zusätzliche Studien unter realen Nutzungsbedingungen“, so Elbitar. Denn in der Untersuchung stellten sich die Teilnehmenden lediglich vor, gerade eine App zu ver-wenden. In echten Nutzungssituationen – mit Zeitdruck oder situativen Einflüssen – könnten Entscheidungen abweichen.

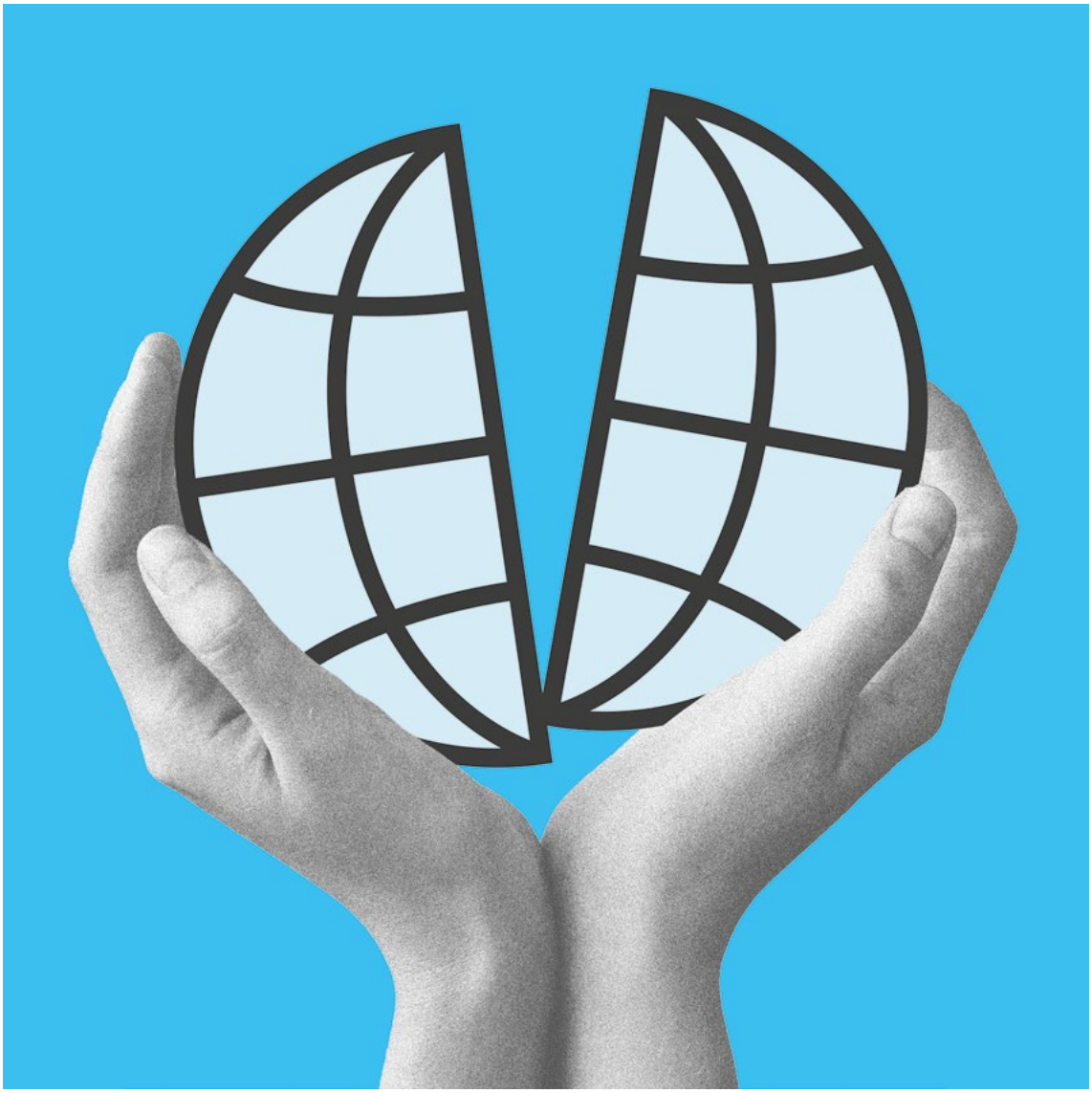
Elbitars Interesse am Thema begann schon in ihrer Masterarbeit. Damals untersuchte sie, ob auch das Timing der Berechtigungsanfrage eine Rolle spielt. „Die Stichprobe war damals noch klein – nur 46 Personen unter Laborbedingungen. Diese neue Studie sollte das erweitern.“ In einer weiteren Arbeit beschäftigte sie sich mit Berechtigungsanfragen auf Webseiten – ein bislang kaum untersuchtes Feld. „Webseiten sind heute oft sehr interaktiv. Hier stellt sich zum Beispiel die Frage: Wird die Anfrage als Banner, Button oder Overlay präsentiert? Auch das kann die Entscheidung beeinflussen.“

Zwar liefert ihre Forschung vor allem Entwickler:innen konkrete Hinweise. Doch für Elbitar steht etwas anderes im Mittelpunkt: „Wir wollen, dass App-Nutzende eine informierte Entscheidung treffen können – wann und wem sie Zugriff auf ihre Daten gewähren.“

Elbitar, Yusra; Hart, Alexander; Bugiel, Sven (2025): The Power of Words: A Comprehensive Analysis of Rationales and Their Effects on Users' Permission Decisions. In: NDSS 2025, 24–28 Febr, 2025, San Diego CA, USA, Conference: Network and Distributed System Security Symposium (NDSS)

Forscherin: ***Yusra Elbitar***
Autorin: ***Annabelle Theobald***

Veröffentlichung
14.04.2025



© Chiara Schwarz

Das Internet ist zwar weltweit verbreitet, doch der oft beschworene globale Charakter wird durch den ‚Digital Divide' relativiert. Denn digitale Teilhabe hängt noch immer stark von ökonomischen Voraussetzungen ab. CISPA-Forscher Masudul Bhuiyan aus dem Team von CISPA-Faculty Dr. Cristian-Alexander Staicu wollte wissen, ob sich auch Unterschiede bei Sicherheit und Datenschutz von Websites finden lassen. Seine Untersuchung von 200.000 Websites aus 20 Industrie-, Entwicklungs- und Schwellenländern zeigt: Websites aus Entwicklungs- und Schwellenländern sind tendenziell kleiner und einfacher, haben häufiger Effizienzprobleme, sind umgekehrt aber vermutlich weniger anfällig für Sicherheitslücken. Die vollständigen Ergebnisse werden im Paper „Digital Disparities: A Comparative Web Measurement Study Across Economic Boundaries" auf der ACM Web Conference 2025 vorgestellt.

Ungleiches Internet: Unterschiede zwischen Websites aus Industrie- und Schwellenländern

Masudul Bhuiyan

Unterschiede in der Digitalisierung zwischen Industrieländern sowie Entwicklungs- und Schwellenländern betreffen eine Vielzahl von Faktoren. „Bisherige Studien haben vor allem Makro-Level-Indikatoren untersucht, wie das Vorhandensein von Internet, die Smartphone-Nutzung oder die generelle technologische Infrastruktur“, erklärt CISPA-Forscher Masudul Bhuiyan. Aus diesen Studien geht hervor, dass in Entwicklungsländern nur 60 Prozent der Bevölkerung online sind, während es in den Industrienationen 93 Prozent sind. Umgekehrt greifen die Menschen in Schwellen- und Entwicklungsländern stärker auf das mobile Internet zu. Darüber hinaus ist die technische Entwicklung dort oft rasant und überspringt ganze Entwicklungsstufen, was auch als Leapfrogging-Phänomen bekannt ist. „Anekdotische Erzählungen in der Community legen darüber hinaus nahe, dass sich Websites in Industrienationen und Entwicklungs- und Schwellenländern signifikant unterscheiden. Wir wollten wissen, ob an dieser Hypothese etwas dran ist“, erzählt Bhuiyan.

Weltweit einzigartiger Datensatz als Datenbasis

Für die Studie haben Bhuiyan und seine Kollegen je 10.000 Websites der zehn bevölkerungsreichsten Industrienationen sowie der zehn bevölkerungsreichsten Entwicklungs- und Schwellenländer untersucht. „Wir haben uns dabei auf die Definition des Internationalen Währungsfonds (IWF) gestützt“, so der CISPA-Forscher. Ausgehend von der Klassifizierung des IWF wurden die Entwicklungs- und Schwellenländer China, Indien, Pakistan, Brasilien, Nigeria, Bangladesch, Russland, Mexiko und Philippinen sowie die Industrienationen USA, Japan, Deutschland, Frankreich, Großbritannien, Italien, Süd-korea, Spanien, Kanada und Australien für die Studie ausgewählt. Pro Land suchten die Forscher dann die 10.000 populärsten Websites heraus. Diese Bedingung hatte zur Folge, dass die Philippinen anstatt Äthiopien ins Sample kamen, da sich dort nicht genug Websites ausreichender Größe fanden. Die Websites wurden einem Land zugerechnet, wenn sie entweder länderspezifische Top-Level-Domains wie .de nutzten oder im WHOIS Protokoll eine dem Land zuordenbare Adresse angegeben war. „Wir haben die

Tools Google Lighthouse und Puppeteer genutzt, um die insgesamt 200.000 Websites zu crawlen“, erzählt Bhuiyan. „Konkret untersucht wurde dann die Websitegröße und Komplexität, die Performance-Optimierung, Sicherheitsmaßnahmen wie die Nutzung von https statt http, Datenschutzanwendungen wie der Rückgriff auf Cookies sowie die Integration aktueller technologischer Features.“

Unklares Bild: Unterschiede geringer und weniger deutlich als erwartet

„Das Ergebnis unserer Untersuchung ist, dass Websites in Entwicklungs- und Schwellenländern grundsätzlich kleiner und einfacher aufgebaut sind als die in Industrienationen“, erklärt Bhuiyan. „Das kommt der Nutzung aus dem mobilen Internet entgegen, das in diesen Ländern weit verbreitet ist. Gleichwohl sind die Websites in einigen Punkten weniger optimal programmiert. So finden sich ineffiziente Bildformate sowie unnötiger Code, und es fehlt oft ein responsives Design. Auch die Nutzung von https, was den verschlüsselten Verbindungsaufbau ermöglicht, ist weniger verbreitet.“ Erstaunt hat den CISPA-Forscher, dass die Unterschiede zwischen beiden Gruppen gleichwohl nicht so stark waren wie erwartet: „Zum Teil sind die Unterschiede zwischen den einzelnen Ländern innerhalb einer Gruppe größer als zwischen den Gruppen“, so Bhuiyan.

Aufschlussreich waren auch die Ergebnisse zur Nutzung von Trackern und Cookie-Bannern. „Wir haben auf den Websites aus Industrienationen mehr Tracker gefunden als bei denen aus Entwicklungs- und Schwellenländern“, so Bhuiyan. „Der Grund dafür ist, dass Industrieländer tendenziell ausgeklügeltere Werbestrategien anwenden, die stark auf Tracker setzen, selbst bei strengeren Datenschutzgesetzen.“ Ein unerwarteter Trend fand sich bei Schwachstellen. So enthalten Websites in Industrieländern tendenziell mehr anfällige Bibliotheken, die theoretisch ausgenutzt werden könnten. „Eine Erklärung dafür könnte die größere Bedeutung von JavaScript-Bibliotheken bei den Websites in den Industrienationen sein“, erklärt der CISPA-Forscher. „Denn diese verbessern nicht nur die Funktionalität der Websites, sondern erhöhen auch deren Angriffsflächen.“

Ausblick und weitere Forschungsdesiderate

„Das interessante an unserer Studie ist“, erklärt Bhuiyan, „dass wir nicht diesen einen großen Faktor gefunden haben, der sich zwischen den Ländern unterscheidet. Aus diesem Grund ist weitergehende Forschung erforderlich. Ein großer Erfolg ist jedoch, dass wir diesen riesigen Datensatz zusammenstellen konnten, den andere Forschende jetzt nutzen können.“ Auf der Entwicklerplattform Github steht der Datensatz mit den 200.000 gecrawlten Websites Interessierten zum Download zur Verfügung. Zum ersten Mal finden sich dort umfassende Datensätze mit Websites aus Ländern wie Nigeria, Bangladesch oder den Philippinen:

alles Regionen, die bisher eher nicht im Fokus der IT-Sicherheitsforschung standen.

Bhuiyan möchte sich in Zukunft in seiner Forschung unter anderem auf Websites aus Südost-Asien konzentrieren. „In der aktuellen Studie ist uns aufgefallen, dass viele Websites aus Indien, Pakistan und Bangladesch auf Englisch sind“, erklärt er. „Das Problem dabei ist, dass nur ein kleiner Teil der Bevölkerung dort Englisch spricht. Wir wollen untersuchen, welche Auswirkungen die Sprachnutzung auf die Barrierefreiheit der Websites sowie den Umgang mit Sicherheitshinweisen hat.“ Damit kann der CISPA-Forscher seinem Antrieb weiter nachgehen, das Internet so inklusiv wie möglich zu gestalten.

»Wir haben auf den Websites aus Industrienationen mehr Tracker gefunden als bei denen aus Entwicklungs- und Schwellenländern.«

Bhuiyan, Masudul Hasan Masud; Varvello, Matteo; Staicu, Cristian-Alexandru; Zaki, Yasir (2025): Digital Disparities: A Comparative Web Measurement Study Across Economic Boundaries. In: WWW 2025, 28 April–2 May, 2025, Sydney, Australia, Conference: The ACM Web Conference

Forscher: ***Masudul Bhuiyan***
Autor: ***Felix Koltermann***

Veröffentlichung
29.04.2025

© Chiara Schwarz

Informationen zur Handyeinrichtung und zu Cybersicherheitsthemen liegen vor allem in verschriftlicher Form vor. Aber was passiert, wenn die Zielgruppe arm und wenig alphabetisiert ist? CISPA-Forscher Sumair Hashmi und seine Kolleg:innen haben in einer qualitativen Interviewstudie untersucht, wo und wie sich Menschen mit niedrigem sozioökonomischem Status in Pakistan Informationen beschaffen, um sich vor Cyberangriffen zu schützen. Ihr Paper „Understanding the Security Advice Mechanisms of Low Socioeconomic Pakistanis“ präsentierten Hashmi und seine Kolleg:innen auf der Conference on Human Factors in Computing Systems (CHI) 2025. Dort erhielten sie auch eine lobende Erwähnung für den Best Paper Award.

Cybersicherheits-praktiken von Menschen mit niedrigem sozioökonomischem Status in Pakistan

Sumair Hashmi

Pakistan ist ein Land mit sehr großen sozialen Unterschieden, insbesondere was das Einkommen und das Bildungsniveau angeht. „Am unteren Ende der sozialen Leiter stehen Menschen, die als Reinigungspersonal, im Haushalt der Mittel- und Oberschicht oder in Fabriken arbeiten“, erklärt CISPA-Forscher Sumair Hashmi. „Gleichzeitig stellen sie die Mehrheit der Gesellschaft dar. Geringes Einkommen geht dabei oft auch mit einer niedrigen Alphabetisierungsrate einher. Mich hat interessiert, welchen Cybersicherheitsgefahren diese Menschen ausgesetzt sind, wie sie sich vor Angriffen schützen und was Cybersicherheit und Datenschutz für sie bedeutet.“ Denn anders als das Verhalten von Menschen aus den Industrienationen sind Cybersicherheitspraktiken von Menschen aus dem globalen Süden aus nicht-englischsprachigen Kontexten bisher nur wenig erforscht.

Hashmis Forschungsinteresse war dabei von alltäglichen Fragestellungen angetrieben. Wie informieren sich die Menschen über Sicherheit und Datenschutz? Was machen sie, wenn sie einen Betrugsanruf erhalten? Das waren Fragen, die ihn interessierten. „Informationen zu diesen Themen sind in der Regel nur schriftlich und auf Englisch verfügbar“, so der Forscher weiter. „Und weil Menschen aus einkommensschwachen Bevölkerungsgruppen in vielen Fällen weder lesen noch schreiben können und meist nur die lokale Sprache Urdu sprechen, haben sie keinen Zugang zu diesen Informationen.“ Hashmi und seine Kolleg:innen vom CISPA und der Lahore University of Management Sciences in Pakistan entschieden sich, das Thema mit einer qualitativen Interviewstudie zu untersuchen, für die sie 20 pakistanische Teilnehmer:innen aus den oben erwähnten Branchen rekrutierten. Der eigentlichen Interviewstudie voraus ging eine Phase ethnographischer Feldbeobachtungen, um die Lebens- und Arbeitsumstände der Zielgruppe besser zu verstehen.

Die Rolle besser informierter Personen und des Arbeitsumfeldes

Konkret befragt wurden für die Studie zehn Männer und zehn Frauen im Alter zwischen 20 und 49 Jahren und einem mittleren Monatseinkommen von 30.500 Pakistanischen Rupien (ca. 96 Euro). Nicht alle Befragten besaßen eigene Geräte, sondern nutzten etwa Handys von Familienangehörigen oder Bekannten mit. „Als die bedeutendsten vorherrschenden Sicherheitsrisiken haben wir Finanzbetrug und digitale Erpressung identifiziert“, erklärt Hashmi „Wir haben herausgefunden, dass alle Befragten auf besser informierte Personen angewiesen waren, sei es um Benutzerkonten für ihre Telefone und Anwendungen einzurichten oder bei Problemen um Rat zu fragen“, so Hashmi weiter. Hilfestellung kam entweder als Ratschlag oder konkretes Hilfsangebot. „Viele Befragte gaben einer Bezugsperson einfach ihr Handy und baten diese, darauf bestimmte Handlungen wie etwa die Passworteinrichtung auszuführen, anstatt um Rat zu fragen“, erklärt Hashmi die soziale Interaktion. Als Quellen der Ratschläge fungierten vor allem Familienmitglieder und nahe Freund:innen sowie Arbeitskolleg:innen. Das konkrete Arbeitsumfeld beeinflusst dabei die Ratschläge: „Das Reinigungspersonal an Universitäten gab differenziertere und vielfältigere Ratschläge als Fabrikarbeiter:innen. Sie hatten außerdem mehr Möglichkeiten, um sich Rat zu holen, etwa von Kolleg:innen, Vorgesetzten und sogar von Professor:innen und Studierenden auf dem Campus“, erklärt der Forscher. Die Sicherheitshinweise, die die Befragten bekamen, lassen sich in Handlungsaufforderungen und Aufklärungshinweise unterscheiden. Als wichtigste konkrete Tipps, die die Befragten erhielten, identifizierte Hashmi unbekannte Nummern zu vermeiden und zu blockieren, Nachrichten und deren Absender zu überprüfen und zu verifizieren sowie keine persönlichen Vermögenswerte preiszugeben.

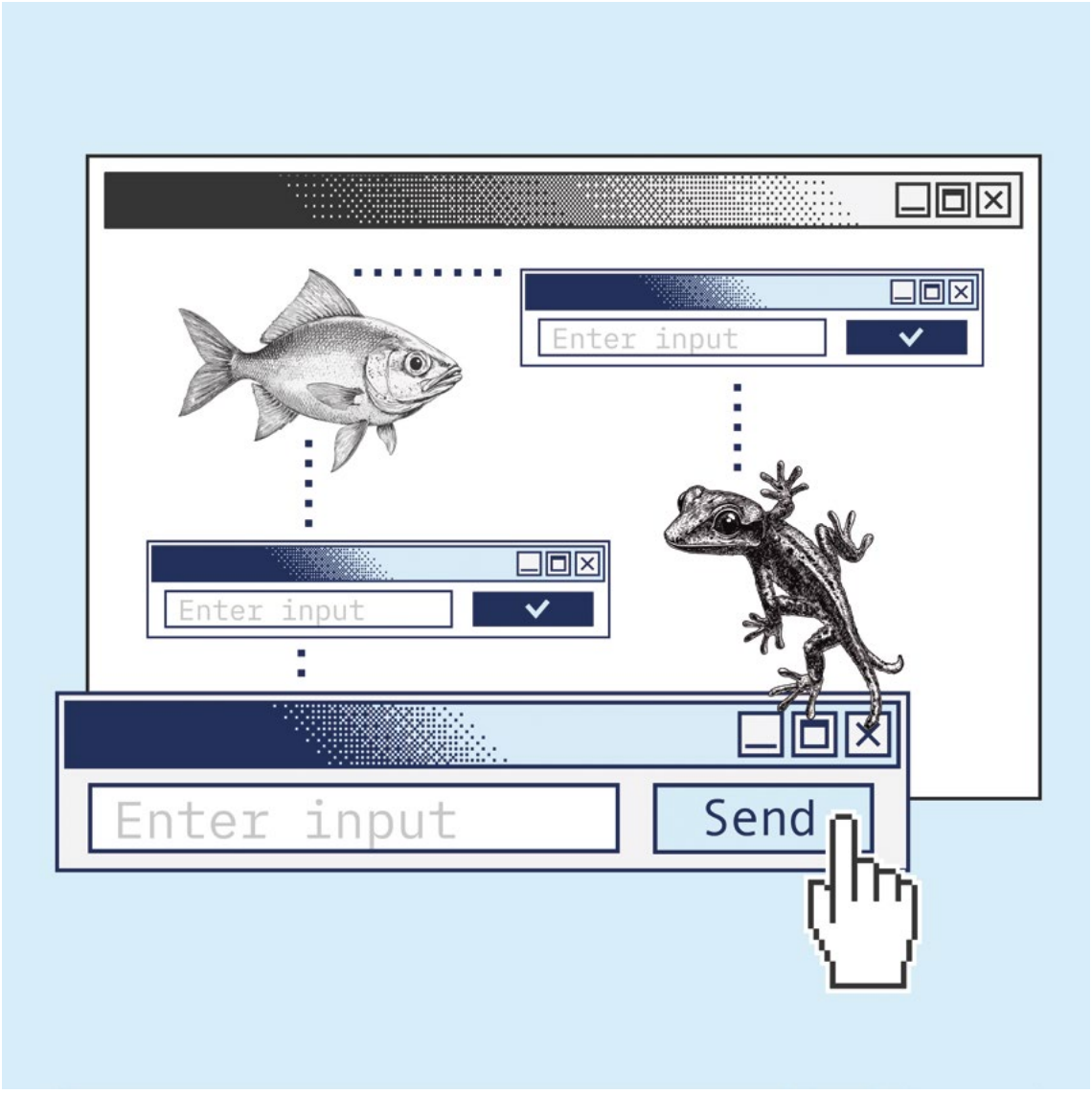
Zusammenfassend verdeutlicht die Studie die extreme soziale Verankerung sicherheitsrelevanter Praktiken von Pakistaner:innen mit niedrigem sozioökonomischem Status. „Unsere Studie zeigt, dass die Art und Weise, wie Ratschläge innerhalb solcher sozioökonomisch benachteiligten Gemeinschaften weitergegeben werden, Einfluss darauf hat, wie Menschen mit digitaler Sicherheit umgehen“, so Hashmi. Eine besondere Rolle spielen dabei die starren Genderrollen sowie die festen Regeln der pakistanischen Klassengesellschaft. So erschweren familiäre Dynamiken und die Angst, verspottet zu werden, das Einholen von Rat. Ratschläge werden dann angenommen, wenn die Ratgeber:innen als kompetent wahrgenommen werden und ein Vertrauensverhältnis besteht.“ Ein weiteres wichtiges Ergebnis war die Besonderheit der Bedrohungslage für die erforschte Zielgruppe in Pakistan: „Aufgrund ihrer schlechten finanziellen Situation sind unsere Studienteilnehmer:innen besonders anfällig für Betrugsmaschen, bei denen sie mit leicht rückzahlbaren Krediten oder

Lotteriegewinnen gelockt werden“, erklärt Hashmi. Damit geht einher, dass die Bedrohungslage für die Befragten eher auf der Ausnutzung menschlicher Schwächen als auf technologischen Schwachstellen beruht.

Menschen mit niedrigem sozio-ökonomischem Status erreichen

Die Bedrohungslage für Cyberangriffe in Entwicklungsländern wie Pakistan ist einzigartig, da Angreifende die finanzielle Notlage der Menschen und deren soziokulturellen Normen ausnutzen. „Um dieser Situation zu begegnen und um Sicherheits- und Datenschutzprobleme von Bevölkerungsgruppen mit niedrigem sozioökonomischem Status zu mindern, müssen neue kontextspezifische Ratschläge und Technologien entwickelt werden“, so Hashmi. „Künftige Forschung sollte prüfen, wie Sicherheitshinweise gezielt besonders benachteiligte Menschen in Pakistan erreichen können, um ihnen mehr Sicherheit zu bieten“, so der Forscher weiter. Er selbst will sich weiter diesem Themenfeld widmen. Hashmi kann sich vorstellen, die Untersuchung auch auf andere Weltregionen auszuweiten. „Wichtig ist, dass wir die Bedürfnisse und die Cybersicherheitspraktiken der Menschen im globalen Süden, die nicht zur westlichen, gebildeten, industrialisierten, reichen und demokratischen (W.E.I.R.D) Bevölkerung gehören, besser verstehen, um passende Empfehlungen zu entwickeln“, so seine Überzeugung.

Hashmi, Sumair Ijaz; Sarfaraz, Rimsha; Gröber, Lea; Javed, Mobin; Krombholz, Katharina (2025): Understanding the Security Advice Mechanisms of Low Socioeconomic Pakistanis. In: CHI 2025, 26 April–1 May 2025, Yokohama, Japan, Conference: Conference on Human Factors in Computing Systems



© Alexandra Goweiler

Das Prinzip vom Überleben der am besten Angepassten, von Charles Darwin im 19. Jahrhundert beschrieben, ist jetzt auf das Testen von Software angewandt worden: FANDANGO, ein neuer Open-Source-Fuzzer, nutzt einen evolutionären Algorithmus, um hochwertige Testeingaben zu erzeugen, die zuvor definierten Anforderungen entsprechen. FANDANGO bringt das sprachbasierte Testen um einen entscheidenden Schritt voran: Der Fuzzer verwendet ein iteratives Verfahren, das der biologischen Evolution nachempfunden ist, um individuell angepasste Eingaben zu liefern. Die CISPA-Forscher José Antonio Zamudio Amaya und Prof. Dr. Andreas Zeller präsentieren das Paper „FANDANGO: Evolving Language-Based Testing“ auf dem International Symposium on Software Testing and Analysis (ISSTA) 2025.

Open-Source-Fuzzer mit evolutionärem Algorithmus erzeugt individualisierte Inputs

José Antonio Zamudio Amaya

In den vergangenen zehn Jahren sind Fuzzer zu den gängigsten Tools für das Testen von Softwaresicherheit und -robustheit geworden. Sie generieren zufällige Testeingaben, speisen diese in Anwendungen ein und helfen so dabei, unerwünschtes Programmverhalten wie Fehler und Schwachstellen aufzudecken. Mit FANDANGO haben José Antonio Zamudio Amaya und CISPA-Faculty Prof. Dr. Andreas Zeller einen bioinspirierten Algorithmus in das Software-Fuzzing eingeführt. Dem Vorbild der biologischen Evolution folgend, führt ihr Algorithmus einen Prozess der Mutation und Selektion aus, um Testeingaben zu erzeugen, die den Anforderungen des Testers genau entsprechen. Zamudio erläutert: „Der evolutionäre Algorithmus ist recht einfach. Wir beginnen mit einer Population von Eingaben, die aus den Spezifikationen des Programms stammen. Dann machen wir zwei Dinge: Erstens mutieren wir diese Eingaben, um verschiedene Veränderungen auszulösen, und zweitens kreuzen wir diese Eingaben, das heißt wir kombinieren Teile von zwei Eingaben, um Nachkommen zu erzeugen. Wir wieder-holen diesen Prozess und bewerten bei jeder Wiederholung die Qualität der Eingaben hinsichtlich ihrer Übereinstimmung mit den Anforderungen des Testers.“ Dieser Prozess führt zu gültigen Testeingaben, die speziell darauf zugeschnitten sind, bestimmte Teile des Programms zu untersuchen.

FANDANGO bietet volle Kontrolle über Test-Inputs

FANDANGO ist das erste Fuzzing-Tool, das Softwaretestern die volle Kontrolle über die Eigenschaften der von ihnen generierten Testeingaben verleiht. Zeller erklärt: „Im Gegensatz zu einem normalen Fuzzer erzeugt FANDANGO Eingaben, die unter der Kontrolle des Testers stehen, da wir davon ausgehen, dass die Tester a) wissen, wie eine typische Eingabe aussieht, und b) in der Regel eine Vorstellung davon haben, wo typische Fehler auftreten könnten. Sie sind diejenigen mit dem Fachwissen, und wir möchten, dass sie dieses Fachwissen beim Testen eines Programms nutzen können.“ Mit FANDANGO können die Tester nicht nur die Syntax der Eingabe, also die gewünschte Struktur, bestimmen, sondern auch die Semantik der Eingabe, also ihre Bedeutung und spezifischen Eigenschaften.

Um die Vorteile von FANDANGO für das Testen von Software zu veranschaulichen, nutzt Zeller das Beispiel eines Online-Shops für maßgefertigte Möbel, in dem Kunden individuelle Werte für Höhe, Länge und Tiefe eingeben müssen, um die Größe eines Möbelstücks zu bestimmen. „In diesem Fall“, erklärt Zeller, „wäre es interessant zu sehen, was das Programm macht, wenn ich beispielsweise sage: ‚Dieses Möbelstück soll eine Länge von weniger als null oder eine Sitzfläche von einem Quadratkilometer haben.‘ Mit unserem evolutionären Algorithmus könnte FANDANGO automatisch Werte für alle diese einzelnen Felder – Höhe, Länge, Tiefe – berechnen, die die Anforderung dieser immensen Fläche von einem Quadratkilometer genau erfüllen würden.“

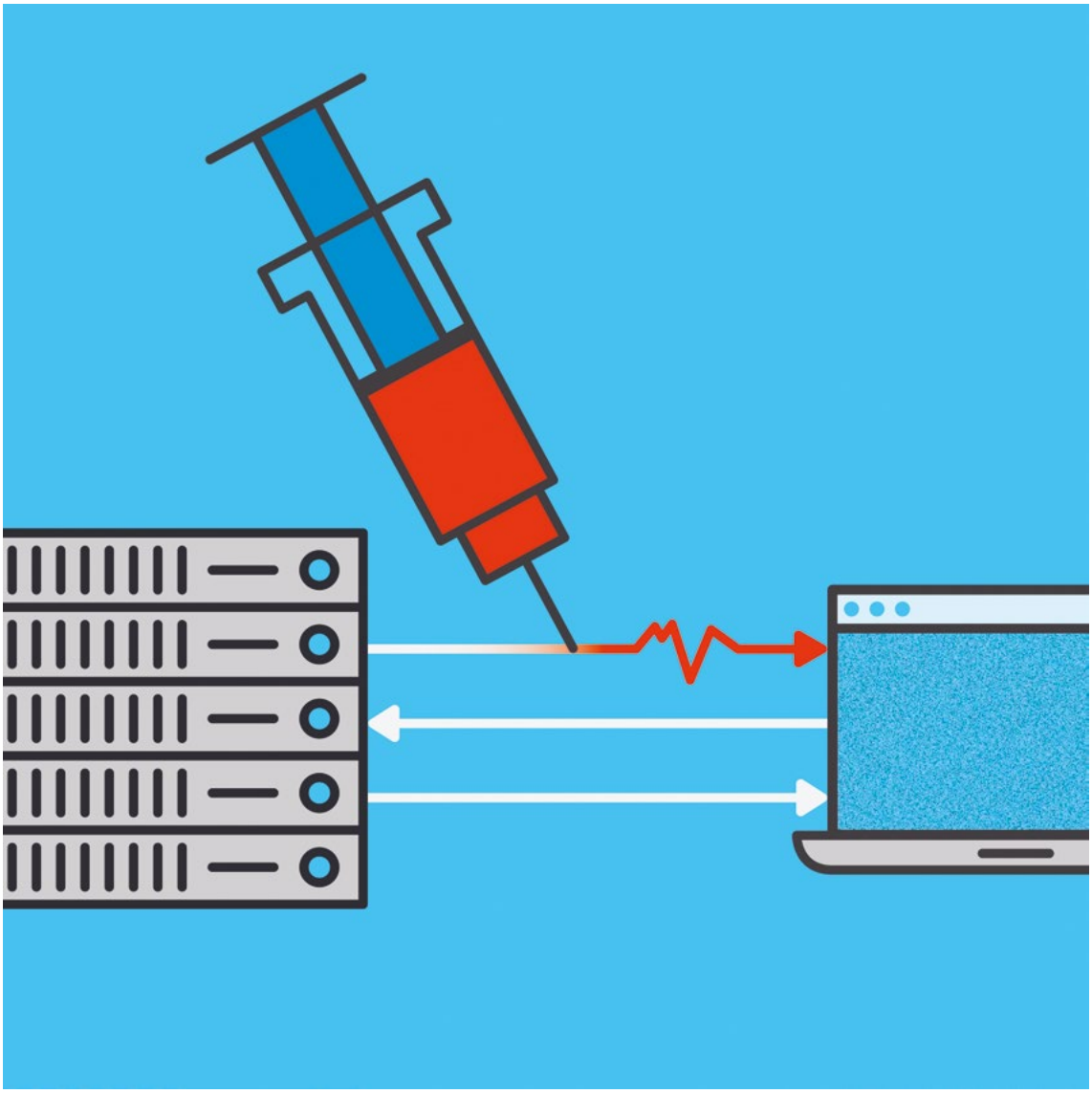
FANDANGO ist auf GitHub verfügbar

Um Softwaretestende und Programmierende von ihrer Forschung profitieren zu lassen, haben Zamudio und Zeller FANDANGO auf GitHub zur Verfügung gestellt. Das Programm ist Open Source und wird in Form eines einfachen Befehlszeilentools samt Tutorials und umfangreicher Dokumentation bereitgestellt. Um ihren Fuzzer noch weiter zu verbessern, bitten die beiden CISPA-Forscher auch um Feedback. „Ich bin schon sehr gespannt darauf, wie die Leute FANDANGO nutzen und welche Verbesserungsvorschläge sie haben. Ich habe bereits mit Leuten aus verschiedenen Unternehmen gesprochen. Die Vorstellung, dass sie die Kontrolle darüber haben, was getestet werden soll, und die Ergebnisse einer Berechnung überprüfen können, ist für sie ein echter Vorteil“, sagt Zeller.

Förderhinweise auf Seite 78

»Der evolutionäre Algorithmus ist recht einfach. Wir beginnen mit einer Population von Eingaben, die aus den Spezifikationen des Programms stammen.«

Amaya, José Antonio Zamudio; Smytzek, Marius; Zeller, Andreas (2025): FANDANGO: Evolving Language-Based Testing. In: ISSTA 2025, 25–28 June, 2025, Trondheim, Norway, Conference: ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)



© Stephanie Bremerich

Programme wie Webbrowser oder Webserver tauschen ständig Daten über das Internet aus. Dadurch sind sie besonders interessant für Angreifer:innen, weswegen wir diese Programme gründlich auf Schwachstellen untersuchen müssen. Doch viele herkömmliche Testmethoden stoßen dabei an Grenzen: Sobald Nachrichten verschlüsselt sind oder die Kommunikation zu kompliziert wird, versagen sie. Genau hier setzt „Fuzztruction-Net“ an. Entwickelt von CISPA-Forscher Nils Bars und seinem Team, verfolgt das neue Verfahren einen cleveren Ansatz: Statt Nachrichten direkt zu verändern, wird einer der Gesprächspartner leicht aus dem Takt gebracht. So lassen sich selbst in weit verbreiteter und gut getesteter Software neue Fehler aufspüren. Sein Paper „No Peer, no Cry: Network Application Fuzzing via Fault Injection“ hat Bars auf der Conference on Computer and Communications Security (CCS) 2024 vorgestellt.

Fuzzing reloaded: mit gezielter Manipulation zu mehr Sicherheit im Netz

Nils Bars

Sogenannte Fuzzer kommen beim Testen von Software ständig zum Einsatz: Es handelt sich dabei um automatisierte Testwerkzeuge, die Programme mit zufälligen oder speziell erzeugten Eingaben füttern, um so unerwartetes Verhalten, Abstürze oder Sicherheitslücken zu finden. Sie sind vor allem bei der Suche nach Fehlern nützlich, die durch eher ungewöhnliche Eingaben entstehen – also Dinge, die bei normalen Tests oft übersehen werden. „Bei Programmen, die sehr klar strukturiert Eingaben verarbeiten oder Dateien einlesen, funktioniert das auch schon ganz gut. Aber Netzwerkprogramme mit Fuzzern zu testen, ist viel komplizierter“, sagt Bars.

Netzwerk-Fuzzing mit Fehlerinjektion: ein neuer Testansatz

Woran liegt das? „Klassische Fuzzer versuchen einen der Gesprächspartner im Netzwerk zu ersetzen, stellen sich dabei aber sehr ungeschickt an: Sie haben kein echtes Verständnis davon, wie sie vorgehen müssen. Sie verstehen weder, wann sie welche Nachrichten senden müssen, noch welche Schlüssel oder Sitzungsdaten gebraucht werden, und können sich auch nicht an frühere Nachrichten erinnern. Das Gegenüber, also entweder Client oder Server, merkt das früher oder später und bricht die Kommunikation ab, bevor die Nachrichten einen Fehler oder eine Sicherheitslücke aufdecken können“, so der Forscher.

Sein Ansatz ist daher nicht wie bei klassischen Fuzzern einen der Gesprächspartner zu ersetzen, sondern ihn stattdessen so geschickt zu manipulieren, dass er sinnvolle, richtig verschlüsselte, aber eben unerwartete Nachrichten produziert. Dieser Trick nennt sich „fault injection“; dabei werden gezielt Fehler in den Programmablauf des Gespächspartners eingebaut. „Das Beste ist, dass wir damit sowohl Server als auch Clients testen können. FuzzstructionNet ist damit der erste Netzwerk-Fuzzer, der das kann. Bisher gibt es solche Fuzzer im Prinzip nur für Server“, sagt Bars.

»Wir haben eine bestimmte Art von Fehlern in den Fokus genommen, sogenannte Memory Corruption. Das sind Programmfehler, bei denen ein Programm Daten in einem Speicherbereich verändert,auf den es gar nicht zugreifen sollte. Das kann zu Abstürzen, Datenverlusten und kritischen Sicherheitslücken führen.«

Mehr Testtiefe und neue Sicherheitslücken

In Tests zeigte der Prototyp von Fuzztruction-Net beeindruckende Ergebnisse: Im Vergleich zu bisherigen Methoden deckte das neue Verfahren im Schnitt 16 Prozent mehr Programmcode ab – ein wichtiges Maß für Testtiefe – und entdeckte dabei dreimal so viele Fehler wie der bisher beste Netzwerk-Fuzzer. Fuzztruction-Net fand dabei auch Schwachstellen in gut getesteten Programmen wie Nginx, dem OpenSSH-Client und cURL. „Wir haben eine bestimmte Art von Fehlern in den Fokus genommen, sogenannte Memory Corruption. Das sind Programmfehler, bei denen ein Programm Daten in einem Speicherbereich verändert, auf den es gar nicht zugreifen sollte. Das kann zu Abstürzen, Datenverlusten und kritischen Sicherheitslücken führen“, so Bars. Insgesamt entdeckte das Forschungsteam 23 neue, auch über das Netzwerk ausnutzbare Sicherheitslücken in verbreiteter Netzwerkinfrastruktur – ein starkes Signal für die Relevanz des neuen Ansatzes.

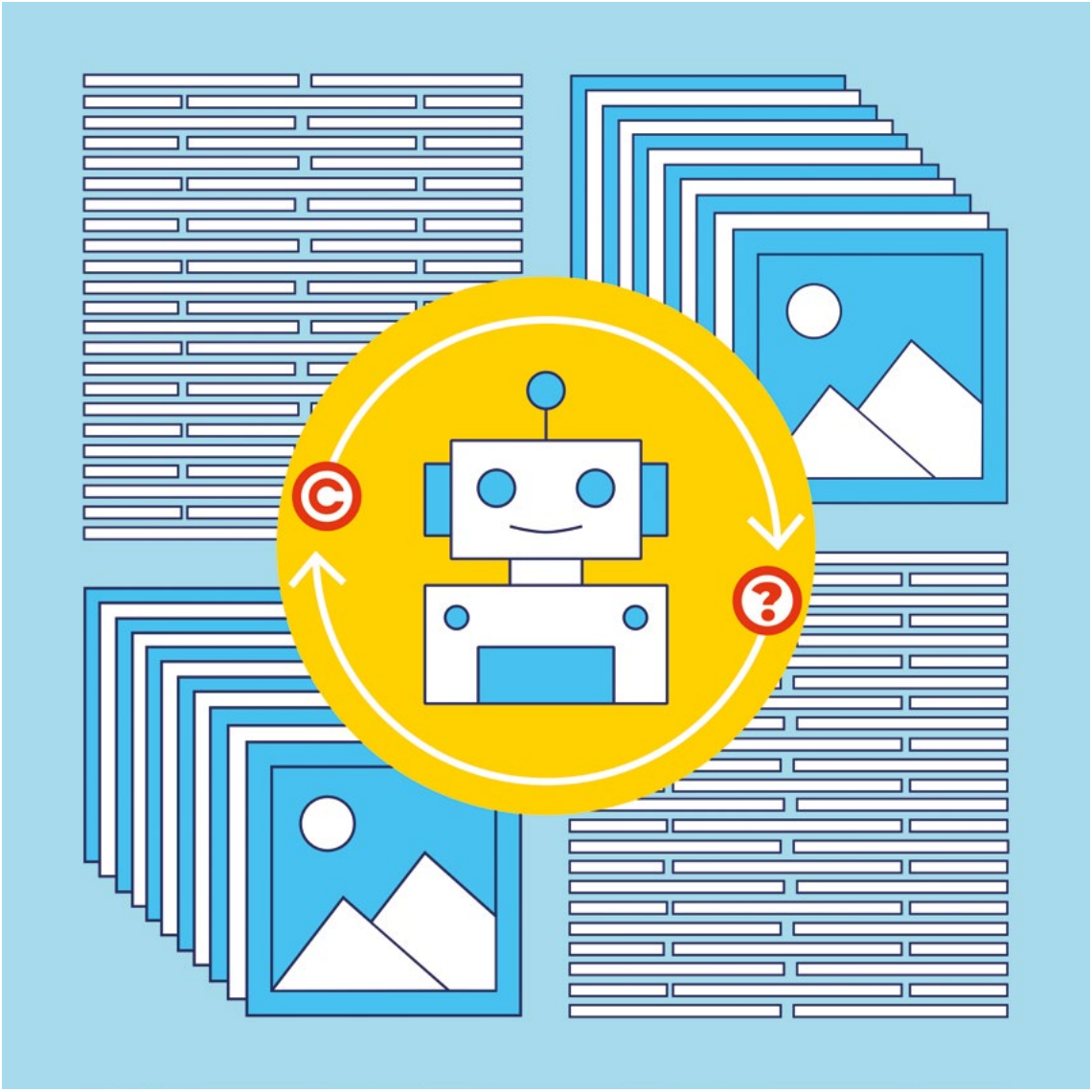
Hilfreiche Hinweise für Entwickler:innen: Fuzzstruction-Net macht Sicherheitslücken nachvollziehbar

Selbst beheben können die Fuzzer die Fehler leider noch nicht. Es braucht also weiterhin die Arbeit von Entwickler:innen, die diesen Fehlern dann nachspüren müssen. „Das Wesentliche ist aber, dass die Fehler reproduzierbar sind. Damit haben Entwickler:innen einen klaren Anhaltspunkt, wo das Problem entsteht, und können es dann beheben“, erklärt Bars. Nginx hat bereits Interesse an der Verwendung von Fuzzstruction-Net gezeigt. „Unser Prototyp funktioniert gut. Für den dauerhaften Einsatz lässt er sich aber sicher noch optimieren“, so Bars. Der Prototyp steht jetzt schon allen Interessierten als Open Source zur Verfügung.

Bars, Nils; Schloegel, Moritz; Schiller, Nico; Bernhard, Lukas; Holz, Thorsten (2024): No Peer, no Cry: Network Application Fuzzing via Fault Injection. In: CCS 2024, 14–18 Oct, 2024, Salt Like City, USA, Conference: ACM Conference on Computer and Communications Security (CCS)

Forscher: **Nils Bars**
Autorin: **Annabelle Theobald**

Veröffentlichung
08.07.2025



© Alexandra Goweiler

In wenigen Jahren ist aus dem wissenschaftlichen Projekt, KI-Modelle zur Generierung von Bildern zu verwenden, eine Alltagsanwendung geworden. Damit tauchen auch neue Probleme auf. Immer mehr Urheber:innen, wie Fotograf:innen oder Illustrator:innen, fragen sich, ob ihre Bilder für das Training von KI-Modellen verwendet wurden. CISPA-Forscher Antoni Kowalczuk hat nun ein Verfahren entwickelt, mit dem nachgewiesen werden kann, ob bestimmte Bilder zum Training eines KI-Modells benutzt wurden. Seine Ergebnisse hat er im Paper „CDI: Copyrighted Data Identification in Diffusion Models“ auf der IEEE Conference on Computer Vision and Pattern Recognition 2025 publiziert.

Neues Verfahren erkennt Nutzung urheberrechtlich geschützter Bilder im KI-Training

Antoni Kowalczuk

KI-Bildgeneratoren haben in den letzten Jahren ein rasantes Wachstum erfahren. Viele der Generatoren wie etwa DALL·E, Midjourney oder Stable Diffusion basieren auf sogenannten Diffusion Models. „Ein Diffusion Model ist ein tiefes neuronales Netz, das lernt, Bilder schrittweise zu erzeugen, indem es nach und nach Rauschen aus dem Bild entfernt“, erklärt Antoni Kowalczuk, PhD-Student und CISPA-Forscher. Trainiert wurden diese Systeme mit Millionen von Bildern aus dem Internet. Dies geschah angeblich ohne Zustimmung der Urheber:innen, was rechtliche und ethische Probleme aufwirft. „Als die Modelle noch rein wissenschaftlichen Zwecken dienten, hat die Urheberrechtsfrage niemanden so wirklich interessiert“, erzählt Kowalczuk. „Aber ab dem Moment, in dem die Leute anfingen, mit den Modellen Geld zu verdienen, wurde das Thema plötzlich relevant. Ich dachte, dass ich da mit meiner Forschung etwas bewirken kann.“

Warum bisherige Methoden versagen

Bisherige Anwendungen, die herausfinden, ob KI-Modelle bestimmte Bilder als Trainingsmaterial verwenden, basieren auf einer Methode namens „Membership Inference Attacks“ (MIA). Diese versuchen zu beurteilen, ob ein einzelnes Bild zum Training eines KI-Modells verwendet wurde. Die Forschung zeigt jedoch, dass die Wirksamkeit solcher Angriffe (MIAs) gegen null geht, sobald die Modelle und ihre Trainingsdaten größer werden, was in der Regel der Fall ist. „Aus diesem Grund habe ich mit meinen Kolleg:innen eine neue Methode namens ‚Copyrighted Data Identification’ (CDI) entwickelt“, erzählt der CISPA-Forscher. „Grundlegend für CDI ist, dass wir nicht einzelne Bilder, sondern ganze Datensätze untersuchen, wie zum Beispiel eine Sammlung von Stockfotos oder ein digitales Kunstportfolio.“

Wie CDI funktioniert

Um zu überprüfen, ob urheberrechtlich geschütztes Material zum Training eines KI-Modells verwendet wurde, hat Kowalczuk für CDI ein vierstufiges Verfahren konzipiert. Zuerst müssen zwei Datensets zusammengestellt werden: „Im ersten sind Bilder enthalten, von denen der Dateninhaber glaubt, dass sie zum Training dieses spezifischen

Modells verwendet wurden. Das Zweite ist ein sogenannter Validierungssatz, der aus Bildern besteht, bei denen wir uns zu 100 % sicher sind, dass sie nicht beim Training verwendet wurden“, erklärt der Forscher. Anschließend lässt man beide Datensätze durch das KI-Modell laufen, um dessen Reaktionen zu beobachten. Auf Grundlage dieser Reaktionen wird ein Werkzeug trainiert, das erkennen kann, ob der betroffene Datensatz wahrscheinlich Teil der Trainingsdaten war. „Am Ende wird ein statistischer Test durchgeführt, um zu prüfen, ob die betroffenen Daten systematisch höhere Werte erzielen als die unveröffentlichten“, so der Forscher. Ist das der Fall, spricht das stark dafür, dass die KI mit diesen Daten trainiert wurde; ist das nicht der Fall, bleibt das Ergebnis offen.

Der CISPA-Forscher testete CDI an einer Reihe bestehender KI-Modelle, für die Informationen über die Trainingsdaten vorliegen: zum Beispiel Modelle, die mit dem Image Net-Datensatz trainiert wurden. Dabei nutzte er sowohl echte Bilddatensätze (etwa aus der Open-Images-Datenbank) als auch gezielt manipulierte Testdaten. Die Ergebnisse sind vielversprechend, erzählt Kowalczuk: „CDI kann mit hoher Genauigkeit erkennen, ob ein Datensatz zum Training benutzt wurde, auch bei komplexen, großen Modellen. Selbst wenn wir die exakten Bilder, die zum Training verwendet wurden, nicht eindeutig identifizieren können, lässt sich dennoch zuverlässig erkennen, ob Daten aus dem Datensatz zum Training des Modells verwendet wurden. CDI liefert auch dann zuverlässige Ergebnisse, wenn nur ein Teil des Gesamtwerks im Training genutzt wurde.“

Hürden für die Anwendung und den Transfer in die Praxis

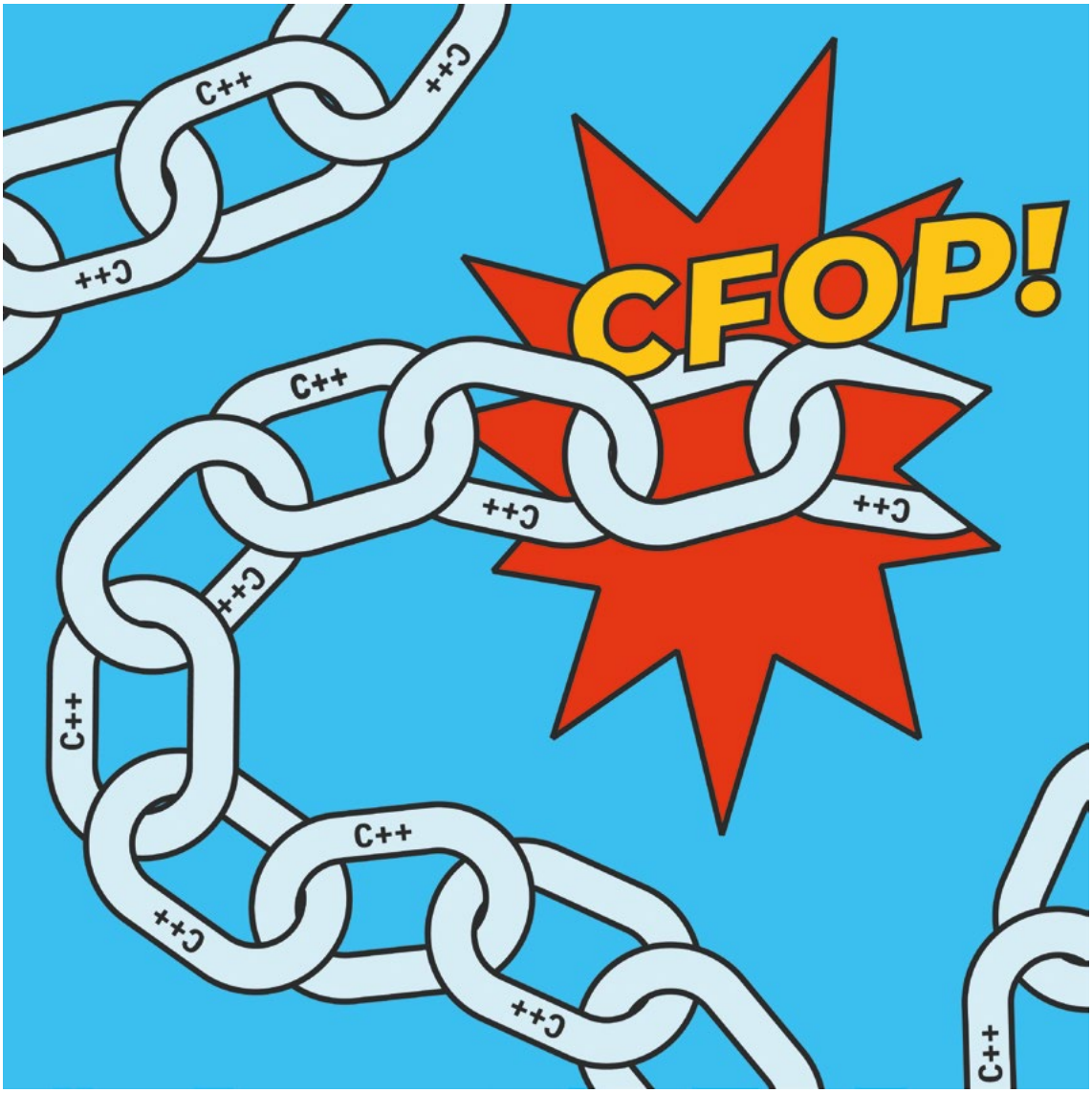
Im Moment ist CDI noch eine Methode, deren Anwendung aufgrund ihrer Komplexität vor allem Wissenschaftler:innen vorbehalten ist. „Einige der von uns extrahierten Merkmale erfordern vollständigen Zugriff auf das Modell und seinen Code“, so Kowalczuk. „Darüber hinaus gibt es einige sehr wichtige Kriterien für die von uns verwendeten Datensamples.“ Insofern liefert CDI im Moment vor allem einen theoretischen Nachweis, dass es möglich ist herauszufinden, ob ein bestimmter Satz von Bildern zum Training von KI-Modellen verwendet wurde. Zur Entwicklung einer Anwendung, die auch Urheber:innen ohne großes technisches Know-how nutzen können, wären weitere Modifikationen und Entwicklungen notwendig, die im Moment jedoch technisch (noch) nicht lösbar erscheinen. „CDI ist noch ziemlich jung, und es gibt noch viel zu tun. Aber eines ist klar: Wenn wir bessere Methoden haben, werden wir vielleicht irgendwann die Brücke von der Theorie zur Umsetzung überschreiten“, zeigt sich der CISPA-Forscher überzeugt.

»Grundlegend für CDI ist, dass wir nicht einzelne Bilder, sondern ganze Datensätze untersuchen, wie zum Beispiel eine Sammlung von Stockfotos oder ein digitales Kunstportfolio.«

Dubiński, Jan; Kowalczuk, Antoni; Boenisch, Franziska; Dziedzic, Adam (2025): CDI: Copyrighted Data Identification in Diffusion Models. In: CVPR 2025, 11–15 June, 2025, Nashville, USA, Conference: IEEE Conference on Computer Vision and Pattern Recognition

Forscher: ***Antoni Kowalczuk***
Autor: ***Felix Koltermann***

Veröffentlichung
17.07.2025



© Chiara Schwarz

Ein Code-Reuse-Angriff namens „Coroutine Frame-Oriented Programming (CFOP)“ kann C++-Coroutinen über drei wichtige Compiler hinweg ausbeuten, nämlich Clang/LLVM, GCC und MSVC. CFOP ist sogar in Umgebungen erfolgreich, die durch Control Flow Integrity (CFI) geschützt sind, und zeigt somit relevante Sicherheitslücken in 15 dieser Abwehrmechanismen auf. Anstatt neuen Code einzuschleusen, verkettet CFOP bereits vorhandene Funktionen miteinander und erreicht die Ausführung beliebigen Codes, nachdem er coroutine-interne Speicherstrukturen beschädigt hat. Die Angriffsart wurde von den CISPA-Forschern Marcos Sanchez Bajo und Prof. Dr. Christian Rossow entdeckt. Ihr Paper „Await() a Second: Evading Control Flow Integrity by Hijacking C++ Coroutines” stellen sie auf dem Usenix Security Symposium 2025 vor.

C++-Coroutinen: anfällig für Code-Reuse-Angriffe trotz CFI

Marcos Sanchez Bajo

Mit einem neuartigen Code-Reuse-Angriff haben Marcos Sanchez Bajo und CISPA-Faculty Prof. Dr. Christian Rossow gezeigt, dass alle bestehenden Implementierungen von C++-Coroutinen ausgenutzt werden können, um moderne CFI-Schutzmaßnahmen in Linux und Windows zu umgehen. Der Angriff namens „Coroutine Frame-Oriented Programming“ (CFOP) führt zu einer Beschädigung des Heap-Speichers, wodurch die Angreifenden Daten manipulieren und die vollständige Kontrolle über Anwendungen übernehmen können. Coroutinen sind eine relativ rezente Ergänzung in C++, aber bereits in mehr als 130 beliebten GitHub-Repositorys vorhanden. „Sie werden verwendet, um Funktionen anzuhalten und wiederaufzunehmen“, erklärt Bajo, „was für die asynchrone Programmierung, wie beispielsweise in Servern, Datenbanken und Webbrowsern, sehr nützlich ist.“

C++-Coroutinen verketten, um den Heap-Speicher zu beschädigen

Mithilfe von Coroutinen können beispielsweise Generatoren erstellt werden, die eine Folge von Elementen erzeugen, wie etwa die Fibonacci-Folge. In der Fibonacci-Folge ist jede neue Zahl die Summe der beiden vorangegangenen Zahlen. Nach jeder neuen Zahl in der Folge wird die Coroutine angehalten, bis sie dazu aufgerufen wird, die nächste Zahl zu generieren. Bei CFOP werden ganze C++-Coroutinen und andere vorhandene Funktionen dazu genutzt, einen Code-Reuse-Angriff zu erstellen, wie Bajo erklärt: „Bei Code-Reuse-Angriffen verwenden Angreifende in der Regel Code-Schnipsel, die ohnehin schon zur Anwendung gehören, sodass kein neuer Code eingeschleust wird. Aus diesen Code-Schnipseln bilden sie dann Ketten, um den Ausführungsfluss des Programms zu manipulieren. Die Umgehung von CFI-Schutzmaßnahmen ist allerdings etwas schwieriger. Anstatt nur Code-Schnipsel zu nehmen und Ketten zu bilden, muss man vollständige Coroutine-Funktionen nehmen und sie auf intelligente Weise miteinander verbinden.“ Sobald die CFI-Schutzmaßnahmen auf diese Weise durch das Hijacking einer Coroutine-Funktion umgangen sind, kann jede andere vorhandene Funktion einem Code-Reuse-Angriff unterzogen werden.

Control Flow Integrity (CFI) versagt beim Schutz von C++-Coroutinen

CFI-Mechanismen wurden zum Schutz vor Code-Reuse-Angriffen eingeführt und sollen sicherstellen, dass der korrekte Programmablauf eingehalten wird. Programmiersprachen entwickeln sich jedoch dynamisch weiter, während CFI-Mechanismen nur die zum Zeitpunkt ihrer Erstellung vorhandenen Programmierparadigmen schützen. Bajo betont: „Das Hauptproblem bei CFI ist, dass es eine statische Abwehr ist, das heißt, es deckt nur die Möglichkeiten einer Programmiersprache in ihrer bestehenden Form ab. Wenn zu einem späteren Zeitpunkt neue Funktionen in die Programmiersprache eingeführt werden, kann CFI diese nicht erkennen. Es kann nicht mit ihnen umgehen, da es auf einer älteren Version der Programmiersprache basiert.“ In ihrer Studie fanden Bajo und Rossow heraus, dass nur sieben von 15 ursprünglich betrachteten CFI-Mechanismen mit Coroutinen kompatibel waren. Von diesen sieben boten nur zwei (IBT und Control Flow Guard) einen teilweisen Schutz vor der Ausnutzung von Coroutinen, während die übrigen fünf gar keinen Schutz boten. „Letztendlich“, fasst Bajo zusammen, „waren wir in der Lage, alle zu umgehen. Mit CFOP kann man weiterhin all das tun, was vor CFI möglich war.“

Das Patchen von CFOP ist ein strukturelles Problem

Die Tatsache, dass C++-Coroutinen sich zunehmender Beliebtheit erfreuen, verstärkt das Potenzial von CFOP. Bajo erklärt: „Coroutinen wurden 2020 in C++ eingeführt und werden seitdem immer häufiger von Entwickler:innen verwendet. Leider haben wir festgestellt, dass Coroutinen bestimmte Strukturen im Speicher aufweisen, auf die Angreifende abzielen können. Soweit wir wissen, wurde dies bisher aber noch nicht in der Praxis ausgenutzt.“ Im Wesentlichen ist CFOP möglich, weil die drei wichtigsten Compiler C++-Coroutinen auf eine Art und Weise implementieren, die sie strukturell anfällig macht. Bajo fährt fort: „Diese Angriffstechnik zu entschärfen, ist nicht so einfach wie das Patchen eines Codes – es handelt sich um ein strukturelles Problem, und man muss die interne Funktionsweise der Anwendung überdenken.“ Bajo und Rossow haben erfolgreiche Implementierungsalternativen für C++-Coroutinen entwickelt und diese Abhilfemaßnahmen im November 2024 an Clang/LLVM, GCC und MVSC gemeldet.

»Das Hauptproblem bei CFI ist, dass es eine statische Abwehr ist, das heißt, es deckt nur die Möglichkeiten einer Programmiersprache in ihrer bestehenden Form ab. Wenn zu einem späteren Zeitpunkt neue Funktionen in die Programmiersprache eingeführt werden, kann CFI diese nicht erkennen.«

Sanchez Bajo, Marcos; Rossow, Christian (2025): "Await() a Second: Evading Control Flow Integrity by Hijacking C++ Coroutines". In: 34th Usenix Security Symposium, 13–15 Aug, 2025, Seattle, USA, Conference: USENIX Security Symposium

© Janine Paulus

Stellt man sich Software als Gebäude vor, so könnte man sagen, sie besteht aus Code-Bausteinen. Viele dieser Bausteine sind gebäudespezifisch und werden eigens dafür programmiert. Andere hingegen sind Standardbausteine, die bei vielen Gebäuden benötigt werden, wie etwa kryptographische Algorithmen. Werden diese Krypto-Bausteine mit der Zeit porös, leidet auch die Sicherheit der Software. In einer qualitativen Interviewstudie hat CISPA-Forscher Alexander Krause erhoben, vor welchen Hindernissen Softwareentwickler:innen stehen, wenn sie Krypto-Implementierungen in einer Software erneuern oder gleich bessere Bausteine entwickeln wollen. Das Paper „'That's my perspective from 30 years of doing this': An Interview Study on Practices, Experiences, and Challenges of Updating Cryptographic Code" präsentiert er auf dem Usenix Security Symposium 2025.

Wie agil ist deine Krypto? Interviewstudie zu kryptographischen Updateprozessen

Alexander Krause

Ein fundamentaler Baustein in der Programmierung neuer Anwendungen sind kryptographische Algorithmen. Sie sorgen dafür, dass Daten und Informationen verschlüsselt kommuniziert werden können. Anders als andere Code-Sequenzen veralten bestimmte kryptographische Implementierungen mit der Zeit: Schreitet die Entwicklung auf anderen technologischen Gebieten voran, gewinnen Computer etwa signifikant an Rechenleistung, werden beispielsweise asymmetrische Verschlüsselungen potenziell brechbar. Quanten-Computing ist dafür ein Paradebeispiel. Das Rechnen mit drei statt zwei möglichen Zuständen befähigt Quantencomputer, mathematische Probleme in geringerer Zeit zu lösen.

Das Aktualisieren von Krypto-Implementierungen ist also eine wiederkehrende Aufgabe, und eine mit weitreichender Bedeutung: Gehen Krypto-Updates schief, hat das mitunter gravierende Folgen für die allgemeine Sicherheit der Software. Alexander Krause erwähnt in diesem Kontext das Konzept der „Krypto-Agilität“: „Dieser wiederkehrende Updateprozess von kryptographischen Implementierungen beginnt idealerweise mit etwas, das man „crypto agility“ nennt. Das bedeutet, dass Entwickler:innen schon während des Software-Designs daran denken, dass sie die kryptographische Implementierung vielleicht irgendwann einmal austauschen oder updaten müssen.“ Dieses Vorausdenken soll die spätere Aktualisierung der Software erleichtern. Allerdings erfordern Krypto-Updates ein bereichsspezifisches Wissen, über das längst nicht alle Software-Entwicker:innen verfügen.

Krypto-Bibliotheken wollen gepflegt werden

In der Regel stammen kryptographische Implementationen aus öffentlich zugänglichen, kostenfreien Krypto-Bibliotheken, die von spezialisierten Entwickler-Communities gepflegt werden. Hinter diesen Open-Source-Projekten, von denen Entwickler:innen auf der ganzen Welt profitieren, stehen oft nur wenige Köpfe, die sich in der Regel unentgeltlich für das Projekt engagieren. Während die Wiederverwendung bestehender Algorithmen und Funktionen effiziente Programmierarbeit erlaubt, birgt sie für die Kryptographie besondere Sicherheitsrisiken.

Denn wenn Krypto-Bibliotheken nicht richtig gepflegt und Fehler nicht behoben werden, proliferieren diese Schwachstellen in einer Vielzahl von Anwendungen.

Krypto-Updates und ihre Herausforderungen

Vor welchen Fragen und Herausforderungen Software-Entwickler:innen – die selbst zumeist keine Krypto-Expert:innen sind – beim Updaten von Krypto-Implementierungen stehen, haben Krause und seine Kolleg:innen anhand einer qualitativen Interviewstudie mit 21 Teilnehmenden erhoben. So sollten Antworten auf vier eng definierte Forschungsfragen gefunden werden: Wie erlangen Entwickler:innen Kenntnis von einem empfohlenen Krypto-Update? Welche Ziele verfolgen sie damit? Welche Prozesse durchlaufen sie bei der Planung und Durchführung eines Krypto-Updates? Und schließlich: Welche Erfahrungen haben sie beim Durchführen dieser Updates gemacht? „Zu dem reinen Updaten von Softwareprojekten gibt es schon viel Forschung. Aber wir haben uns hier die Frage gestellt, ob auch Experten-populationen, die über ein hochspezialisiertes Wissen verfügen, besondere Anforderungen haben“, so Krause.

Heterogene Studienergebnisse: Krypto-Updates sind kontext-abhängig

Es zählt zu den wichtigsten Ergebnissen der Interviewstudie, dass der Informationsfluss rund um Krypto-Updates uneinheitlich und zum Teil lückenhaft ausfällt. Zu den Auslösern eines Updates zählten vorrangig Informationen, die den Entwickler:innen über Blogs, Social Media und GitHub zugingen. Anhängig von institutioneller Zugehörigkeit erhalten einige Entwicklergruppen jedoch eher relevante Informationen über Updates als andere Kolleg:innen. „Wenn man in einem großen Unternehmen arbeitet, dann bestehen häufig Absprachen. Die erhalten oft vorab Informationen zu Sicherheitslücken und können die als erstes schließen, zum Beispiel im Rahmen eines Disclosure-Prozesses. Diese Informationen kommen dann über private Mailinglisten, auf die nur wenige Zugriff haben“, fasst Krause zusammen. „Es war ein wichtiges Takeaway für uns, dass es schwer ist, in diese Communities reinzukommen.“

Die Interviewstudie ergab zudem, dass es in Unternehmen und Projekten kaum strukturierte Prozesse gibt, um Krypto-Updates zu regeln. Die Priorisierung solcher Updates hing mitunter von Arbeitsressourcen wie beispielsweise Teamgröße ab. Zuweilen waren die Entscheidungsprozesse und Zuständigkeiten rund um Krypto-Updates unklar. „Wer entscheidet darüber, wer die Verantwortlichkeit für ein Krypto-Update übernimmt? Das ist sehr unterschiedlich gewesen“, sagt Krause. „Manchmal gab es tatsächlich Führungskräfte, die dafür verantwortlich waren. In anderen Situationen hieß es aber, ‚du hast doch gerade selber festgestellt, dass wir diese Sicherheitslücke haben, dann ist es auch deine Aufgabe, das Problem zu beheben‘.

“ Als einen zentralen Forschungsbeitrag haben die Forschenden einen solchen Update-Prozess skizziert und so die heterogenen Aussagen der Teilnehmenden zusammengeführt.

Andere Studienergebnisse fielen positiver und erwartbarer für die Forschenden aus. Ein Beispiel hierfür sind die Ziele, die mit Krypto-Updates verfolgt werden. „Hier waren wir insgesamt positiv überrascht, dass viele Entwickler:innen das aus der intrinsischen Motivation heraus machen, dass ihre Software zukunftssicher sein soll“, erklärt Krause. Zudem wurden präventive Updates vorgenommen, um einen Sicherheitsvorsprung gegenüber zukünftigen Bedrohungen zugewinnen. Recht einheitlich war auch die Rückmeldung, dass Krypto-Updates als beschwerlich und komplex wahrgenommen werden. Krause fasst zusammen: „Alle unsere Teilnehmenden hatten einen sehr individuellen Background und sehr individuelle Projekte, aber im Großen und Ganzen macht das Updaten von Krypto schwer, dass man das Wissen braucht, und das haben ganz viele letztlich nicht.“

Netzwerk ist alles: Eine Lücke zwischen Forschung und Praxis

Die Frage, wie man diese Wissenslücke im Sinne der IT-Sicherheit schließen könnte, beschäftigt Krause nachhaltig. „Die größte Herausforderung, die wir sehen – und das bezieht sich nicht nur auf unser Paper, sondern allgemein auf die Krypto-Forschung – ist, neue Forschungserkenntnisse in ein Format zu übertragen, in dem sie die Developer:innen auch erreichen.“ Während der Zugang zu den einschlägigen Mailinglisten schwer zu erlangen ist, haben die Antworten aus der Interviewstudie zudem gezeigt, dass Software-Entwickler:innen sich kaum in wissenschaftlichen Publikationsdatenbanken informieren. „In unserer Studie hatten hier diejenigen einen Vorteil, die einen hohen Bildungsabschluss hatten, einen Master oder einen PhD, denn die bringen das Skillset dafür mit“, erklärt Krause. Letzen Endes bleibt die Beschaffung relevanter Informationen zu einem großen Teil abhängig von der persönlichen Initiative der einzelnen Entwickler:innen. In dieser Hinsicht besteht eine Lücke zwischen Forschung und Praxis, die es zu überwinden gilt. Die CISPA-Forschenden haben ihre Ergebnisse allen Entwickler:innen, die an ihrer Interviewstudie teilgenommen haben, zur Verfügung gestellt.

Förderhinweise auf Seite 78

Krause, Alexander; Kaur, Harjot; Klemmer, Jan; Wiese, Oliver; Fahl, Sascha (2025): “That’s my perspective from 30 years of doing this”: An Interview Study on Practices, Experiences, and Challenges of Updating Cryptographic Code. In: 34th Usenix Security Symposium, 13–15 Aug, 2025, Seattle, USA, Conference: USENIX Security Symposium

© Chiara Schwarz

CISPA-Forscher Sarath Sivaprasad hat zusammen mit Hui-Po Wang und Mario Fritz vom CISPA sowie weiteren Kolleg:innen vom Helmholtz-Institut für Pharmazeutische Forschung (HIPS) ein KI-Modell entwickelt, das Entwicklungsstörungen in der Embryonalentwicklung von Zebrafischen automatisch erkennen kann. Der Ansatz kombiniert einen groß angelegten, hochauflösenden Bilddatensatz mit einem Transformer-basierten Modell des maschinellen Lernens, um Toxizitätseffekte und Fruchtbarkeit mit hoher Genauigkeit und Effizienz zu identifizieren. Dieser Fortschritt könnte die Prozesse der Wirkstoffentwicklung für Arzneimittel deutlich beschleunigen. Das Paper „Automated Detection of Abnormalities in Zebrafish Development" wird auf der Konferenz für Medical Image Computing and Computer Assisted Intervention (MICCAI) 2025 vorgestellt.

KI beschleunigt Medikamentenentwicklung durch automatische Analyse von Zebrafisch-Embryonen

Sarath Sivaprasad

Die Anomalieerkennung ist schon seit einiger Zeit ein Schwerpunkt von Sivaprasads Forschung. „Im Maschinellen Lernen versteht man unter Anomalieerkennung den Prozess, Datenpunkte, Ereignisse oder Muster zu identifizieren, die erheblich vom erwarteten Verhalten abweichen“, erklärt er „Während des Trainings lernt das System, wie ‚normal‘ aussieht. Bei der Auswertung wird jede Probe danach bewertet, wie stark sie von diesem Normalbild abweicht. Im Gegensatz zur klassischen Klassifikation, die Eingaben bestimmten Kategorien zuordnet (z. B. Katze, Hund oder Auto), geht es bei der Anomalieerkennung darum, zwischen ‚A‘ und ‚nicht A‘ zu unterscheiden.“ In seiner aktuellen Veröffentlichung wird ein ähnliches Konzept auf die biomedizinische Forschung angewandt. „In diesem Fall haben wir eine Variante der Anomalieerkennung eingesetzt, um die Entwicklung von Zebrafisch-Embryonen zu beobachten“, erläutert der Forscher.

Zebrafisch: ein kleiner Kraftprotz in der Wirkstoffforschung

„Zebrafische sind ein hervorragender Modellorganismus für die biomedizinische Forschung“, sagt Sivaprasad. „Das liegt an ihren durchsichtigen Körpern und den genetischen Ähnlichkeiten zum Menschen.“ Ihre schnelle Entwicklung und ihre Reaktionsfähigkeit auf Chemikalien machen sie ideal für Hochdurchsatz-Screenings zur Toxizitätsbestimmung, eine wichtige Methode in der Wirkstoffforschung. „Die Analyse ihrer Entwicklung beruht jedoch weiterhin stark auf fachkundiger manueller Begutachtung, was ein zeitaufwändiger und subjektiver Prozess ist.“ Die Herausforderung besteht darin, subtile Entwicklungsstörungen, die sich im Laufe von Bildsequenzen herausbilden, zuverlässig zu erkennen. „Bei bestehenden Datensätzen fehlt es sowohl an zeitlicher Abdeckung als auch an dem Umfang, der erforderlich ist, um groß angelegte Modelle zu trainieren“, ergänzt Sivaprasad.

Ein bahnbrechender Datensatz und ein Modell für die Zukunft

Um diesen Engpass zu überwinden, stellten Sivaprasads Kolleg:innen am HIPS zunächst einen der umfassendsten Bilddatensätze zur embryonalen Entwicklung des Zebrafischs zusammen – bestehend aus mehr als 185.000 mikroskopischen Aufnahmen. „Sie haben Zebrafisch-Embryonen auf Objektträgern platziert, sie unter dem Mikroskop beobachtet und ihre Entwicklung kontinuierlich erfasst“, erklärt er. Der Datensatz deckt zwei entscheidende Experimente ab:

- Fruchtbarkeitsklassifikation: 130.368 Bilder über 8-Stunden-Sequenzen zur Bestimmung der Lebens fähigkeit von Eiern.

- Toxizitätsbestimmung: 55.296 Bilder über 48 Stunden, um die Wirkung toxischer Verbindungen zu erkennen.

Bilder zur Fruchtbarkeitsbestimmung wurden mit Sequenz-Labels versehen, während Entwicklungsanomalien fein abgestufte zeitliche Markierungen erhielten. Damit wurde ein wertvoller Benchmark für die Entwicklung und Testung automatisierter Tools geschaffen.

Transformer-basierte KI steigert die Genauigkeit

Im zweiten Schritt wurde ein KI-Modell mit diesem Datensatz trainiert. Sivaprasad entwickelte ein neues, Transformer-basiertes neuronales Netz, das sowohl die Struktur jedes einzelnen Bildes als auch die zeitlichen Veränderungen in den Sequenzen interpretieren kann. Die KI erreichte eine Genauigkeit von 98 % bei der Erkennung, ob ein Embryo befruchtet war, und 92 % bei der Erkennung von Entwicklungsanomalien, die durch die Exposition gegenüber einer toxischen Verbindung (3,4-Dichloranilin) verursacht wurden. Besonders wichtig ist, dass das Modell nachahmt, wie menschliche Expert:innen die Entwicklungsverläufe über die Zeit analysieren und so frühe Vorhersagen zur Toxizität ermöglicht.

Eine Plattform für zukünftige Innovationen

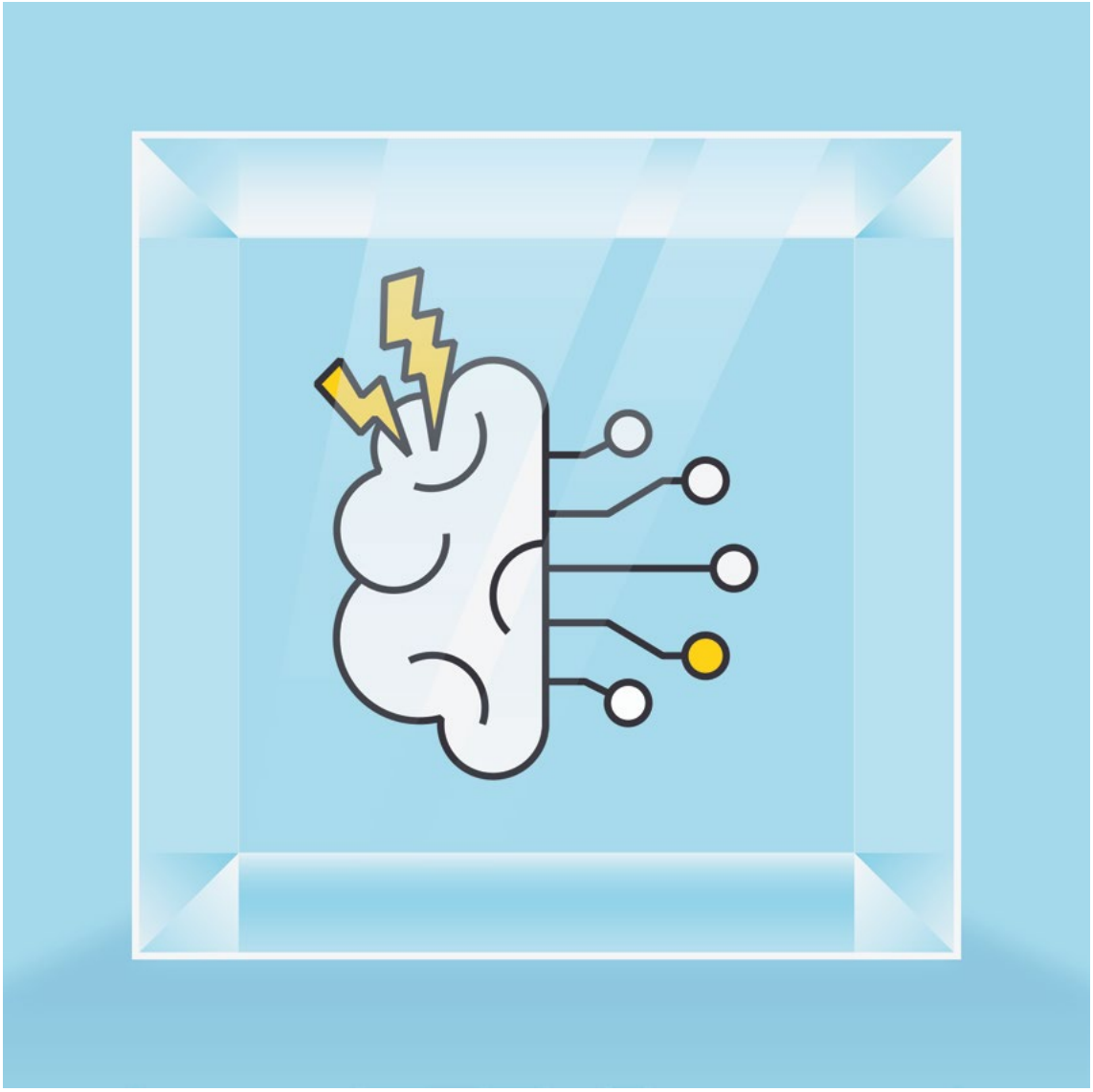
Der Datensatz und das Modell schaffen die Grundlage für zukünftige Forschung zur Toxizität in frühen Entwicklungsstadien und verbessern sowohl die Sensitivität als auch die Geschwindigkeit von Vorhersagen. „Derzeit untersuchen wir nur eine einzelne Chemikalie, um zu verstehen, wie Anomalien sich entwickeln. Unser Ziel ist es jedoch, dies auf eine gesamte Bibliothek von Chemikalien auszuweiten“, sagt Sivaprasad. Der vollständige Datensatz wird frei auf GitHub zur Verfügung gestellt, sodass andere Forschende ihn ohne Kosten nutzen und erweitern können. Ziel ist es, die Community sowohl der biomedizinischen als auch der KI-Forschung zu befähigen, fortschrittlichere, effizientere und ethischere Methoden für das Toxizitäts-Screening zu entwickeln. „Der Daten-

satz ist eine wertvolle Ressource sowohl für die Machine-Learning-Community, um ihre Methoden zu bewerten, als auch für die biomedizinische Forschung, um die Wirkung verschiedener Wirkstoffe besser zu verstehen“, ergänzt Sivaprasad.

Förderhinweise auf Seite 78

»Derzeit untersuchen wir nur eine einzelne Chemikalie, um zu verstehen, wie Anomalien sich entwickeln. Unser Ziel ist es jedoch, dies auf eine gesamte Bibliothek von Chemikalien auszuweiten.«

Sivaprasad, Sarath; Wang, Hui-Po; Jäckel, Anna-Lisa; Baumann, Jonas; Baumann, Carola; Herrmann, Jennifer; Fritz, Mario (2025): Automated Detection of Abnormalities in Zebrafish Development. In: MICCAI 2025, 23–27 Sept, Daejeon, Korea, Conference: Medical Image Computing and Computer Assisted Intervention



© Chiara Schwarz

Das Forschungsprojekt Liberate AI vereint Expertisen aus Medizin, Informatik und vertrauenswürdiger KI, um ein KI-Modell zu entwickeln, das Ärzt:innen bei der Behandlung des ischämischen Schlaganfalls unterstützt. Als digitales Assistenzsystem soll es den langfristigen Behandlungserfolg einer mechanischen Thrombektomie sowie mögliche Komplikationen vorhersagen. Das KI-Modell wird privatsphäreschonend anhand medizinischer Daten trainiert, die an verschiedenen Standorten in Deutschland vorliegen. Liberate AI widmet sich zudem der Erklärbarkeit der KI sowie ihrer Fähigkeit, differenzierte Vorhersagen für Patientensubgruppen zu treffen. Liberate AI ist ein gemeinsames Projekt des DZNE, des Universitätsklinikums Bonn und des CISPA. Es wird von der Helmholtz-Gemeinschaft mit 250.000 Euro gefördert.

Von Black Box zu Glasbox: erklärbare KI in der Schlaganfallbehandlung

Jilles Vreeken

Ein ischämischer Schlaganfall tritt auf, wenn sich Blutgerinnsel sich in Hirngefäßen festsetzen und den Blutfluss im Gehirn und somit dessen Sauerstoffversorgung unterbrechen. Eine mögliche Maßnahme in diesem Fall ist die mechanische Thrombektomie, ein minimalinvasiver Eingriff, bei dem das Gefäß mit einem speziellen Katheter wieder geöffnet wird. Ob die mechanische Thrombektomie für die betroffene Person jedoch die vielversprechendste Option darstellt, hängt von einer Vielzahl individueller Faktoren ab. Um Ärzt:innen bei dieser zeitkritischen Entscheidung zu unterstützen, wollen die Forschenden in Liberate AI ein KI-Modell mit medizinischen Daten aus dem Deutschen Schlaganfall-Register sowie den zugehörigen MRT- und CT-Aufnahmen aus verschiedenen deutschen Krankenhäusern trainieren. Hierfür nutzen sie Swarm Learning, eine vom DZNE in Kooperation mit Hewlett Packard Enterprise entwickelte KI-Technologie. Swarm Learning ermöglicht es der KI, dezentral zu lernen: Sie reist virtuell zu allen Datenquellen im Netzwerk und sammelt dort Wissen ein, ohne dass die Daten selbst die Standorte verlassen, an denen sie gespeichert sind.

Auf dem Weg zur Glasbox: Erklärbarkeit ist entscheidend

In Liberate AI werden zudem technologische Herausforderungen adressiert, die über das eigentliche Training des KI-Modells hinausgehen. Die erste große dieser Herausforderungen betrifft die Erklärbarkeit des KI-Modells. Im Gegensatz zu Deep-Learning-Anwendungen, die meist wie eine Black Box funktionieren, muss das KI-Modell in Liberate AI seine Entscheidungsfindung für die behandelnden Ärzt:innen transparent nachvollziehbar machen. Prof. Dr.Jilles Vreeken, Experte für vertrauenswürdige Informationsverarbeitung am CISPA, erklärt: „Wir wollen eine Glasbox-KI entwickeln, die genauso gute Vorhersagen trifft wie eine Black-Box-KI. Denn wenn man Mediziner:in ist und die KI sagt ‚Ja' oder ‚Nein', dann ist die erste Frage, die man stellt: ‚Warum sollte ich dir vertrauen?'. Das bedeutet, dass wir erklärbare KI einsetzen müssen – also den KI-Forschungszweig, in dem wir KI-Modelle entwickeln, bei denen wir nachvollziehen können, auf Grundlage welcher Beweise sie ihre Aussagen treffen. Das ist die Art von KI, die Expert:innen wirklich unterstützen kann, denn Mediziner:innen sind dann in der Lage zu unterscheiden, ob die Vorhersage auf zufälligen Beweisen oder auf echten Biomarkern beruht."

Vreeken und seine Forschungsgruppe haben es sich zum Ziel gesetzt, ein transparentes KI-Modell zu entwickeln. Im Kontext von Swarm Learning bringt Erklärbarkeit jedoch besondere technologische Herausforderungen mit sich. „Wir müssen bedenken, dass wir zwar diese Glasbox-KI entwickeln können, sie muss aber immer noch in der Lage sein, in einer Swarm-Learning-Umgebung zu lernen und genauso zuverlässige Vorhersagen zu treffen wie eine Black-Box-KI. Es ist nicht trivial, das möglich zu machen“, so der CISPA-Forscher. Im Zuge des Projekts müssen die Forschenden daher ein Gleichgewicht finden zwischen dem Transparenzgrad des KI-Modells und seiner Fähigkeit, erfolgreich am Swarm Learning teilzunehmen.

Auf der Suche nach Subpopulationen und kausalen Schlussfolgerungen

Die zweite große Herausforderung, mit der sich die CISPA-Forschenden befassen, betrifft die Identifizierung solcher Patientengruppen, die hinsichtlich ihrer langfristigen Lebensqualität positiv oder negativ auf eine mechanische Thrombektomie reagieren. Im Idealfall wird das KI-Modell in der Lage sein, diese statistischen Subgruppen automatisch anhand bestimmter Muster zu identifizieren, die es aus den gesammelten medizinischen Daten extrahiert. „Die Frage ist: Können wir eine Glasbox-KI entwickeln, die die Bedingungen erkennen kann, unter denen Menschen ein außergewöhnliches Überlebensverhalten zeigen? Zum Beispiel könnte das von der Größe des Blutgerinnsels, hohem oder niedrigem Blutdruck, genetischen Faktoren oder der Einnahme von Blutverdünnern abhängen. Man kann sich verschiedene Bedingungen vorstellen, die auf einige, aber nicht auf alle Patient:innen zutreffen“, erklärt Vreeken. Diese Subgruppen, betont er, können selbst dann noch identifiziert werden, wenn sich das Training eines erklärbaren Glasbox-Modells im Swarm Learning als unmöglich herausstellen sollte. „Das Schöne an unserer Glasbox-KI ist, dass wir sie zusätzlich zu einer Black-Box-KI nutzen können. Wir können nämlich fragen: ‚Für welche Menschen trifft die Black-Box-KI besonders zuverlässige Vorhersagen?‘ Selbst wenn wir also letzten Endes eine Black-Box-KI verwenden, weil sie akkurater ist als jedes transparente Modell, das wir entwickeln können, sind wir immer noch in der Lage, die Subgruppen zu bestimmen, für die wir sie befragen sollten oder nicht.“

Liberate AI: Fachexpertise und Maschinelles Lernen verbinden

Letztendlich möchten die CISPA-Forschenden ein transparentes KI-System entwickeln, das kausale Garantien für seine Vorhersagen geben kann. Wenn es beispielsweise vorhersagen sollte, dass Bluthochdruck die Wirksamkeit der Behandlung verringert, soll es auch die Gründe dafür nennen können. „Das ist sehr schwierig umzusetzen", erklärt Vreeken, „denn man braucht eine randomisierte Kontrollstudie, um festzustellen, ob Bluthochdruck tatsächlich der alleinige Faktor ist oder nur ein Störfaktor – also etwas, das relevant erscheint, es aber nicht ist. Die ultimative KI, die wir entwickeln möchten, ist also eine Glasbox-KI, die sagen kann: ‚Basierend auf allen verfügbaren Schlaganfalldaten gibt es einen klaren Unterschied zwischen ansonsten vergleichbaren Patient:innen, der sich allein durch den Blutdruck erklären lässt.'"

Selbst wenn sich die dreifache Herausforderung – Erklärbarkeit, Identifizierung von Subgruppen und kausale Garantien – am Ende als zu ambitioniert herausstellen sollte, ist Vreeken überzeugt, dass Liberate AI einen bedeutenden Beitrag zur Anwendbarkeit von KI in der Medizin leisten wird. Besonders die Interdisziplinarität des Projektteams eröffnet neue Möglichkeiten für die Behandlung akuter Schlaganfälle, wie er hervorhebt: „Wenn man Fachleute fragt, was sie wollen, dann wollen sie eine bessere Maschine X. Vielleicht brauchen sie aber etwas ganz anderes, von dem sie gar nicht wissen, dass es möglich ist. Das gegenteilige Problem ist, dass Informatiker:innen oft neue Maschinen entwickeln, von denen die Fachleute vielleicht sagen: ‚Das löst ein Problem, das wir gar nicht haben.' Ich bin sehr froh, dass wir in diesem Projekt eine hervorragende Konstellation von Menschen mit Informatikexpertise, Menschen mit rein medizinischer Expertise und Menschen dazwischen haben. In Liberate AI werden wir keine Maschine entwickeln, auf die niemand wartet – sondern eine Maschine, von der die Menschen überhaupt nicht wissen, dass sie sie brauchen."

Förderhinweise auf Seite 78

© Janine Paulus

Passwörter bleiben „das Go-to-Authentifizierungs-Tool“ im Alltag, sagt CISPA-Forscher Alexander Ponticello. Gleichzeitig sind Passwörter oft eine Sicherheitsschwachstelle: zu kurz, zu einfach und zu oft wiederverwendet. Für blinde und sehbehinderte Menschen kommt eine zusätzliche Hürde hinzu: Systeme müssen sinnvoll zusammenarbeiten, damit Authentifizierungsprozesse problemlos ablaufen. Eine neue qualitative Studie mit 33 US-Teilnehmenden zeigt, wie diese Gruppe Passwörter verwaltet und wo Nachholbedarf besteht. Sein Paper „How Blind and Low-Vision Users Manage Their Passwords“ hat Alexander Ponticello auf der Conference on Computer and Communications Security (CCS) 2025 vorgestellt.

So verwalten blinde und sehbehinderte Menschen ihre Passwörter

Alexander Ponticello

Passwörter sind immer noch das Standardwerkzeug für Sicherheit im Netz – aber auch eine ständige Quelle von Problemen. Viele Menschen haben heute Hunderte von Accounts, für die sie mehr oder minder komplexe Passwörter verwalten müssen. Passwortmanager können dabei unterstützen: Sie erstellen sichere Passwörter, speichern sie ab und füllen Anmeldedaten automatisch aus – Problem gelöst, oder? Leider nein, denn Passwortmanager werden längst nicht von allen Menschen konsequent genutzt. Frühere Studien zeigen, dass die Gründe vor allem die Scheu vor komplizierter Einrichtung, mangelndes Vertrauen und fehlendes Wissen über existierende Tools sind. Bei älteren Nutzergruppen kommt zudem eine generelle Zurückhaltung gegenüber digitalen Tools hinzu. Alexander Ponticellos neue Studie erweitert die Forschung zur Verwaltung von Passwörtern und zur Nutzung von Passwortmanagern nun auf eine bisher kaum betrachtete Gruppe: blinde und sehbehinderte Nutzer:innen.

Breite Nutzung von Passwort-managern in der Community

Passwortmanager sind für blinde und sehbehinderte Menschen ein wichtiges Tool zur Verwaltung ihrer Anmeldeinformationen. „Tatsächlich nutzten alle 33 Befragte unserer Studie Passwortmanager – teils bewusst, teils unbewusst, einfach weil ihnen der Browser oder ihr Gerät die Verwaltung anbietet." Darunter waren sowohl sogenannte Drittanbieter-Programme wie LastPass oder 1Password, als auch browserintegrierte Passwortmanager wie sie etwa in Google Chrome eingebaut sind sowie systemintegrierte Passwortmanager wie beispielsweise Apple Passwords. „Wer sich gezielt einen Passwortmana-ger angeschafft hat, verließ sich dabei meist auf Empfehlungen von Bekannten oder Ratschläge in entsprechenden Foren. Dabei spielte die Barrierefreiheit eine mindestens ebenso große Rolle wie die Sicherheit der Systeme", erklärt Ponticello.

Echte Barriere-freiheit nur, wenn Systeme zusammenarbeiten

Blinde und sehbehinderte Nutzer:innen sind je nach Grad der Einschränkung im Alltag vor allem auf Bildschirmleser angewiesen, um ihre Geräte nutzen zu können. „Unsere erste Intuition war, dass es ein großes Problem sein muss, dass die Bildschirmleser Passwörter in der Öffentlichkeit laut vorlesen. Das erwies sich aber als weniger gravierend, denn fast alle Studienteilnehmer:innen sagten uns, dass sie Kopfhörer nutzen", sagt der Forscher. Zudem

laufe die Sprachausgabe meist so schnell, dass Außenstehende kaum etwas verstehen könnten. „Damit blinde und sehbehinderte Menschen Passwortmanager reibungslos nutzen können, müssen allerdings Bildschirmleser, Passwortmanager, Apps und Websites entsprechend zusammenarbeiten. „Wenn eine dieser Parteien versagt, bricht das ganze System zusammen“, so Ponticello. Und leider gibt es immer noch Programme, bei denen Barrierefreiheit eher ein Nachgedanke zu sein scheint. Denn spätestens wenn Updates eingespielt werden müssen, haben einige Nutzer:innen die Erfahrung gemacht, dass Programme nicht mehr richtig funktionieren. Die Folge: Nutzer:innen haben das Gefühl, sich nicht richtig auf die Systeme verlassen zu können.

Sicherheit versus Alltag: Kompromisse sind üblich

Viele der befragten Nutzer:innen kombinieren daher Passwortmanager mit Backup-Strategien. Manche führen sogar Passwortlisten in Braille – sicher aufbewahrt, aber eben analog. „Das ist nicht per se unsicher“, erklärt der Forscher. „Aber man muss sich bewusst sein, wer Zugriff auf diese Liste haben könnte.“ Andere Studienteilnehmer:innen gaben an, absichtlich einfachere Passwörter zu erstellen, um sie notfalls auch ohne Tool eingeben zu können. „Das widerspricht dem Sicherheitsgedanken“, sagt er, „es zeigt aber vor allem auch: Systeme müssen zuverlässiger werden.“

Was (noch) fehlt – und wie es besser geht

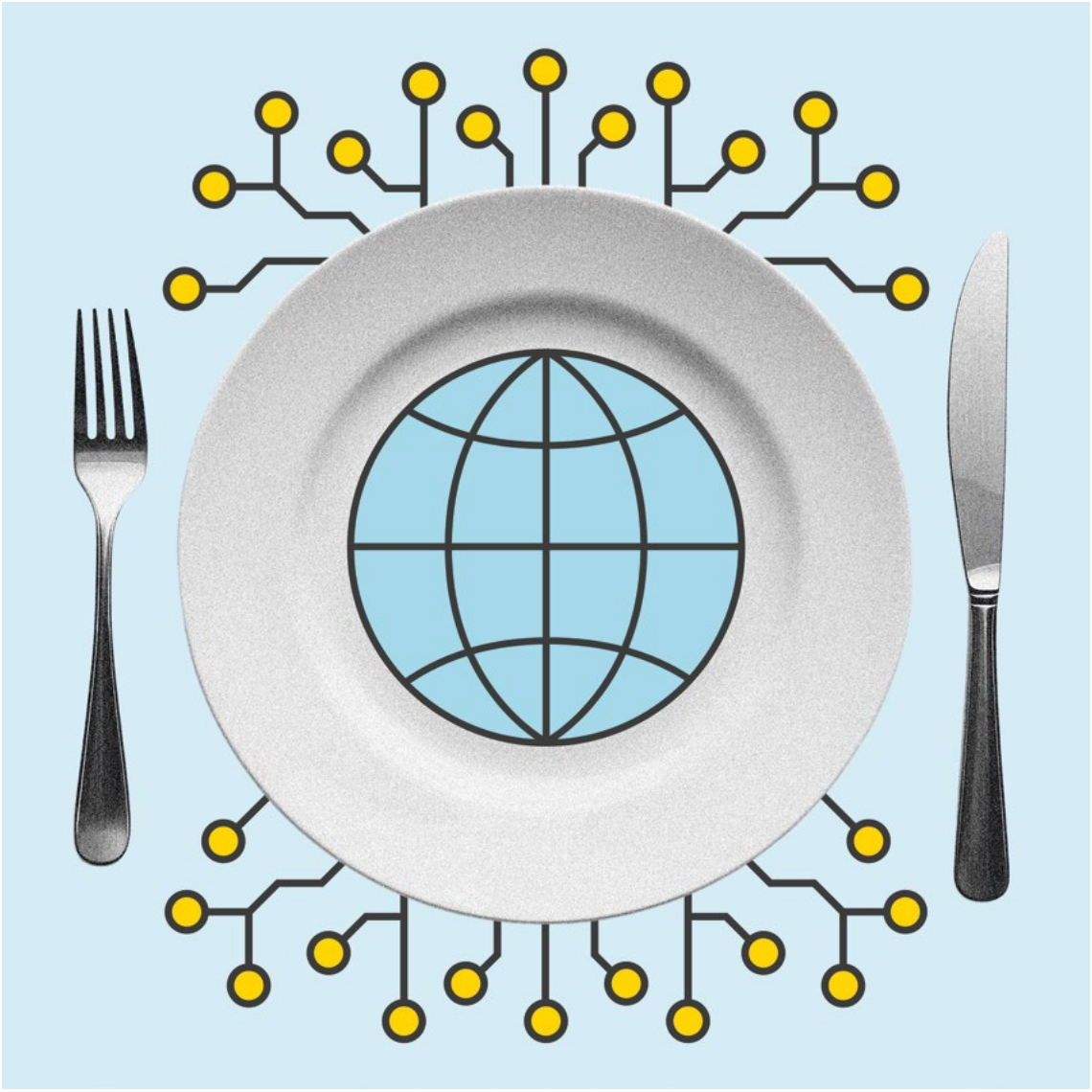
Ein Problem ist laut Ponticello die Passwortgenerierung von Passwortmanagern: Zufallspasswörter mit Sonderzeichen sind für Blinde auf Tastaturen oft schwer zu finden. Eine bessere Alternative wären Passphrasen, bei denen ganze Wörter aneinandergereiht werden. „Leider lesen Bildschirmleser diese Passwörter dann aber Buchstabe für Buchstabe vor, statt die Wörter zu erkennen. Hier ist die Integration nicht zu Ende gedacht“, so der Forscher. Auch App-Stores könnten helfen, indem sie die Barrierefreiheit von Tools klar kennzeichnen und spezielle Review-Kategorien für Betroffene einführen, bei denen sich blinde und sehbehinderte Nutzer:innen direkt informieren können. „Aber das Wichtigste ist: Wir brauchen Barrierefreiheit by Design – korrekte Labels für Buttons, eine sinnvolle Fokus-Reihenfolge und konsistente Bildschirmleser-Flows.“

Ausblick

Eine ähnliche Studie mit deutschen Nutzer:innen durchzuführen, könnte für Ponticello ein nächster Schritt sein. Bislang war die Gesetzgebung in den USA strenger als in der EU. Durch Gesetze wie den „Americans with Disabilites Act“ gelten dort längst strenge Barrierefreiheit-Standards auf Websites und für digitale Dienste. Die EU zieht mit dem European Accessibility Act (EAA) nach. In Deutschland ist daraus das Barrierefreiheitsstärkungsgesetz entstanden, das seit 28. Juni 2025 verbindlich angewendet werden muss. „Ich bin gespannt, welche Effekte das in Zukunft haben wird“, so Ponticello.

Seine Studie zeigt: Barrierefreiheit ist kein Luxus, sondern Grundvoraussetzung für digitale Sicherheit. Viele Hürden – von fehlender Kennzeichnung bis zu brüchigen Integrationen – sind lösbar, wenn Plattformen, Entwickler:innen und Gesetzgeber sie ernst nehmen. „Man muss die Systeme anpassen, nicht die Menschen“, sagt der Forscher. „Nur dann sind Passwörter wirklich für alle sicher nutzbar.“

Ponticello, Alexander; Sharevski, Filipo; Anell, Simon; Krombholz, Katharina (2025): How Blind and Low-Vision Users Manage Their Passwords. In: CCS 2025, 13–17 Oct, 2025, Taipei, Taiwan, Conference: ACM Conference on Computer and Communications Security (CCS)



© Chiara Schwarz

CISPA-Forscherin Tejúmádé Àfọ̀njá hat an einer neuen internationalen Studie mitgearbeitet, die ausgehend vom Thema Essen erhebliche kulturbezogene blinde Flecken in heutigen KI-Systemen aufdeckt. In der Studie wird auch ein neuer partizipativer Forschungsansatz vorgestellt, um inklusivere Datensätze zu erstellen und Verzerrungen in KI-Modellen zu bewerten. Das Paper „The World Wide Recipe: A Community-Centred Framework for Fine-Grained Data Collection and Regional Bias Operationalisation“ wurde im Juni 2025 auf der ACM Conference on Fairness, Accountability, and Transparency (FAccT ’25) in Athen vorgestellt und erhielt dort eine Best Paper Honorable Mention.

World Wide Dishes: mit Essen die kulturellen Blindspots von KI aufdecken

Ṭejúmádé Àfọ̀njá

„Essen stellt einen wichtigen Zugang zur Kultur dar“, erklärt CISPA-Forscherin Ṭejúmádé Àfọ̀njá, Doktorandin im Team von CISPA-Faculty Dr. Mario Fritz. „Wir wollten untersuchen, wie generative KI die Esskulturen der Menschen in generierten Bildern darstellt.“ Dahinter stand der Wunsch mögliche kulturelle Verzerrungen von KI-Modellen zu untersuchen. „Die leitende Projektkoordinatorin unseres Papers, Siobhan Mackenzie Hall, hatte in vorherigen Studien festgestellt, dass viele Modelle in der einen oder anderen Form voreingenommen sind“, fährt Àfọ̀njá fort. „Bei der Frage, durch welche Linse wir dieses Problem betrachten könnten, erwies sich das Thema Essen als guter Zugang, da es für Menschen auf der ganzen Welt von Bedeutung ist.“ Konkret hat das Team untersucht, wie bestimmte Gerichte in KI-generierten Bildern dargestellt werden. Dafür wurde in einem ersten Schritt ein neuer Referenzdatensatz entwickelt und mit diesem in einem zweiten Schritt bestehende Modelle getestet.

Ein neuer Datensatz mit Gerichten aus der ganzen Welt

Das Autor:innenteam entschied sich dabei für einen Community-orientierten Forschungsansatz und nannte diesen World Wide Recipe. Menschen aus der ganzen Welt wurden eingeladen, ihr Wissen zur Verfügung zu stellen. „Wir wollten den Menschen Mitbestimmung darüber geben, wie ihre Kulturen in KI-Systemen repräsentiert werden“ so Àfọ̀njá. Als erste Fallstudie entstand der Datensatz World Wide Dishes (WWD): eine Sammlung von 765 Gerichten aus 106 Ländern, beschrieben in 131 lokalen Sprachen. Die einzelnen Gerichte wurden von Menschen aus den jeweiligen Communities beigesteuert, die den kulturellen, sprachlichen und kulinarischen Kontext erklärten und Fotos lieferten. „Wir haben WWD mit bestehenden, aus dem Internet gesammelten Datensätzen verglichen“, erklärt Àfọ̀njá. „Mehr als die Hälfte der Gerichte im Datensatz tauchen dort nicht auf, was seinen einzigartigen Charakter ausmacht.“ Der Datensatz und der gesamte Code wurden unter einer offenen Lizenz veröffentlicht, um Transparenz und Zusammenarbeit zu fördern.

Fehlrepräsentationen in bestehenden Modellen

In einem zweiten Schritt nutzten Àfọ̀njá und ihre Kolleg:innen WWD, um die Bilder der darin enthaltenen Gerichte mit KI-generierten Bildern dieser Gerichte zu vergleichen. Die vergleichende Analyse wurde wiederum von Mitgliedern der Communities durchgeführt. „Wir haben festgestellt, dass viele der getesteten Modelle stereotype Ergebnisse liefern. Als wir beispielsweise ein Modell baten, ein Bild des nigerianischen Gerichts Amala zu generieren, waren die Ergebnisse oft unappetitlich oder schlicht falsch“, erklärt Àfọ̀njá. „Wenn wir dagegen ein Gericht wie einen Hotdog aus den USA anfragten, war das Ergebnis deutlich realistischer.“ Das galt für alle getesteten Modelle: DALL·E 2, DALL·E 3 und Stable Diffusion. „Die Bildqualität war im Allgemeinen schlecht, und die kulturelle Darstellung oft verfälscht“, fährt sie fort. „Der Grund ist, dass viele Modelle mit Internetdaten trainiert werden. Wenn Gerichte aus bestimmten Regionen online kaum vorkommen, werden diese Regionen in der KI einfach übersehen.“

»Es reicht nicht aus, ein Modell im Silicon Valley oder in Deutschland zu entwickeln und zu erwarten, dass es überall funktioniert.«

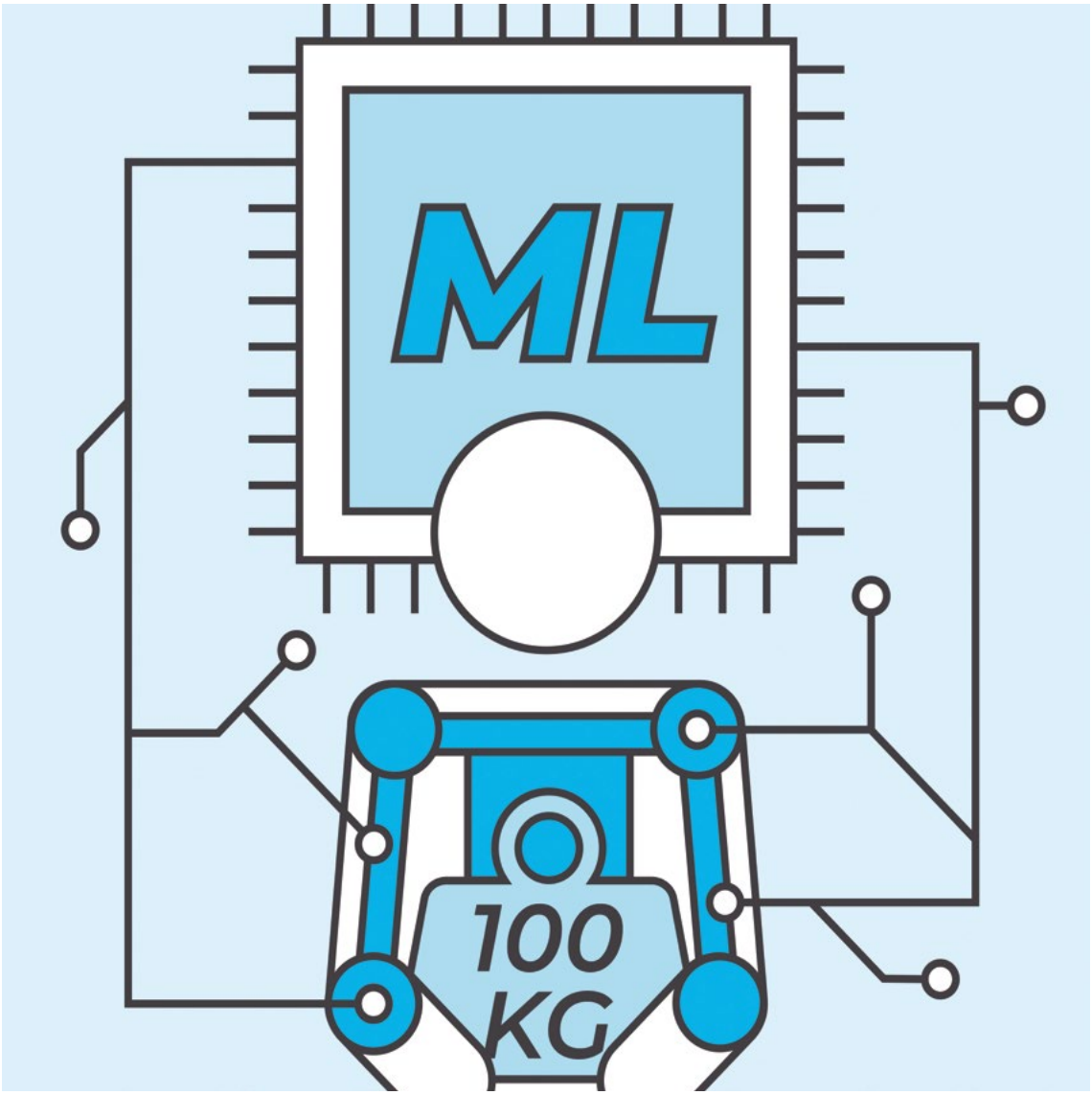
Ein globales Werkzeug braucht globale Beteiligung

Àfọ̀njá und ihre Kolleg:innen ziehen aus dieser Erkenntnis den Schluss, dass die hinter den Modellen stehenden Unternehmen mehr in Long-Tail-Training und die Datensammlung für große Sprachmodelle investieren müssen. „Wir argumentieren, dass diese Unternehmen alle Regionen weltweit gleichberechtigt berücksichtigen müssen, wenn sie Modelle bauen wollen, die dem Anspruch gerecht werden, Kultur auf der ganzen Welt zu repräsentieren“, sagt Àfọ̀njá. „Es reicht nicht aus, ein Modell im Silicon Valley oder in Deutschland zu entwickeln und zu erwarten, dass es überall funktioniert. Entscheidend ist, mehr Daten zu sammeln. Aber das muss in Zusammenarbeit mit den Communities geschehen, nicht durch bloßes Datensammeln über sie hinweg.“ Ein wichtiges Stichwort ist in diesem Zusammenhang die Datenhoheit. „Wenn Daten aus Communities erhoben werden, stellt sich immer die Frage, wem diese Daten gehören: der Community oder der Organisation, die die Datensammlung finanziert hat“, so die CISPA-Forscherin.

Datensammlung und der Kampf gegen kulturelle Verzerrungen

Àfọ̀njá würde das Projekt World Wide Dishes gerne ausbauen. Doch das ist teuer. Bisher wurde WWD komplett ehrenamtlich getragen. „Keiner der Mitwirkenden wurde bezahlt“, sagt sie. „Mit ausreichender Förderung könnten wir sie jedoch bezahlen, damit sie noch mehr lokale Daten sammeln und zum Beispiel Familien nach Rezepten fragen, die online bisher nicht zu finden sind. Solche Daten sind unschätzbar wertvoll, aber aufwendig zu beschaffen.“ Weil die Methode der Datensammlung für das Projekt so wichtig war, entstand daraus eine weitere Publikation: „Wir haben ein Paper mit dem Titel ‚The Human Labour of Data Work‘ veröffentlicht, das dokumentiert, wie wir den Datensatz gesammelt haben und welche Herausforderungen es dabei gab“, so Àfọ̀njá. „Es konzentriert sich auf den menschlichen Aufwand, kulturelles Vertrauen und die Lehren, die andere aus unserer Arbeit ziehen können, wenn sie ähnliche Datensätze aufbauen wollen.“ Wer Àfọ̀njá zuhört, merkt schnell, wie sehr ihr dieses Thema am Herzen liegt und dass sie weiterhin dafür kämpfen wird, dass KI-Modelle ihre kulturelle Voreingenommenheit verlieren und dabei Community-bezogene Ansätze verfolgt werden.

Magomere, Jabez; Ishida, Shu; Afonja, Tejumade; Salama, Aya; Kochin, Daniel; Foutse, Yuehgoh; Hamzaoui, Imane; Sefala, Raesetje; Alaagib, Aisha; Dalal, Samantha; Marchegiani, Beatrice; Semenova, Elizaveta; Crais, Lauren; Mackenzie Hall, Siobhan (2025): The World Wide Recipe: A Community-Centred Framework for Fine-Grained Data Collection and Regional Bias Operationalisation. In: FAccT '25, 23–26 June, 2025, Athens, Greece, Conference: ACM Conference on Fairness, Accountability, and Transparency



© Janine Paulus

Exoskelette gelten als Technologie der Zukunft, dabei entlasten sie heute schon Arbeitskräfte in Logistikbetrieben und Produktionsstätten. Den komplexen Anforderungen des Alltags werden die tragbaren Assistenzsysteme bislang jedoch nicht immer gerecht. CISPA-Forscher Julian Rodemann arbeitet gemeinsam mit Kolleg:innen der LMU München und des Biodesign Lab der Harvard University daran, das zu ändern: mit einem Machine-Learning-Verfahren, das nicht nur optimale Unterstützungseinstellungen findet, sondern auch erklärt, warum es eine bestimmte Konfiguration empfiehlt. Sein Paper „Explaining Bayesian Optimization by Shapley Values Facilitates Human-AI Collaboration For Exosuit Personalization“ hat Rodemann auf der European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) vorgestellt.

Erklärbare KI macht Exoskelette verständlich – und damit alltagstauglich

Julian Rodemann

„Exoskelette sind schon im Einsatz – aber bisher meist in sehr klar definierten Arbeitsumgebungen. Typischerweise unterstützen sie monotone Bewegungen in Logistik oder Produktion und sind genau dafür voreingestellt“, sagt Julian Rodemann. Soft-Exosuits – die leichtere Variante – funktionieren ähnlich: Sie sind meist taskspezifisch konfiguriert, etwa für wiederholtes Heben oder Sortieren. Für wechselnde Aktivitäten eignen sie sich hingegen kaum. Genau hier liegt laut Philipp Arens, Doktorand an der John A. Paulson School in Harvard, die eigentliche Herausforderung: „Ein Schlüsselproblem liegt darin, Exosuits für sehr unterschiedliche Körper und Bewegungsmuster nutzbar zu machen. Designseitig kann man Geräte leichter und weniger störend gestalten – aber die eigentliche Herausforderung liegt darin, wann und wie viel Unterstützung eine Person braucht. Das variiert individuell. Deshalb wird nutzerbasiertes Feedback in der Assistenz selbst so wichtig“, so Arens.

Warum die optimale Einstellung so schwer zu finden ist

Um herauszufinden, welche Einstellungen für welche Person optimal sind, setzen die Forschenden auf maschinelle Unterstützung. „Reale Testreihen dauern oft mehrere Stunden“, erklärt Rodemann. „Die Probanden führen verschiedene Bewegungen aus, während meine Kolleg:innen vom Biodesign Lab in Harvard kontinuierlich physiologische Daten erfassen. Dadurch ist die Zahl der realistisch testbaren Kombinationen stark begrenzt.“ Das Team nutzt deshalb bayesianische Optimierung – ein Verfahren des maschinellen Lernens, das durch gezieltes Probieren und schrittweises Reduzieren von Unsicherheiten effizient zum Optimum findet.

Warum die KI nicht nur das Beste sucht – sondern auch testet, was sie noch nicht weiß

Der Algorithmus sucht dabei nicht einfach in jeder Runde die vermeintlich beste Einstellung. Er muss auch unbekannte Bereiche des Parameterraums erkunden – selbst wenn eine Konfiguration kurzfristig für den Nutzenden weniger komfortabel ist. „Wir unterscheiden zwischen Exploitation, also dem Nutzen des vorhandenen Wissens, und Exploration, dem gezielten Ausprobieren zur Schließung von Wissenslücken“, sagt Rodemann.

Diese Balance ist entscheidend für adaptive Exosuits. „Viele Optimierungsverfahren sind für die Nutzenden Black Boxes. Sie wissen nicht, welche Einstellung vorgeschlagen wird – und erst recht nicht warum. Wenn wir ihnen klar und granular erklären können, was das System als Nächstes tun will und aus welchem Grund, steigt nicht nur das Vertrauen. Die Person kann auch selbst beurteilen, welche Bereiche des Parameterraums nicht sinnvoll sind – und wir vermeiden unnötige Tests“, sagt Arens.

»Human-in-the-loop-Ansätze sind sehr leistungsfähig, aber sie sind aufwändig und belasten Teilnehmende. Ein vielversprechender nächster Schritt ist, nutzergruppenspezifische Startpunkte zu entwickeln – also ‚warm starts‘, die auf typischen Profilen basieren.«

ShapleyBO: ein Verfahren, das zeigt, warum die KI entscheidet, was sie entscheidet

Um diese Transparenz herzustellen, haben Rodemann und seine Kolleg:innen ShapleyBO entwickelt – eine Methode, die Optimierungsentscheidungen nachvollziehbar macht. „Bisher sah der Mensch nur das Ergebnis, etwa dass jetzt ‚Unterstützungsstärke 7 beim Bücken und Stärke 3 beim Heben' eingestellt ist. Mit ShapleyBO zeigen wir, welche der Parameter zur Empfehlung geführt haben. Wir erklären zusätzlich, ob der Vorschlag der Optimierung oder der gezielten Erkundung neuer Einstellungen dient. So können Nutzende einschätzen, ob der Vorschlag in der Situation wirklich sinnvoll ist und bei Bedarf eingreifen", erklärt Rodemann. Damit wird die zuvor abstrakte Balance aus Exploration und Exploitation für Nutzer:innen sichtbar.

Weitere Studien nötig, damit Mensch und KI gezielt Wissen teilen können

„Der Algorithmus kennt Muster aus vielen Nutzerdaten, der Mensch kennt seine aktuelle Situation am besten. Die Frage ist, wie sich beides effizient kombinieren lässt", sagt Rodemann. Das Verfahren befindet sich noch in der Entwicklung und wurde bislang anhand von Simulationsdaten eines realen Soft-Exosuits getestet.

Als Nächstes sollen Nutzerstudien untersuchen, wie Menschen mit erklärbaren Optimierungsvorschlägen interagieren. „Human-in-the-loop-Ansätze sind sehr leistungsfähig, aber sie sind aufwändig und belasten Teilnehmende. Ein vielversprechender nächster Schritt ist, nutzergruppenspezifische Startpunkte zu entwickeln – also ‚warm starts', die auf typischen Profilen basieren. Dadurch könnten wir Optimierungen beschleunigen oder in manchen Fällen sogar ganz vermeiden", so Philipp Arens. Rodemann sieht darin einen grundlegenden Fortschritt: „Unser Ansatz hat das Potenzial, nicht nur die Personalisierung von Exosuits zu verbessern, sondern auch das Vertrauen in KI-basierte Unterstützungssysteme insgesamt zu stärken."

Förderhinweise auf Seite 78

Rodemann, Julian; Croppi, Federico; Arens, Philipp; Sale, Yusuf; Herbinger, Julia; Bischl, Bernd; Hüllermeier, Eyke; Augustin, Thomas; Walsh, Conor J; Casalicchio, Giuseppe (2025): Explaining Bayesian Optimization by Shapley Values Facilitates Human-AI Collaboration for Exosuit Personalization. In: ECML-PKDD 2025, 15–19 Sept, 2025, Porto, Portugal, Conference: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

FÖRDERHINWEISE

LLM-basierter Scanner für Webanwendungen erkennt Tasks und Workflows ***14***

Die in diesem Artikel beschriebene Forschung wurde von der Deutschen Forschungsgemeinschaft (DFG, German Research Foundation) im Rahmen des Projekts „Semantische Modelle und Agenten für die Verwundbarkeitsprüfung von Web-Anwendungen“ mit der Projektnummer 452850842 gefördert.

Open-Source-Fuzzer mit evolutionärem Algorithmus erzeugt individualisierte Inputs ***38***

Die in diesem Artikel beschriebene Forschung wurde von der Europäischen Union (ERC S3, 101093186) gefördert. Die geäußerten Ansichten und Meinungen liegen jedoch ausschließlich bei den Autorinnen und Autoren und spiegeln nicht zwangsläufig die Auffassungen der Europäischen Union oder des Europäischen Forschungsrats wider. Weder die Europäische Union noch die Förderorganisation können hierfür verantwortlich gemacht werden.

Neues Verfahren erkennt Nutzung urheberrechtlich geschützter Bilder im KI-Training ***46***

Die in diesem Artikel beschriebene Forschung wurde von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Weave-Programms unter dem Projekt „Protecting Creativity: On the Way to Safe Generative Models“ (Projekt-Nr. 545047250) unterstützt. Die Forschung wurde zudem durch die polnische National Science Centre (NCN) unter den Fördernummern 2023/51/I/ST6/02854 und 2020/39/O/ST6/01478 gefördert sowie von der Technischen Universität Warschau im Rahmen des Excellence Initiative Research University (IDUB)-Programms unterstützt.

Wie agil ist deine Krypto? Interviewstudie zu kryptographischen Updateprozessen ***54***

Die in diesem Artikel beschriebene Forschung wurde durch die VolkswagenStiftung Niedersächsisches Vorab (ZN3695) gefördert. Alle in diesem Material dargestellten Ergebnisse und Meinungen liegen in der Verantwortung der Autorinnen und Autoren und spiegeln nicht zwangsläufig die Auffassungen der Förderorganisationen wider.

KI beschleunigt Medikamentenentwicklung durch automatische Analyse von Zebrafisch-Embryonen ***58***

Die in diesem Artikel beschriebeneForschung wurde teilweise durch Image-Tox (ZT-I-PF-4-037) gefördert, unterstützt durch den Impuls- und Vernetzungsfonds der Helmholtz-Gemeinschaft. Die geäußerten Ansichten und Meinungen liegen ausschließlich bei den Autorinnen und Autoren und spiegeln nicht notwendigerweise die der Förderorganisationen wider, die hierfür keine Verantwortung übernehmen können. Sie wurde außerdem teilweise durch ELSA – European Lighthouse on Secure and Safe AI (Fördervereinbarung Nr. 101070617) finanziert. Das dieser Veröffentlichung zugrunde liegende Projekt wurde vom Bundesministerium für Bildung und Forschung gefördert (Förderkennzeichen 16KIS2012).

Von Black Box zu Glasbox: erklärbare KI in der Schlaganfallbehandlung ***62***

Die in diesem Artikel beschriebene Forschung wird von der Helmholtz-Gemeinschaft aus dem Helmholtz Impuls- und Vernetzungsfondgefördert gefördert.

Erklärbare KI macht Exoskelette verständlich – und damit alltagstauglich ***74***

Die in diesem Artikel beschriebene Forschung wurde vom Statistischen Bundesamt im Rahmen des Kooperationsprojekts „Maschinelles Lernen in der amtlichen Statistik“ unterstützt. JR dankt zudem der Bayerischen Akademie der Wissenschaften (BAdW) über das Bayerische Forschungsinstitut für digitale Transformation (bidt) sowie dem Mentoringprogramm der LMU-Fakultät für Mathematik, Informatik und Statistik. YS wird durch das DAAD-Programm „Konrad Zuse Schools of Excellence in Artificial Intelligence“, gefördert vom Bundesministerium für Bildung und Forschung, unterstützt.

ÜBER DAS CISPA

Das CISPA Helmholtz-Zentrum für Informationssicherheit ist eine Großforschungseinrichtung des Bundes innerhalb der Helmholtz-Gemeinschaft. CISPA-Wissenschaftler:innen erforschen die Informationssicherheit in all ihren Facetten. Sie betreiben modernste Grundlagenforschung sowie innovative anwendungsorientierte Forschung und arbeiten an den drängenden Herausforderungen der Cybersicherheit, der künstlichen Intelligenz und des Datenschutzes. CISPA-Forschungsergebnisse finden Einzug in industrielle Anwendungen und Produkte, die weltweit verfügbar sind. Damit stärkt das CISPA die Konkurrenzfähigkeit Deutschlands und Europas.

Das CISPA bietet ein Forschungsumfeld von Weltrang und stellt einer großen Zahl an Forscher:innen umfangreiche Ressourcen zur Verfügung. Darüber hinaus fördert das CISPA in besonderem Maße auch die grundständige und postgraduale Bildung von Cybersicherheitsstudierenden. Das Zentrum hat sich zum Ziel gesetzt, eine Kaderschmiede für die nächste Generation an Cybersicherheitsexpert:innen und wissenschaftlichen Führungskräften in diesem Bereich zu werden. Das CISPA ist in Saarbrücken und St. Ingbert situiert. Die Lage des Zentrums in direkter Nachbarschaft zu Frankreich und Luxemburg ist ideal für grenzüberschreitende Kollaborationen mit anderen Forschungsinstitutionen.

Aktuell konzentriert sich unsere Forschung auf die folgenden sechs Forschungsbereiche:

Algorithmische Grundlagen und Kryptographie

Vertrauenswürdige Informationsverarbeitung

Verlässliche Sicherheitsgarantien

Erkennung und Vermeidung von Cyberangriffen

Sichere vernetzte und mobile Systeme

Empirische und verhaltensorientierte Sicherheit

IMPRESSUM

CISPA – Helmholtz-Zentrum
für Informationssicherheit gGmbH
Stuhlsatzenhaus 5
66123 Saarbrücken, Deutschland

Herausgeber

Sebastian Klöckner

Verantwortliche Redaktion

Annelies Bourgeois,
Felix Koltermann,
Kevin Meiser,
Eva Michely,
Annabelle Theobald

Redaktion

Stephanie Bremerich,
Alexandra Goweiler,
Janine Paulus,
Chiara Schwarz

Illustration

Alexandra Goweiler,
Janine Paulus,
Chiara Schwarz

Gestaltung

Tobias Ebelshäuser,
David Rohner

Fotografie

Februar 2026

Stand des Impressums

T: +49 681 87083 2867
M: pr@cispa.de
W: https://cispa.de/

Kontakt Corporate Communications

Digitaler Fingerabdruck: CSS eröffnet neue Möglichkeiten zum Nutzer:innen-Tracking

LLM-basierter Scanner für Webanwendungen erkennt Tasks und Workflows

Das unterschätzte Risiko: warum viele WordPress-Websites zu selten aktualisiert werden

Sicherheit läuft nur nebenher mit: Erkenntnisse aus der Videospielbranche

Die Macht der Worte: wie Formulierungen das Zustimmungsverhalten bei App-Berechtigungsanfragen beeinflussen

Ungleiches Internet: Unterschiede zwischen Websites aus Industrie- und Schwellenländern

Cybersicherheitspraktiken von Menschen mit niedrigem sozioökonomischem Status in Pakistan

Open-Source-Fuzzer mit evolutionärem Algorithmus erzeugt individualisierte Inputs

Fuzzing reloaded: mit gezielter Manipulation zu mehr Sicherheit im Netz

Neues Verfahren erkennt Nutzung urheberrechtlich geschützter Bilder im KI-Training

C++-Coroutinen: anfällig für Code-Reuse-Angriffe trotz CFI

Wie agil ist deine Krypto? Interviewstudie zu kryptographischen Updateprozessen

KI beschleunigt Medikamentenentwicklung durch automatische Analyse von Zebrafisch-Embryonen

Von Black Box zu Glasbox: erklärbare KI in der Schlaganfallbehandlung

So verwalten blinde und sehbehinderte Menschen ihre Passwörter

World Wide Dishes: mit Essen die kulturellen Blindspots von KI aufdecken

Erklärbare KI macht Exoskelette verständlich – und damit alltagstauglich

CISPA
HELMHOLTZ-ZENTRUM FÜR INFORMATIONSSICHERHEIT