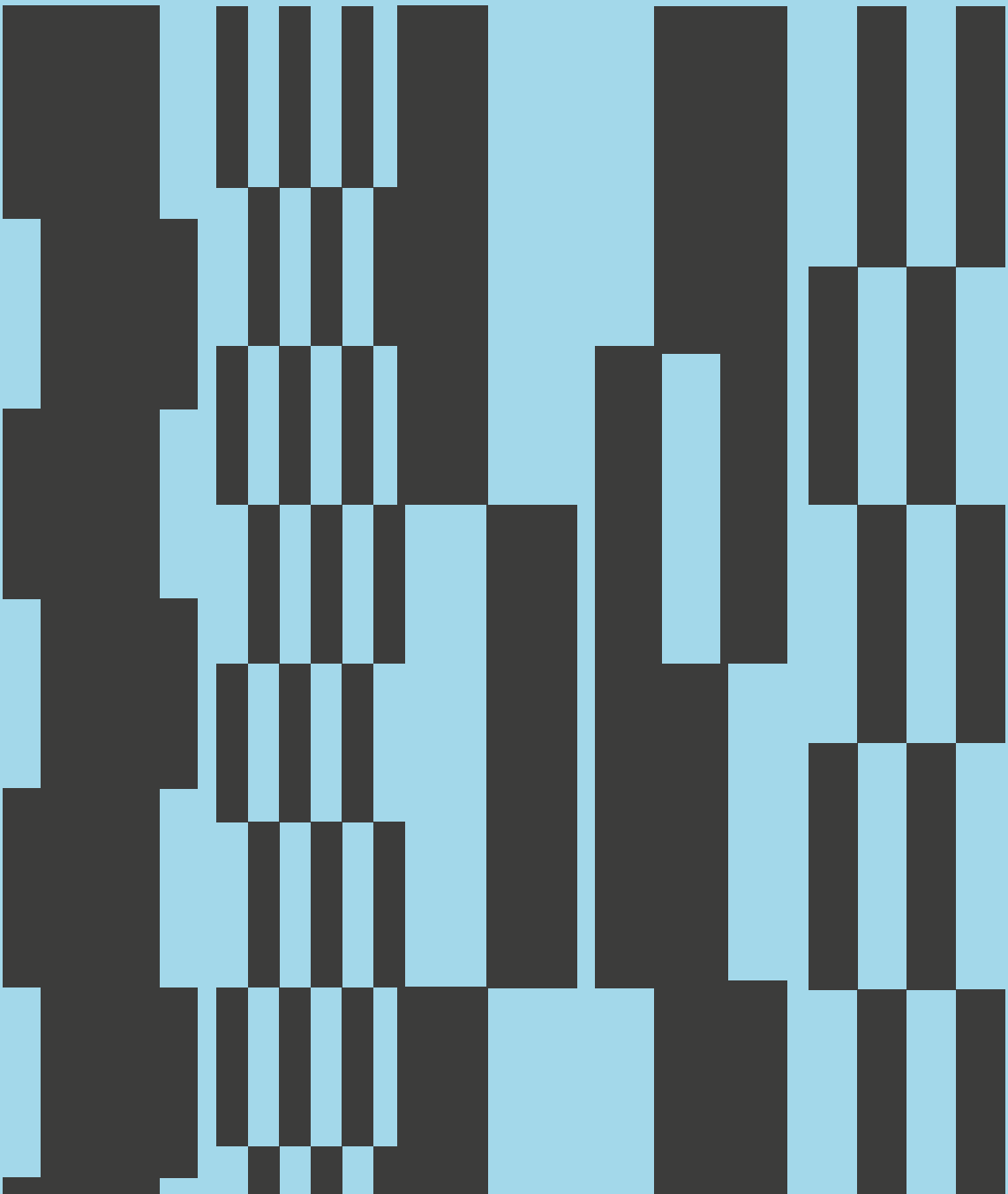
CISPA DISPLAY

EN

EDITION 2026



INTRODUCTION

Making CISPA's research visible and accessible is one of our core missions. We pursue this goal through a wide range of activities, including the dissemination of research news, events hosted by the CISPA Cysec Lab, trade fair appearances, visits by stakeholders and policymakers—such as Federal Chancellor Friedrich Merz in 2025—as well as large-scale research festivals such as CISPA loves IGB. Our formats are diverse, ranging from explanatory texts and conversations to hands-on workshops. Within this broad spectrum, the research yearbook CISPA DISPLAY—now in its third year—has become a permanent pillar of our knowledge transfer activities. It compiles texts published in the previous year that discuss selected scientific papers by our researchers, presented at leading international conferences. In doing so, it offers a concise overview of the research themes pursued at CISPA.

Framework Conditions for Excellent Research

Two fundamental pillars of excellent research are a democratic social order and reliable public funding. Both are firmly established in Germany. In addition, the Helmholtz Association provides the research landscape with a strong network committed to scientific excellence. At the same time, developments in other countries illustrate how fragile these politically supported framework conditions can be: Science and research there often face challenges such as populist debate cultures or political interference in the autonomy and funding structures of universities and research institutions. In addition, growing technological dependencies and evolving digital ecosystems shape how societies envision and design their technological futures.

Digital European Sovereignty

In light of these developments, the importance of digital sovereignty in Germany and Europe has moved to the center of societal and political debate. The issue now extends far beyond economic competitiveness to encompass strategic capacity for action, technological independence, and the question of how democratic societies can shape key technologies autonomously. Information and data security, trustworthy artificial intelligence, resilient infrastructures, and clear regulatory frameworks are among the central building blocks of a European response to global technological dynamics.

For research centers like CISPA, these developments entail a dual responsibility: on the one hand, to actively shape technological innovation, and on the other hand, to contribute to strengthening European values such as

At a time when “digitalization” is omnipresent and neither research on AI and cybersecurity nor knowledge transfer is conceivable without digital tools, a printed research yearbook may seem anachronistic. CISPA DISPLAY, however, represents a deliberate change of medium.

democracy, freedom, and security. CISPA's outstanding scientific evaluation in 2025, the successful growth of the Center, and its international visibility underscore its considerable potential. Research from Germany and Europe not only delivers excellent scientific contributions, but also lays the foundations for a digital future that is independent, secure, and socially responsible.

Artificial Intelligence in Focus

Few technological topics have shaped public debate over the past year as strongly as artificial intelligence. This heightened attention is also reflected in the present edition of CISPA DISPLAY. While AI applications have long since become part of everyday private and professional life and are increasingly used as advisory systems, discussions about their implications for democratic societies are intensifying. CISPA's researchers work every day to ensure that AI is not only trustworthy, but that its potential becomes a societal benefit rather than a liability, particularly in areas where AI and cybersecurity are inextricably linked.

CISPA DISPLAY: Bridging the Analog and Digital Worlds

At a time when "digitalization" is omnipresent and neither research on AI and cybersecurity nor knowledge transfer is conceivable without digital tools, a printed research yearbook may seem anachronistic. CISPA DISPLAY, however, represents a deliberate change of medium. We see it as a productive bridge between the analog and the digital worlds. Handing over a printed publication, consciously turning its pages, and lingering over compelling content and visualizations create moments of focused reflection—moments that are becoming increasingly rare in our accelerated daily lives. All the more, we wish our readers an engaging and inspiring experience with the 2026 edition of CISPA DISPLAY.

CONTENTS

CONTENTS

© Chiara Schwarz

Processor type, IP address, browser in use, installed fonts—by collecting these and other characteristics of browser settings and the underlying operating system, it is possible to create a highly detailed and, in some cases, even unique profile of users. This phenomenon is known as browser fingerprinting. A study by CISPA researcher Leon Trampert and colleagues now suggests that this tracking method can be applied not only when browsing the web but also in emails, through a previously underexplored method: the use of CSS (Cascading Style Sheets), a language for designing websites. The paper "Cascading Spy Sheets: Exploiting the Complexity of Modern CSS for Email and Browser Fingerprinting" will be presented at the Network and Distributed System Security Symposium (NDSS) 2025.

Digital Fingerprint: CSS Opens New Possibilities for User Tracking

Leon Trampert

Even in a large group of website visitors, you are likely uniquely identifiable. Why? Wherever the programming language JavaScript is in use—which is virtually the entire web—specific attributes of your devices and their settings can also be collected. These details are primarily intended to help web developers create better user experiences and functionalities. But as always, knowledge is power, and not everyone wants this knowledge about them to be out in the world. “By now, fingerprinting via JavaScript is pretty well known. People particularly concerned with privacy can protect themselves by blocking JavaScript. This can be done either with plugins or by using the Tor browser. For example, this can be useful for journalists afraid of persecution,” explains Leon Trampert.

Modern CSS Leaks Data

Where one door closes, another opens—and the same seems true for fingerprinting. “Researchers recently discovered that information about users can also leak through the use of CSS,” says Trampert. CSS (short for Cascading Style Sheets) ensures that text, images, and menus are displayed correctly, determines fonts, colors, and the sizes of elements on websites, and allows views to adapt to different screen sizes. “CSS has become increasingly popular and has gained many new functions in recent years. Some of these have already been analyzed by research colleagues for their potential to violate privacy. However, a holistic review was still missing.” So, a few months ago, Trampert decided to systematically study modern CSS functions. “We wanted to see how much we could uncover with it and whether CSS allows tracking outside of the web as well.”

Telltale Fonts

Trampert examined multiple fingerprinting approaches and identified three techniques for creating user fingerprints using CSS. “We initially analyzed 1,176 combinations of browsers and operating systems with various settings and were able to infer the users’ systems in 97.95% of cases. Installed fonts, for instance, can be revealing. They provide clues about the browser, operating system, and installed programs,” Trampert explains. The researchers identified the fonts with a few tricks: “We can’t see this information in plain text, but we can, for example, measure the heights and widths of words by exploiting certain

otherwise useful CSS functions. From this, we can infer not just the font but also the system language," says Trampert.

CSS Enables Tracking Beyond the Web

Even more exciting for him was testing email applications. While JavaScript is often blocked by default in many email clients, the use of CSS remains largely unrestricted. "We tested 21 email clients, including Android, iOS, desktop, and web-based clients. In nine cases, we were able to apply all our techniques successfully and gather information about the users. Eighteen of the 21 email clients were vulnerable to at least one of the techniques," Trampert explains. According to him, this could open up entirely new threat scenarios. "For example, attacks could aim to link web sessions of visitors to their email accounts or identify all email addresses of specific users," Trampert explains.

What Now?

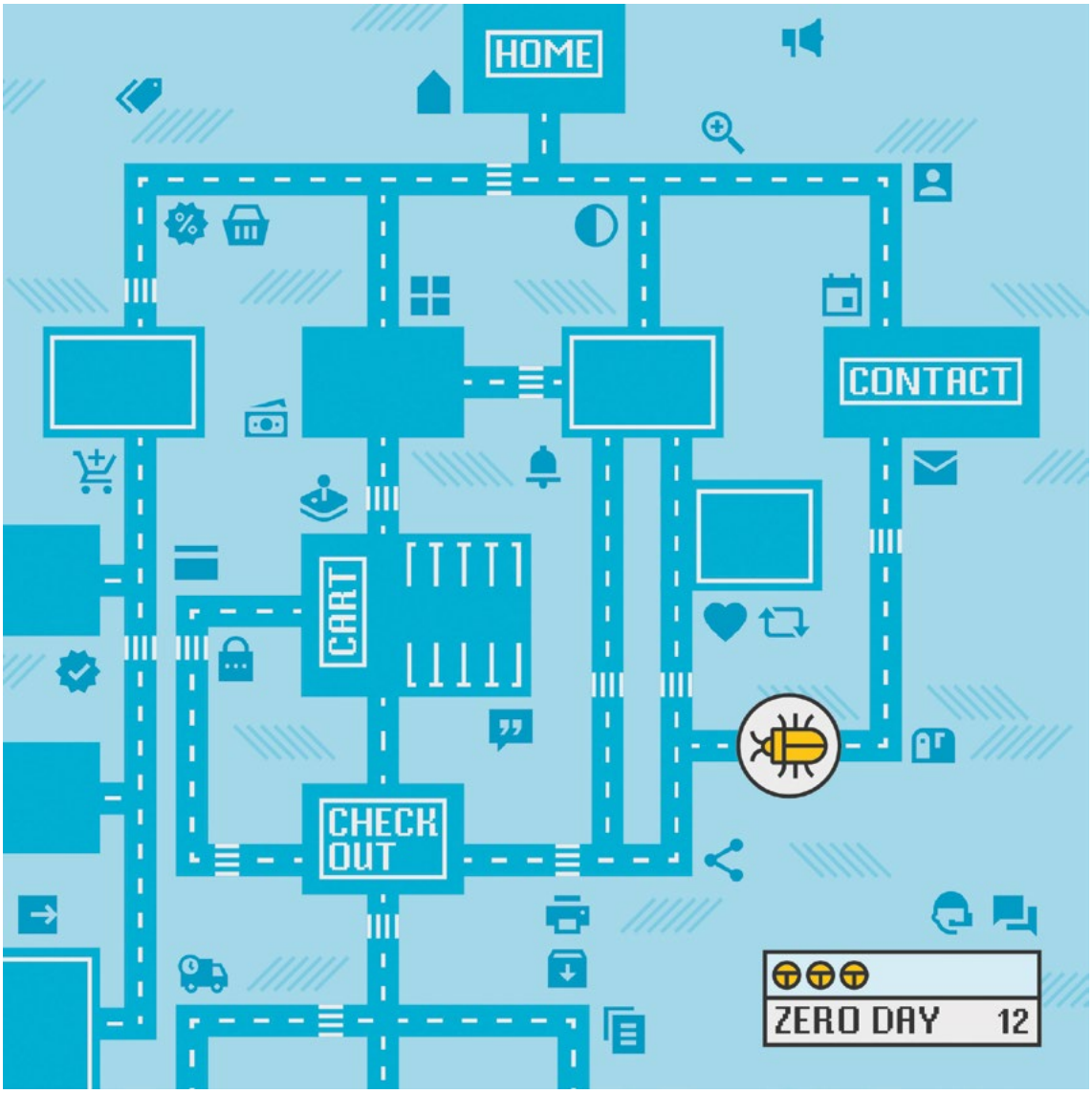
Anyone browsing the web is already being measured involuntarily due to tracking cookies and JavaScript. "Nevertheless, it is important to demonstrate what technical possibilities exist and where new opportunities for abuse arise—as seen here, suddenly even in email programs. Only then can we develop robust defense mechanisms," says Trampert. The PhD student conducts research at CISPA, supervised by CISPA-Faculty Dr. Michael Schwarz and Prof. Dr. Christian Rossow, and intends to continue working on email security issues in the future.

»CSS has become increasingly popular and has gained many new functions in recent years. Some of these have already been analyzed by research colleagues for their potential to violate privacy.«

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)

Researcher: ***Leon Trampert***
Author: ***Annabelle Theobald***

Publication date
January 3, 2025



© Chiara Schwarz

A new automated web application scanner autonomously understands and executes tasks and workflows on web applications. YuraScanner harnesses the world knowledge stored in Large Language Models (LLMs) to navigate through web applications in the same way a human user would. It is capable of working through tasks in a coherent fashion, performing the correct sequence of steps as required by, for example, an online shop. YuraScanner was tested against 20 web applications, unearthing 12 zero-day cross-site scripting (XSS) vulnerabilities. The technique behind YuraScanner as well as the tool itself have been developed by CISPA researcher Aleksei Stafeev and his colleagues. Their paper "YuraScanner: Leveraging LLMs for Task-driven Web App Scanning" will be presented at the Network and Distributed System Security Symposium (NDSS) 2025.

LLM-Based Web Application Scanner Recognizes Tasks and Workflows

Aleksei Stafeev

Automated web application scanners are commonly used to test the security of online applications such as, for example, online shops, learning platforms, or project management tools. Typically, these scanners consist of two parts: the crawler component, which scans the web application for user interfaces, and the attack module, which then proceeds to test the interfaces identified by the crawler. Aleksei Stafeev, who works in the research group of CISPA-Faculty Dr. Giancarlo Pellegrino, highlights the importance of the crawler component for such automated testing to be successful: “One of the main challenges in security testing is determining the scope of the web application and identifying its functionalities and workflows. We know quite well how to detect the security issues, but how do we identify all the entry points?” Stafeev and his CISPA colleagues have developed YuraScanner with the aim of identifying as much of the attack surface as possible.

YuraScanner: Using LLMs to Navigate Web Applications

The main innovation YuraScanner proposes is enhancing the reach and performance of the scanner’s crawler component by harnessing it to an LLM. “LLMs have been trained on the data from the web, which is rich on documentation on how to interact with websites. We tap into this knowledge by combining a crawler and a LLM to guide the exploration of a web application,” Stafeev explains. For the purpose of their study, Stafeev and his colleagues used the OpenAI API to establish the connection between their crawler component and OpenAI model GPT-4. The attack module on the YuraScanner is identical to Black Widow, an established state-of-the-art cross-site scripting scanner. This parallel setup allowed the CISPA researchers to directly compare the performances of the two crawler components. Testing YuraScanner against 20 web applications, they were in fact able to detect 12 previously unknown XSS vulnerabilities, in comparison to only three detected by Black Widow.

Taking Automated Web Application Scanning to a Deeper Level

Guided by an LLM, YuraScanner operates in a task-driven fashion, which allows it to access the deeper layers of the web application being tested. Not only can it identify the tasks that are offered by the web application, it can also carry them out in a deliberate fashion,

performing the sequence of steps required to finish the task at hand. It proceeds vertically, while other, already established scanners, tend to proceed horizontally. Stafeev explains: “Usually, testing tools don’t distinguish between different kinds of buttons, they just click on whatever is available. The main drawback of that is that if there is some very specific multi-step workflow as in, for example, an online shop, where you have to put an item into a cart, proceed to check-out and fill in a form—the chances of a simple web crawler to succeed at that are very slim.” With YuraScanner, Stafeev and his colleagues have shown that LLMs can be used in web security scanning, paving the way for further research in the field.

Funding acknowledgements on page 78.

»LLMs have been trained on the data from the web, which is rich on documentation on how to interact with websites. We tap into this knowledge by combining a crawler and a LLM to guide the exploration of a web application.«

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)

Researcher: Aleksei Stafeev
Author: Eva Michely

Publication date
February 21, 2025

© Chiara Schwarz

Millions of websites worldwide are based on content management systems (CMS) like WordPress, which enable people without programming knowledge to create their own websites and manage digital content. Their widespread use makes them an attractive target for cyberattacks. Regular security updates are the best way to defend against these attacks, although more than half of the systems are not updated on a regular basis. CISPA researcher and psychologist Dr. Maria Hellenthal, along with her colleagues, explored the reasons for this in a qualitative study. Her paper The (Un)usual Suspects– Studying Reasons for Lacking Updates in WordPress was presented at the Network and Distributed System Security Symposium (NDSS) 2025.

The Underestimated Risk: Why Website Owners Often Neglect Security Updates in WordPress

Maria Hellenthal

Cybercriminals exploit security vulnerabilities in websites to steal data, use servers with outdated WordPress installations for spam campaigns or DDoS attacks, or use foreign websites to build fake online shops. To close vulnerabilities and minimize risks, CMS providers regularly offer security updates to their customers. “Unfortunately, many website owners don’t do these updates, or not on a regular basis,” explains Maria Hellenthal. The researchers explored the reasons for this avoidable risk based on outdated WordPress sites and interviews with their owners. The team also spoke with web developers and hosting providers to include their professional perspectives. “We chose WordPress because with over 60% market share worldwide, it is currently the most widespread CMS,” says Hellenthal.

Missing Updates: Causes and Obstacles

The issue of missing security updates is not limited to WordPress. “We see this phenomenon across the entire online ecosystem,” Hellenthal explains. A frequently mentioned reason, which also appears in Hellenthal’s study, is a lack of risk awareness. “Many website owners do not realize that cyberattacks not only harm them but the entire network community. When a site is hacked, not only can its visitors be harmed, but also other website owners and hosting providers—in other words, the entire online community,” says the researcher. Additional obstacles to regular updates include the fear that updates might cause problems, such as compatibility issues with plugins, or that additional costs might arise due to updates.

Two of the reasons identified in Hellenthal’s study were not explicitly mentioned in previous literature on the lack of updating: “An important factor in update behavior seems to be what the website means to its owners. A business owner running an online store, for example, and relying on the site as their main source of income, values it differently than a small business owner who only provides information and relies more on word of mouth. In both cases, updates can be neglected, but website owners who care more about their site are more likely to be persuaded to update by targeted, clear warnings about

potential vulnerabilities," the researcher explains.

According to the study, another problem arises when website management is outsourced to external parties. "Delegating website management to a more experienced person should bring advantages, but it can also bring disadvantages. For example, we have seen in several cases that there can be a diffusion of responsibility when the maintenance tasks are not clearly defined, compensated, and described in the contract. Nobody really feels responsible," Hellenthal says. One interviewee also mentioned feeling overwhelmed when an external person added plugins with which they were not familiar, causing the system to become more complex. "And of course, money plays a role. Many cannot afford an agency to handle these tasks, and tech-savvy friends are only asked when there is no other option," Hellenthal explains.

Security Warnings Are Often Ignored

Security experts have long tried to push website owners to update by sending vulnerability notifications—often with moderate success. "There are studies that examine why notifications are so often ignored and how they should be designed to have a greater impact. Our study, which investigates why system security is neglected in the first place, can help us better understand whom we can still reach with notifications," says Hellenthal. However, according to her, relying solely on security warnings is not enough to sustainably improve the security of WordPress-based websites.

How Can Risks Be Minimized?

Hellenthal also sees responsibility with CMS providers. "They could make security solutions like static site generators, which do not contain unnecessary security-relevant components, much more user-friendly. They should also better educate their customers, in a way that is more understandable for non-experts, about the risks they take when they disable automatic security updates," Hellenthal suggests. She also thinks that public recognition programs for secure websites could be helpful.

Qualitative Research Provides New Insights

The study was based on 19 interviews. How representative is this? "In qualitative research, it's not about generalizability, but rather about identifying behavioral patterns," explains Hellenthal. "On this basis, we can develop theories and test them in further quantitative studies or—as in this case—develop initial improvement strategies, taking into account the reasons for missing updates." The interdisciplinary project grew out of a shared research idea by IT security researcher and CISPA-Faculty Dr. Ben Stock and Dr. Michael Schilling, psychologist and Head of Empirical Research Support at CISPA. The research is based on the master's theses of Lena Gotsche and Sarah Kugel, both of whom are psychologists. Sociologist

Dr. Rafael Mrowczynski contributed his expertise in qualitative research methodology. “We complemented each other perfectly on a methodological and technical level,” concludes Hellenthal, who works in the CISPA Empirical Research Support team, assisting IT security researchers with methodology and study design. “I come from experimental cognitive psychology and have always done more applied research. For a long time, I was a bit of an outsider at my former university. At CISPA, I can contribute my skills to a highly interesting field.”

»Many website owners do not realize that cyberattacks not only harm them but the entire network community.«

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)

Researcher: ***Maria Hellenthal***
Author: ***Annabelle Theobald***

Publication date
February 28, 2025

The video game industry is a constantly changing market worth billions. In a qualitative interview study with industry experts, CISPA researcher Philip Klostermeyer from the team of CISPA-Faculty Prof. Dr. Sascha Fahl investigated the challenges involved in incorporating security considerations into game development. He published the results in the paper "Skipping the Security Side Quests: A Qualitative Study on Security Practices and Challenges in Game Development," which was presented at the Conference on Computer and Communications Security (CCS) 2024.

Security is Just a Side Quest: Insights From the Video Game Industry

Philip Klostermeyer

Video games have long fascinated Philip Klostermeyer, CISPA researcher and PhD student at the CISPA site in Hanover. And this not only from the perspective of the player. “I was required to develop a video game as part of my bachelor’s degree. That was the first time I realized how many different elements even a simple game has,” he says in the interview. “I suddenly understood how the different types of software work together.” In principle, a video game is nothing more than very complex software, explains Klostermeyer: “We have source code and data in the background. On the user interface, we then add a complex graphic design and elements such as audio. This is supplemented by the respective game logic. For online games, there is also the connection to servers that handle game logic, manage common security issues like login and authentication, and enable the display of advertisements. This shows that almost all topics that are important in computer security are relevant to video games.”

Aim of the Study: Gaining an Overview

The complexity of video game development and the significance of security made it an intriguing research topic for Klostermeyer and his colleagues. “We decided to investigate the security topic in video game development with the help of a qualitative interview study,” explains the CISPA researcher. “The method is well suited to gaining an overview of a subject area. After all, there is already a considerable amount of research that thoroughly covers individual topics within the games industry. What has been missing so far is a coherent overview of the entire field.” Another key aspect for Klostermeyer was the desire to apply their findings to the industry: “Our goal was to translate our insights into practical applications for the industry. That’s why we made it a priority to focus our study on the challenges faced by this target group.”

For the study, 20 individuals from 15 countries were interviewed, all of whom hold different positions in the games industry. “We identified the key stakeholders involved in game development and then carefully selected our interviewees,” explains Klostermeyer. “This included game developers, managers, platform publishers, as well as security experts. Our goal was to gain various perspectives on the topic of security. Through the interviews, we aimed to gather first-hand experience regarding

awareness, priorities, knowledge, and practices related to security within the industry."

Security as a Secondary Factor That Depends on Many Aspects

Analyzing the interviews, the CISPA researchers distilled two key areas that are central to the topic of security in video game development. One of these is the unique circumstances within the games industry that affect game development and, consequently, security. "Factors such as the fast-paced nature of the industry, varying security standards, time and budget constraints, as well as a lack of security consulting are worth mentioning here," explains Klostermeyer. On the other hand, the researchers identified five securityrelevant areas in the game development process. "Specifically, these include measures to prevent in-game cheating, the security of so-called assets like source code or graphics, network security, software stability, and the protection of user data," he continues. The importance of each area depends on the type of game in question. "For example, network security is of little relevance for games that are not played online," says the CISPA researcher.

In terms of whether and how studios integrate security into the video game development process, the study identified time, budget, and team size as the most important factors. While external players such as publishers provide securityrelated input, they mainly prioritize security to protect their company's revenue or public image. Whereas large companies recruit their own security specialists, small studios usually lack the budget for this. And even when developers are aware of security issues, this may be considered less of a priority by management than, for example, the playability of a product. "Basically, it can be said that the games industry is very erratic when it comes to security," says Klostermeyer. "The fast pace of the industry prevents developers from taking indepth security measures and developing threat models for video games that are implemented from the start of game development."

Prospects

For Klostermeyer and his colleagues from the Usable Security research team in Hanover, the current interview study was just the starting point for delving deeper into the subject matter. "The great thing about the study is that we were able to identify these five security-relevant areas in the game development process. With this knowledge, we can start developing proposals for guidelines." However, there are already some concrete recommendations for the industry that Klostermeyer has derived from the results of the study. The key point is to integrate the aspect of security into game development as early as possible and to consider it at every level. Guidelines that each development studio should develop itself based

on the respective requirements and adapted to its own products are helpful here. “This is a crucial cross-sectional task that every studio should take seriously,” says Klostermeyer with conviction.

»The fast pace of the industry prevents developers from taking in-depth security measures and developing threat models for video games that are implemented from the start of game development.«

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

One click–and the app has extensive access rights, for example to the camera, microphone or, contacts. What many people don't know: Whether we tap on "Allow" or "Deny" often depends largely on the wording of this request. This is shown by a recent study by CISPA researcher Yusra Elbitar. She presented her paper "The Power of Words: A Comprehensive Analysis of Rationales and Their Effects on Users' Permission Decisions" at the Network and Distributed System Security Symposium (NDSS) 2025. If an app wants to access a sensitive function such as a camera or location, the operating system displays what is known as a "permission request." Developers can supplement this with an additional, explanatory text–a so-called rationale. This rationale is intended to explain to users why the app requires a certain permission.

The Power of Words: How Wording Influences Consent Behavior in App Permission Requests

Yusra Elbitar

"The app needs storage access" or "Please allow access to your storage to attach photos to your posts"–which of these phrases would be more likely to motivate you to give the app permission? App developers can expect a higher approval rate for the second, more specific variant. At least that's what Elbitar's study suggests: "People who understand why they need access to the camera or storage feel better informed and have a greater sense of control. According to our study, both of these factors increase approval of app permissions," the researcher explains. Together with colleagues, she analyzed more than 9,500 frequently used Android apps to find out how permission requests are formulated and designed.

Central Modules for Rationales

In practice, the explanatory texts in the apps examined varied widely. "There are guidelines for developers–from Apple or Android, for example–on how such requests should be designed. However, these are not binding," says Elbitar. As a result, the researchers were able to identify distinct key elements that, when combined, influence how clear and persuasive an explanation is for users. "A key element is the functionality explanation: Good requests clearly state which function the permission is needed for–for example, to 'add photos to your message.' If this explanation is missing, the justification remains vague and simply states that the app 'won't work properly otherwise'," says Elbitar. The way the consequence is phrased also plays an important role: Some requests highlight positively what the user gains by agreeing, while others make it clear which function will be unavailable if the permission is denied. The latter is often perceived by users as more helpful and understandable.

The tone of the app's communication also varies. Some requests address the user directly, either demanding ("You must allow…") or politely asking ("Please allow us…"). Others are phrased more neutrally ("Access to the camera is required") or from the app's perspective ("This app requires…"). Additionally, some requests include extra information designed to build trust or convey a sense of control, such as security assurances like "We do not store any per-

sonal data," statements like "You can change this at any time in the device settings," or links to the privacy policy. Depending on how these elements are combined, a request can appear either more trustworthy or provoke skepticism.

Extracting, Checking, Sorting–the Elaborate Analysis of App Texts

The study consists of two parts: "on the one hand, we analyzed over 9,500 popular Android apps to capture the wording of the rationales. On the other hand, we surveyed 960 people online to learn how they would react to different formulations," says Elbitar.

Above all, the first part required a great deal of detailed work: "We used a machine-learning model to extract thousands of app texts for potential requests–and were able to identify over 35,000 such texts. However, the model does not always clearly recognize these as rationales. It extracts sentences that potentially match–often out of context." In some cases, it turned out that these were simply generally worded sentences that appeared in other areas of the app. Consequently, a lot of manual post-processing was necessary: In the end, the research team manually evaluated 1,054 clear requests from 709 apps via screenshots.

From Experiment to Everyday Life

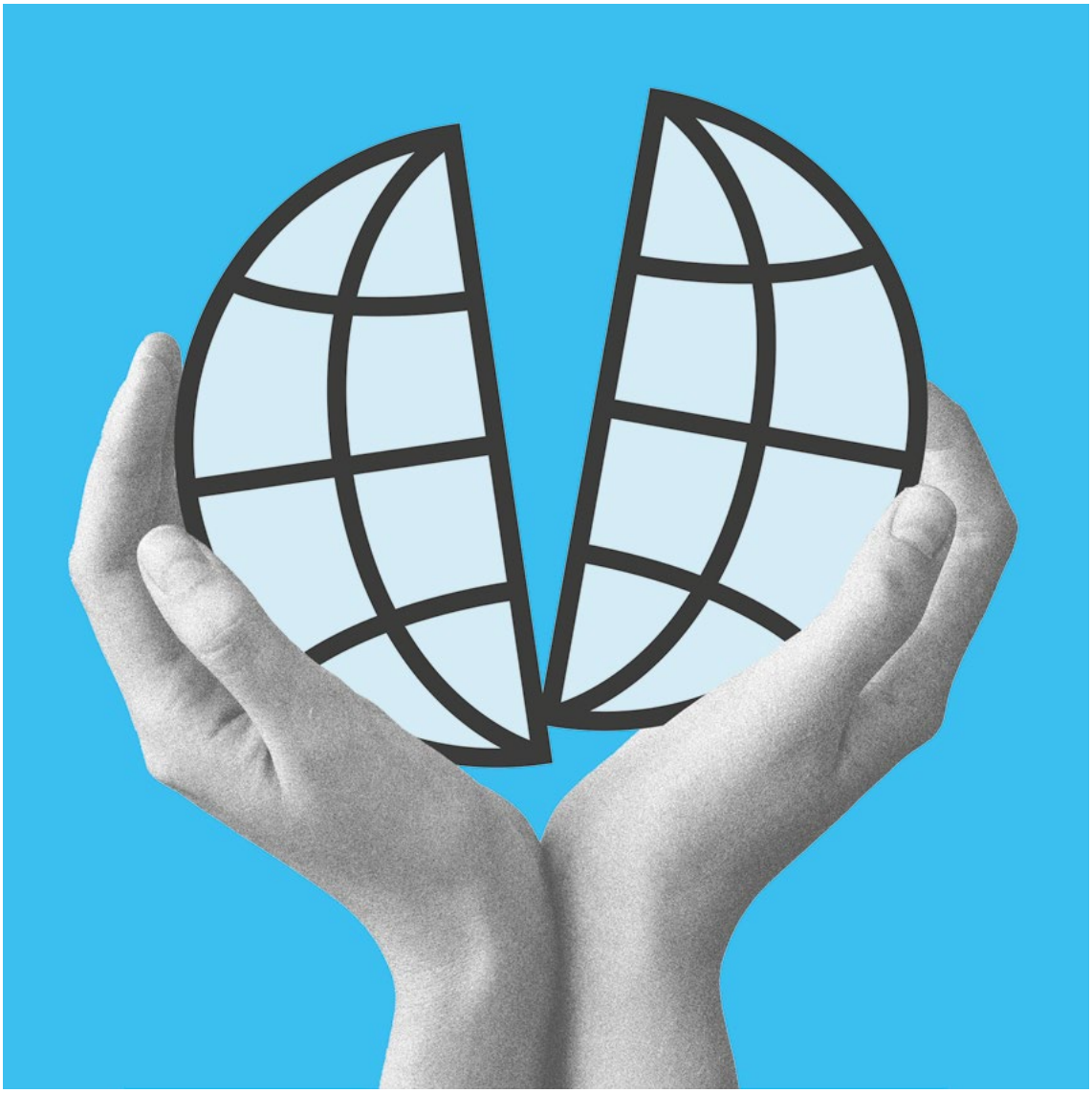
The results provide initial clues for best-practice recommendations to app developers and UX designers. "However, if we want to make reliable predictions about how people respond to certain formulations, we need additional studies under reallife conditions," says Elbitar. In the study, the participants merely imagined that they were using an app. In real usage scenarios–under time pressure or other situational influences–decisions might differ.

Elbitar's interest in the topic began during her master's thesis. At that time, she investigated whether the timing of the permission request also played a role. "The sample size was still small back then–only 46 people under laboratory conditions. This new study was intended to expand on that." In another research project, she focused on permission requests on websites–a field that has been scarcely explored so far. "Websites are often very interactive today. For example, the question arises: Is the request presented as a banner, a button, or an overlay? That, too, can influence the decision."

Although her research primarily provides concrete insights for developers, for Elbitar something else is paramount: "We want app users to be able to make an informed decision about when and to whom they grant access to their data."

»Good requests clearly state which function the permission is needed for– for example, to ‘add photos to your message.’ If this explanation is missing, the justification remains vague and simply states that the app ‘won’t work properly otherwise’«

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)



© Chiara Schwarz

The internet may be a global phenomenon, but its often-claimed global nature is tempered by the 'digital divide'—digital participation still heavily depends on economic conditions. CISPA researcher Masudul Bhuiyan from CISPA-Faculty Dr. Cristian-Alexander Staicu's team explored whether security and data privacy differences can be found on websites as well. His study of 200,000 websites from 20 developing and developed countries concludes that websites in developing and emerging countries are generally smaller and less complex, more prone to efficiency issues, but conversely, potentially less vulnerable to security risks. The complete findings were published in the paper "Digital Disparities: A Comparative Web Measurement Study Across Economic Boundaries," which will be presented at the ACM Web Conference 2025.

Unequal Internet: Differences Between Websites from Industrialized and Emerging Countries

Masudul Bhuiyan

Differences in digitalization between developing and developed countries involve a variety of factors. “Previous studies have primarily examined macro-level indicators, such as the availability of internet, smartphone usage, or general technological infrastructure,” CISPA researcher Masudul Bhuiyan explains. These studies show that only 60 percent of the population in developing countries are online, whereas in developed countries, the figure is 93 percent. Conversely, people in developing countries rely more on mobile internet. Also, technical development there is often rapid and skips entire development stages, which is known as the leapfrogging phenomenon. “Anecdotal stories in the community also suggest that websites in industrialized nations and developing and emerging countries differ significantly. We wanted to know whether there was any truth to this hypothesis,” says Bhuiyan.

Globally Unique Dataset as a Database

For the study, Bhuiyan and his colleagues examined 10,000 websites from each of the ten most populous developing countries and the ten most populous developed countries. “We based this on the definition of the International Monetary Fund (IMF),” says the CISPA researcher. Based on IMF’s classification, the developing countries China, India, Pakistan, Brazil, Nigeria, Bangladesh, Russia, Mexico and the Philippines as well as the developed countries USA, Japan, Germany, France, Great Britain, Italy, South Korea, Spain, Canada and Australia were selected for this study. The researchers then selected the 10,000 most popular websites per country. This criterion meant that the Philippines were included in the sample instead of Ethiopia, as there were not enough websites of sufficient size there. The websites were assigned to a country if they either used the corresponding country code top-level domain, such as .de, or if an address that could be assigned to the country was given in the WHOIS protocol information. “We used the Google Lighthouse and Puppeteer tools to crawl a total of 200,000 websites,” explains

Bhuiyan. “We then specifically examined the website size and complexity, performance optimization, security measures such as the use of https instead of http, data protection applications such as the use of cookies and the integration of current technological features.” Rossow, Ascheman and Pan disclosed their discovery to the affected vendors and a trusted operator community. The CISPA researchers also coordinated a plan for the publication of an attack-specific advisory and started a notification campaign together with The Shadowserver Foundation.

Unclear Picture: Differences Smaller and Less Significant Than Expected

“The result of our investigation is that websites in developing countries are generally smaller and simpler in structure than those in industrialized countries,” explains Bhuiyan. This suits the use of mobile internet, which is widespread in these countries. Nevertheless, the websites incorporate less efficient programming techniques in some respects. For example, inefficient image formats, unnecessary code, and often a lack of responsive design can be found. The use of https, which enables encrypted connections, is also less common.” The CISPA researcher was surprised that the differences between the two groups were not as significant as expected: “In some cases, the differences between individual countries within a group are greater than those between the groups,” Bhuiyan says.

The findings on the use of trackers and cookie were also revealing. “We found more trackers on websites from developed countries than on those from developing countries,” mentions Bhuiyan. “The reason is that developed countries tend to adopt more sophisticated advertising strategies, which rely heavily on trackers—even in the presence of stricter data protection laws.” An unexpected trend was observed when looking at vulnerabilities: Websites in developed countries tend to include more vulnerable libraries, which could potentially be exploitable. “One explanation could be the greater importance of JavaScript libraries for websites in developed countries,” explains the CISPA researcher. “Not only do these improve the functionality of the websites, but they also increase their attack surfaces.”

Outlook and Further Research Desiderata

“The interesting thing about our study,” explains Bhuiyan, “is that we did not find a single major distinguishing factor between the countries. For this reason, further research is necessary. A major success, however, is that we were able to compile this huge dataset, which other researchers can now use.” The dataset with the 200,000 crawled websites is available for download to interested parties on the developer platform GitHub. For the first time, comprehensive datasets featuring websites from countries like Nigeria, Bangladesh, or the Philippines—

regions that have so far not been in the focus of IT security research—can be found there.

Bhuiyan intends to focus his future research partially on websites from Southeast Asia. “In the current study, we noticed that many websites from India, Pakistan, and Bangladesh are in English,” he explains. “The problem with this is that only a small portion of the population there speaks English. We want to investigate which impact language usage on the accessibility of websites as well as the handling of security warnings has.” With this, the CISPA researcher can continue his drive to make the internet as inclusive as possible.

»The result of our investigation is that websites in developing countries are generally smaller and simpler in structure than those in industrialized countries.«

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

Researcher: ***Masudul Bhuiyan***
Author: ***Felix Koltermann***

Publication date
April 29, 2025

Information on setting up a mobile phone and on cybersecurity topics is usually available only in written form. But what happens when the target audience has limited economic resources and low literacy? CISPA researcher Sumair Hashmi and his colleagues explored in a qualitative interview study where and how people from low socioeconomic backgrounds in Pakistan find information to protect themselves against cyberattacks. They presented their paper, "Understanding the Security Advice Mechanisms of Low Socioeconomic Pakistanis," at the Conference on Human Factors in Computing Systems (CHI) 2025, where it received an Honorable Mention.

Cybersecurity Practices of People with Low Socioeconomic Status in Pakistan

Sumair Hashmi

Pakistan is a country with very large social disparities, especially in terms of income and education level. “At the bottom of the social ladder are people who work as cleaning staff, in middle- and upper-class households, or in factories,” explains CISPA researcher Sumair Hashmi. “At the same time, they make up the majority of society. Low income often goes hand in hand with a low literacy rate. I was interested in which cybersecurity threats these people face, how they protect themselves from attacks, and what cybersecurity and data privacy mean to them.” Unlike the behavior of people in industrialized nations, cybersecurity practices among populations in the Global South and in non-English-speaking contexts have so far been little researched.

Hashmi’s research interest was driven by everyday questions: How do people inform themselves about security and privacy? How do they react when they receive a scam call? Those were the kind of questions that intrigued him. “Information on these topics is usually only available in written form and in English,” the researcher continues. “And because people from low-income groups in many cases cannot read or write and generally speak only Urdu, the local language, they have no access to this information.” Hashmi and his fellow researchers from CISPA and the Lahore University of Management Sciences in Pakistan decided to explore the topic through a qualitative interview study, for which they recruited 20 Pakistani participants from the occupational fields mentioned above. The interview study itself was preceded by a phase of ethnographic field observations to gain a deeper understanding of the target group’s living and working conditions.

The Importance of Better Informed Persons and the Work Environment

A total of ten men and ten women between the ages of 20 and 49, with an average monthly income of 30,500 Pakistani rupees (approximately 96 euros), were specifically interviewed for the study. Not all respondents owned personal devices; some used phones belonging to family members or acquaintances. “We identified financial fraud and digital extortion as the most predominant security risks,” Hashmi explained. “We found that all respondents

relied on better-informed individuals—either to set up user accounts on their phones and applications or to seek advice when problems arose," Hashmi continued. Support came either in the form of verbal advice or intermediation.

"Many respondents simply handed their phone to a trusted person and asked them to perform specific actions, such as setting up a password, rather than asking for guidance," Hashmi noted, describing the social dynamics involved. Advice typically came from family members, close friends, or coworkers. The nature of the advice was influenced by the respondents' work environments: "University janitorial staff shared more diverse advice than factory workers. They also had more avenues to seek advice from, such as from their co-workers, supervisors, and even the professors and students on campus," the researcher observed. The security guidance respondents received could be categorized into action-oriented instructions and explanatory advice. The most important practical tips identified were to avoid and block unknown numbers, to check and verify messages and their senders, and to refrain from disclosing personal assets.

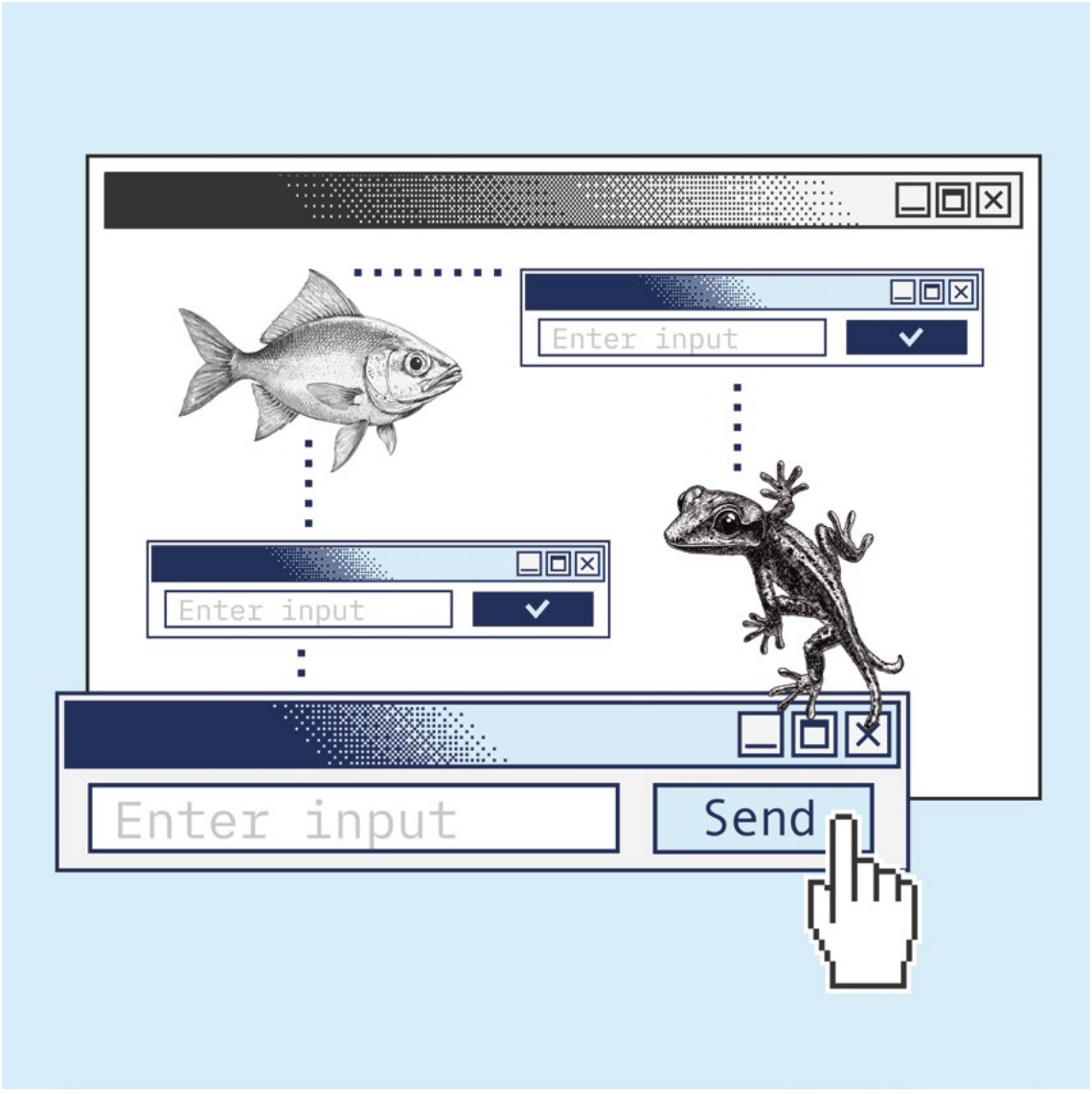
»We identified financial fraud and digital extortion as the most predominant security risks.«

In summary, the study highlights the deeply rooted social embeddedness of security-related practices among Pakistanis with low socioeconomic status. “Our study shows how advice is shared within low socioeconomic communities,” Hashmi explains. Rigid gender roles and the strict norms of Pakistan’s class system play a particularly important role in this context. Family dynamics and the fear of being ridiculed often make it difficult to seek advice. Guidance is typically accepted only when the advisor is perceived as competent and a relationship of trust exists. Another key finding concerned the specific threat landscape faced by the target group in Pakistan: “Due to their precarious financial situation, our study participants are especially vulnerable to scams that lure them with supposedly easy-to-repay loans or lottery winnings,” he notes. Consequently, the nature of the threats faced by respondents tends to exploit human vulnerabilities rather than technological flaws.

Reaching Population Groups with Low Socio-economic Status

The threat landscape for cyberattacks in developing countries such as Pakistan is unique, as attackers exploit people’s financial hardships and their sociocultural norms. “In order to address this situation and to mitigate security and privacy issues for population groups with low socio-economic status, new context-specific guidance and technologies must be developed,” says Hashmi. “Future research should investigate how security advisories can be targeted to reach the most disadvantaged people in Pakistan, providing them with greater protection,” the researcher continues. He plans to continue working in this area and envisions expanding the study to other regions of the world. “It is crucial that we gain a deeper understanding of the needs and cybersecurity practices of people in the Global South—those who do not belong in the Western, Educated, Industrialized, Rich, and Democratic (W.E.I.R.D) population—in order to develop appropriate recommendations,” he asserts.

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



© Alexandra Goweiler

The principle of the survival of the fittest, described by Charles Darwin in the 19th century, has now been applied to software testing: FANDANGO, a new open-source fuzzing tool, uses an evolutionary algorithm to generate myriads of high-quality test inputs that satisfy defined constraints. Advancing language-based testing by a decisive step, FANDANGO employs an iterative procedure that is modeled on biological evolution, yielding customized inputs. The CISPA researchers José Antonio Zamudio Amaya and Professor Dr. Andreas Zeller will present their paper "FANDANGO: Evolving Language-Based Testing" at the International Symposium on Software Testing and Analysis (ISSTA) 2025.

Open-Source Fuzzer with Evolutionary Algorithm Produces Customized Test Inputs

José Antonio Zamudio Amaya

Over the past decade, fuzzers have become the most widely used tools to test software security and robustness. Generating random inputs and feeding them to an application, they help detect undesired program behavior such as bugs and vulnerabilities. With FANDANGO, José Antonio Zamudio Amaya and CISPA-Faculty Prof. Dr. Andreas Zeller have introduced a bio-inspired algorithm to software fuzzing. In an emulation of biological evolution, their algorithm performs a process of mutation and selection to produce inputs that closely correspond to the tester's conditions. Zamudio explains: "The evolutionary algorithm is pretty straightforward. We start with a population of inputs that come from the specifications of a program. And then we do two things: first, mutate those inputs to trigger different changes and second, cross these inputs, which means combining parts of two inputs to produce offspring. We repeat this process and with every iteration, we evaluate the quality of the inputs in terms of meeting the constraints imposed by the tester." This process results in valid test inputs that are customized to specifically explore particular parts of the program that is being tested.

FANDANGO Offers Complete Control Over Test Inputs

While not the first fuzzing tool to automate test generation, FANDANGO is the first tool that gives software testers complete control over the characteristics of the inputs they generate. As Zeller explains: "In contrast to a normal fuzzer, FANDANGO produces inputs which are under the control of the tester, because we assume that the testers a) know what a typical input looks like and b) tend to have an idea where typical bugs might be. They are the ones with the domain knowledge, and we want them to be able to use that domain knowledge when testing a program." FANDANGO enables testers not only to specify the syntax of the input, i.e., the structure they want it to have, but also to define the semantics of the input, i.e., its meaning and specific properties.

To illustrate FANDANGO's benefits for software testing, Zeller uses the example of an online shop for custommade furniture, where customers are required to enter individual

values for height, length, and depth that taken together determine the size of a piece of furniture. “In this case,” Zeller explains, “it would be interesting to see what the program does when I say, for instance, ‘this piece of furniture should have a length of less than zero or a seating surface of one square kilometer.’ Using our evolutionary algorithm, FANDANGO could automatically compute values for all these individual fields—height, length, depth—that would precisely satisfy the condition of this immense surface of one square kilometer.”

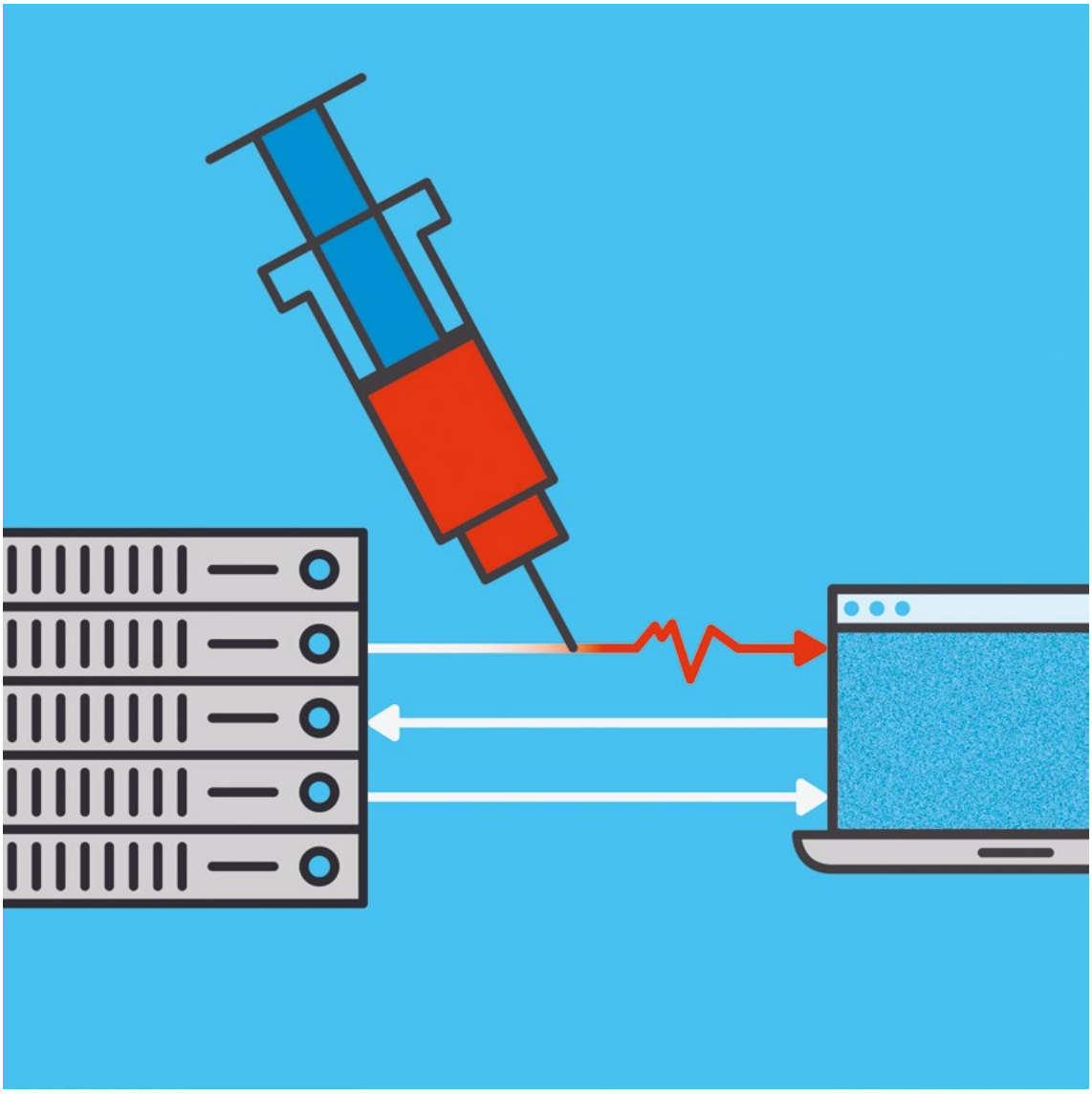
Feedback Invited: FANDANGO is Available on GitHub

To let software testers and programmers benefit from their research, Zamudio and Zeller have made FANDANGO available on GitHub. The program is open-source and comes in the form of a simple command-line tool, accompanied by tutorials and extensive documentation. The CISPA researchers are also openly inviting feedback with the aim of improving their fuzzer even further. “I can’t wait to see how people are using FANDANGO and what they suggest we implement further. I’ve already been talking to people at various companies. The idea of being in control over what should be tested and the idea of being able to check the results of a computation is a real boon to them,” Zeller says.

Funding acknowledgements on page 78.

»The evolutionary algorithm is pretty straightforward. We start with a population of inputs that come from the specifications of a program. And then we do two things: first, mutate those inputs to trigger different changes and second, cross these inputs.«

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

Programs like web browsers and web servers constantly exchange data over the internet. This makes them particularly attractive targets for attackers, which is why we must examine these programs especially thoroughly for vulnerabilities. However, this is where many conventional testing methods reach their limits: As soon as messages are encrypted or communication becomes too complex, they fail. This is precisely where "Fuzztruction-Net" comes in. Developed by the CISPA researcher Nils Bars and his team, this new approach takes a clever route: Instead of directly altering the messages, one of the communication partners is subtly thrown off balance. This makes it possible to uncover new bugs even in widely used and thoroughly tested software. Bars presented his paper "No Peer, no Cry: Network Application Fuzzing via Fault Injection" at the Conference on Computer and Communications Security (CCS) 2024.

Fuzzing Reloaded: Targeted Manipulation for Enhanced Security on the Web

Nils Bars

So-called fuzzers are widely used in software testing. These are automated testing tools that feed programs with random or specially crafted inputs to detect unexpected behavior, crashes, or security vulnerabilities. They are particularly useful for uncovering bugs caused by unusual or edgecase inputs—issues that are often missed during standard testing. “For programs that process inputs or read files in a clearly structured way, fuzzing already works quite well. But testing network programs with fuzzers is much more complicated,” says Bars.

Network Fuzzing with Fault Injection: A New Testing Approach

Why is this the case? “Traditional fuzzers try to replace one of the communication partners in the network, but they do so rather clumsily: They have no real understanding of how to proceed. They don’t know when to send which messages, which keys or session data are required, and they can’t remember previous messages. The other side—whether client or server—eventually notices this and terminates the communication before any message can trigger a bug or expose a security vulnerability,” explains the researcher.

His approach, therefore, is not to replace one of the communication partners like conventional fuzzers do, but instead to subtly manipulate it so that it produces valid, properly encrypted, yet unexpected messages. This technique is known as fault injection. It involves deliberately introducing small errors into the program flow of the communication partner. “The best part is that we can use this method to test both servers and clients. Fuzztruction-Net is the first network fuzzer capable of doing that. Until now, such fuzzers have essentially existed only for servers,” says Bars.

Deeper Testing and New Security Vulnerabilities

In tests, the Fuzztruction-Net prototype delivered impressive results: Compared to previous methods, the new approach achieved, on average, 16 percent greater code coverage—an important metric for test depth—and uncovered three times as many bugs as the best existing network fuzzer. Fuzztruction-Net even found vulnerabilities in well-tested programs such as Nginx, the OpenSSH client,

and cURL. “We focused on a specific class of bugs known as memory corruption. These are programming errors where a program modifies data in a memory region it should not have access to. This can lead to crashes, data loss, and critical security vulnerabilities,” explains Bars.

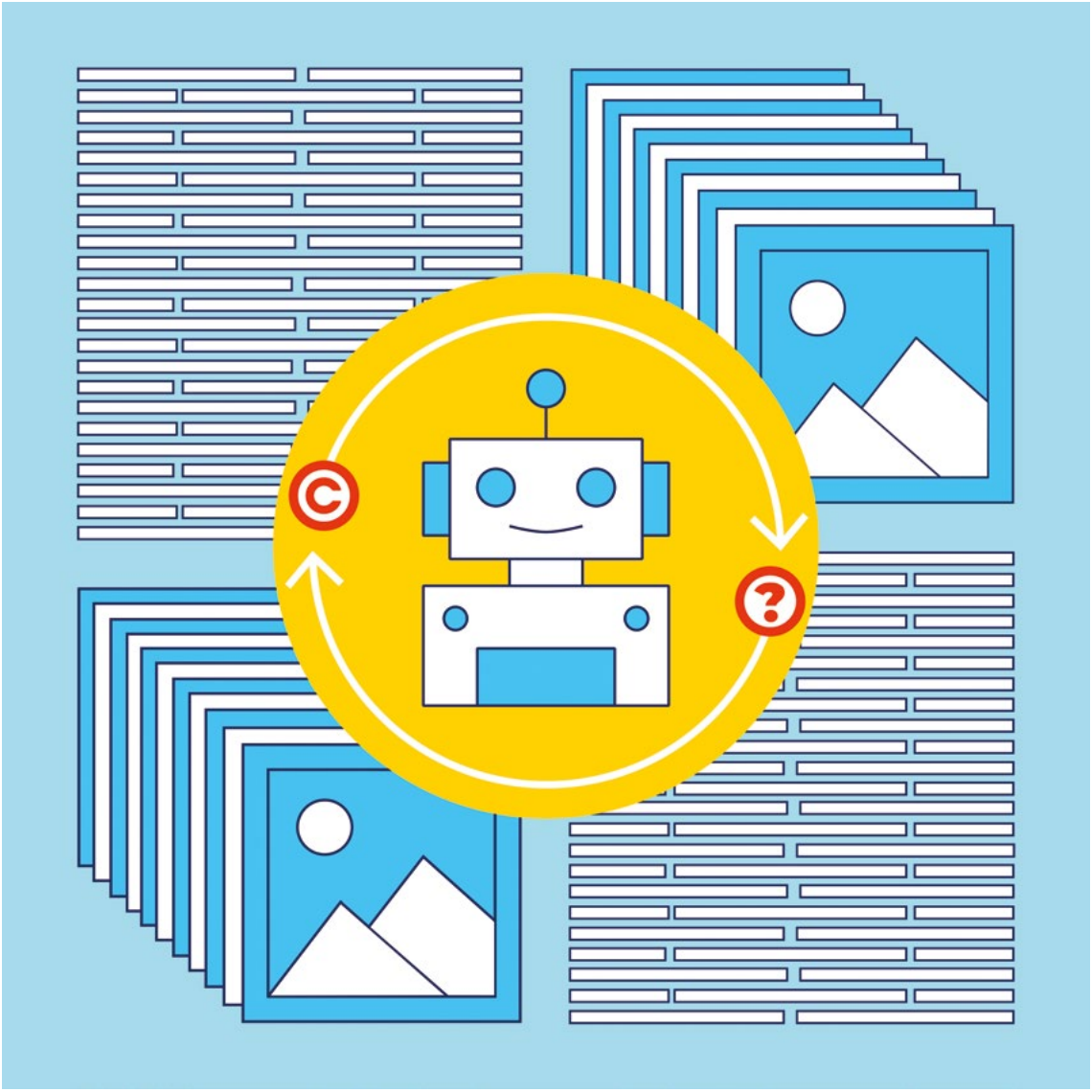
In total, the research team discovered 23 previously unknown security flaws in widely used network infrastructure—many of which can be exploited remotely—underscoring the relevance and potential of this new approach.

Helpful Insights for Developers: Fuzzstruction-Net Makes Security Flaws Reproducible

Unfortunately, fuzzers still cannot fix the bugs themselves. Developer effort is still required to track down and resolve the issues. “The key point, however, is that the bugs are reproducible. This gives developers a clear indication of where the problem arises, making it possible to fix,” explains Bars. Nginx has already expressed interest in using Fuzzstruction-Net. “Our prototype works well, but there’s certainly room for optimization before it can be used in long-term deployment,” says the researcher. The prototype is already available as open source to anyone interested.

»We focused on a specific class of bugs known as memory corruption. These are programming errors where a program modifies data in a memory region it should not have access to. This can lead to crashes, data loss, and critical security vulnerabilities.«

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)



© Alexandra Goweiler

In just a few years, what began as a scientific project—to use AI models for generating images—has become an everyday application. Along with this development, new problems have emerged. Increasingly, creators—such as photographers and illustrators—are asking whether their images have been used to train AI models. CISPA researcher Antoni Kowalczuk has now developed a technique that can prove whether specific images were employed in a model's training. He published his findings in the paper "CDI: Copyrighted Data Identification in Diffusion Models" at the IEEE Conference on Computer Vision and Pattern Recognition 2025.

A New Method Can Detect Whether Copyright-protected Images Were Used to Train AI Models

Antoni Kowalczuk

AI image generators have experienced explosive growth in recent years. Many of these systems—such as DALL·E, Midjourney, and Stable Diffusion—are based on so called diffusion models. “A diffusion model is a deep neural network that learns to generate images step by step by gradually removing noise from an image,” explains Antoni Kowalczuk, a PhD student and researcher at CISPA. These systems were trained on millions of images from the internet, allegedly without the creators’ consent, raising legal and ethical issues. “When the models were still used purely for scientific purposes, nobody really cared about the copyright question,” Kowalczuk recalls. “But once people started making money with these models, the issue suddenly became relevant. I thought my research could make a difference.”

Why Previous Methods Fail

Existing techniques for determining whether AI models used particular images for training rely on a method called “Membership Inference Attacks” (MIA). These try to assess if a single image was used to train an AI model. However, research shows that when the models and their datasets grow—and they only tend to do so—the efficacy of MIAs falls to almost zero. “For this reason, my colleagues and I developed a new method called ‘Copyrighted Data Identification’ (CDI),” says the CISPA researcher. “The key idea behind CDI is that we don’t examine individual images, but entire datasets—for example, a collection of stock photos or a digital art portfolio.”

How CDI works

To check whether copyright protected material was used to train an AI model, Kowalczuk designed a four stage process for CDI. First, two datasets must be assembled: “The first contains images that the data owner believes were used to train this specific model. The second is a so called validation set, made up of images we are 100% certain were not used in training,” explains the researcher. Next, both datasets are run through the AI model to observe its responses. Based on those responses, a model is trained that can predict whether the dataset in question was likely part of the training data. “At the end, a statisti-

cal test is performed to see whether the suspect dataset systematically scores higher than the validation set," says Kowalczuk. If it does, that is strong evidence the AI was trained on those images, and if not, the result remains inconclusive.

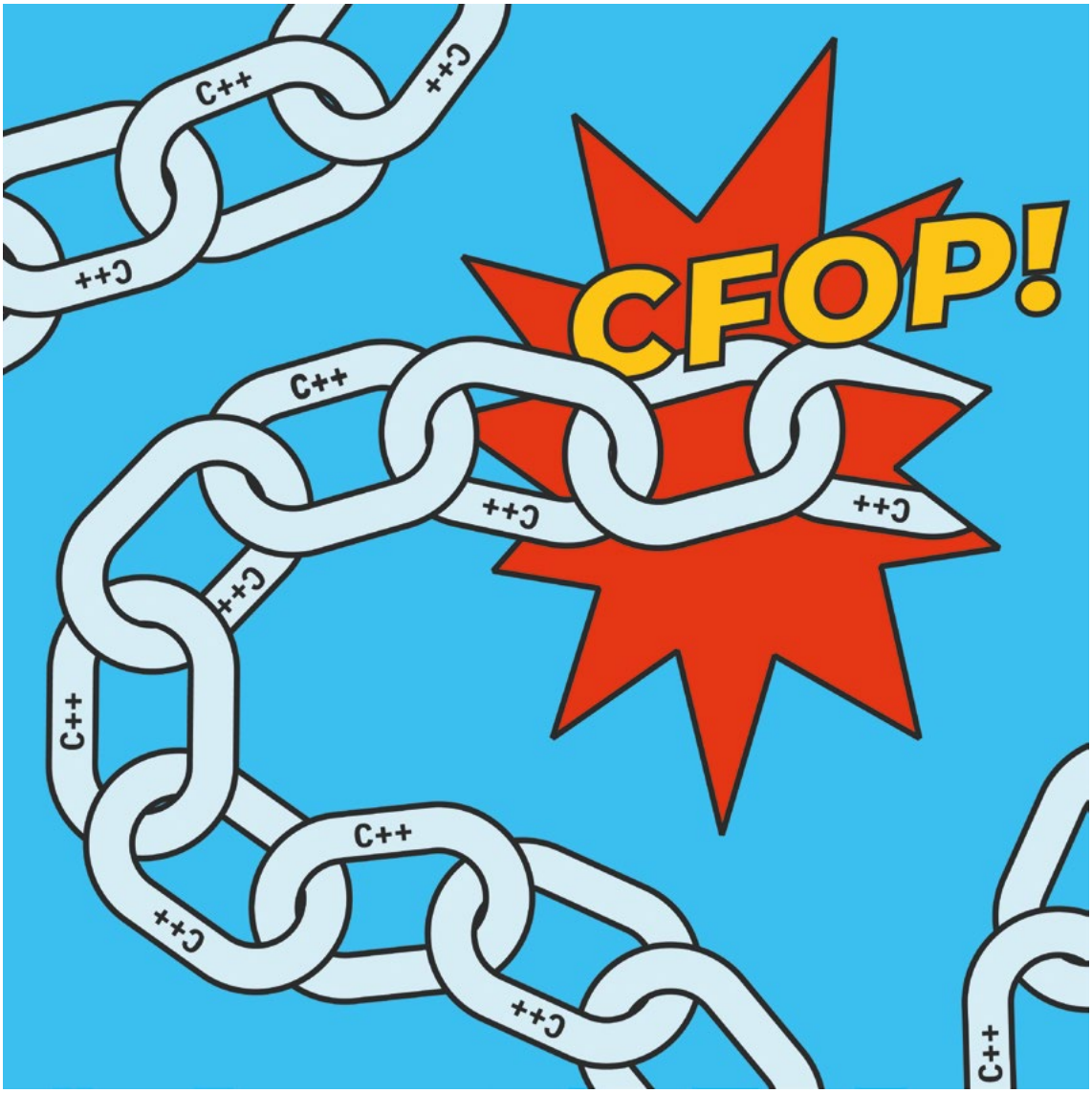
The CISPA researcher tested CDI a suite of existing AI models with available information about training data, for example, models trained on the ImageNet dataset. The results are promising, Kowalczuk reports: "CDI can detect with high accuracy whether a dataset was used in training, even on complex, large scale models. Even when we are unable to pinpoint the exact images used in training, we can still successfully recognize if data in the set was used to train the model. CDI also yields reliable results when only a subset of the entire work was included in training."

»The key idea behind CDI is that we don't examine individual images, but entire datasets—for example, a collection of stock photos or a digital art portfolio.«

Obstacles to Practical Application and Deployment

At present, CDI remains a method whose use—because of its complexity—is largely confined to researchers. “Some of the features we extract require full access to the model and its code,” notes Kowalczuk. “Moreover, there are very stringent criteria for the data samples we employ.” As a result, CDI currently offers mainly a theoretical proof of concept that it is possible to determine whether a particular set of images was used to train AI models. Developing a user friendly application for creators without deep technical expertise would require further modifications and advances that, for now, appear technically out of reach. “CDI is still quite young and there is much work to be done. But one thing is clear: once we have better methods, we may someday bridge the gap between theory and practical implementation,” the CISPA-researcher concludes.

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



© Chiara Schwarz

A code-reuse attack named Coroutine Frame-Oriented Programming (CFOP) is capable of exploiting C++ coroutines across three major compilers, namely Clang/LLVM, GCC and MSVC. CFOP even succeeds in environments that are protected by Control Flow Integrity (CFI), exposing relevant gaps in 15 of these defense schemes. Rather than injecting new code, CFOP chains together existing functions, achieving arbitrary code execution after corrupting coroutine-internal memory structures. This new exploitation technique has been discovered by the CISPA researchers Marcos Sanchez Bajo and Professor Dr. Christian Rossow. Their paper "Await() a Second: Evading Control Flow Integrity by Hijacking C++ Coroutines" will be presented at the Usenix Security Symposium 2025.

C++ Coroutines: Prone to Code-reuse Attack despite CFI

Marcos Sanchez Bajo

Devising a novel code-reuse attack, Marcos Sanchez Bajo and CISPA-Faculty Professor Dr. Christian Rossow have demonstrated that all existing implementations of C++ coroutines can be exploited to bypass state-of-the-art CFI protections in both Linux and Windows. Called Coroutine Frame-Oriented Programming (CFOP), the attack results in a corruption of heap memory, allowing attackers to manipulate data and assume complete control over applications. A relatively recent addition to C++, coroutines are already present in more than 130 unique popular GitHub repositories. “They’re being used to pause and resume functions,” Bajo explains, “which is very useful for asynchronous programming, for example in servers, databases and web browsers.”

Connecting C++ Coroutine Functions to Corrupt Heap Memory

In more concrete terms, coroutines can, for instance, be used to create generators that produce a sequence of elements. Imagine a Fibonacci series, where each new number in the series is the sum of the two numbers that have gone before. After each new number in the series, the coroutine is paused until it is called to generate the next one. In CFOP, entire C++ coroutines and other existing functions are used to create a code-reuse attack, as Bajo explains: “With codereuse attacks in general, attackers take snippets of code that belong to the application anyway, so no new code is injected. They then form chains of these code snippets to manipulate the program’s execution flow. But bypassing CFI protections is a little more difficult. Instead of just taking snippets of code and creating chains, you have to take full coroutine functions and connect them in smart ways.” Once the CFI protections are circumvented by hijacking a coroutine function in this manner, any other existing function can be submitted to a code-reuse attack.

CFI Schemes Fail to Protect C++ Coroutines

Introduced to protect against code-reuse attacks, CFI schemes ensure that the correct program execution flow is observed. Programming languages, however, evolve dynamically, while CFI schemes only protect the programming paradigms that were present at the time of their creation, as Bajo points out: “The main problem with CFI is that this defense is static in time, meaning that it only covers the possibilities of a programming language as is.

If new features are introduced to the programming language later on, CFI does not recognize them and cannot deal with them because it was created based on an older version of the programming language." In their study, Bajo and Rossow found that only 7 out of the 15 CFI schemes they considered initially were compatible with coroutines. Of these 7, only 2 (IBT and Control Flow Guard) provided partial protection against the exploitation of coroutines, while the remaining 5 provided none. "In the end," Bajo summarizes, "we were able to bypass all of them. With CFOP, you can still do all the things that were possible previous to CFI."

Patching CFOP Is a Structural Issue

The fact that C++ coroutines are enjoying increasing popularity exacerbates the potential reach of CFOP. Bajo says: "Coroutines were introduced to C++ in 2020 and, since then, developers have been using them more and more. Unfortunately, we found that coroutines have certain structures in memory that can be targeted by attackers. To the best of our knowledge, this has not yet been exploited in real life." Essentially, CFOP is possible because the three major compilers implement C++ coroutines in a way that renders them structurally vulnerable. Bajo says: "Mitigating this exploitation technique is not as easy as patching the code–this is a structural issue and you need to rethink how the application works internally." Bajo and Rossow have developed successful implementation alternatives for C++ coroutines and reported these mitigations to Clang/LLVM, GCC and MVSC in November 2024.

»The main problem with CFI is that this defense is static in time, meaning that it only covers the possibilities of a programming language as is. If new features are introduced to the programming language later on, CFI does not recognize them.«

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

Researcher: ***Marcos Sanchez Bajo***
Author: ***Eva Michely***

Publication date
August 4, 2025

© Janine Paulus

If you think of software as a building, you might say it's made up of code blocks. Many of these building blocks are custom-built for a specific application; others are standard components and used in many buildings such as, for example, cryptographic algorithms. If these building blocks become brittle with age, the security of the entire application degrades. In a qualitative interview study, CISPA researcher Alexander Krause explored the challenges faced by experienced software developers when they want to renew existing crypto implementations—or even create better cryptographic building blocks from scratch. He will present his paper "'That's my perspective from 30 years of doing this': An Interview Study on Practices, Experiences, and Challenges of Updating Cryptographic Code" at the Usenix Security Symposium 2025.

How Agile is Your Crypto? Interview Study Explores Cryptographic Update Processes

Alexander Krause

Cryptographic algorithms are fundamental building blocks in the development of new applications. They ensure that data and information can be communicated in encrypted form. Unlike most other code sequences, certain cryptographic implementations lose their effectiveness over time: As other technological fields advance, for example, if computers significantly gain processing power, asymmetric encryption can potentially become vulnerable. Quantum computing is a prime example of this. Computing with three instead of only two possible states enables quantum machines to solve mathematical problems much faster, and to use new, more efficient algorithms that aren't available on "conventional" computers.

Updating cryptographic implementations is thus a recurring task—and one with far-reaching implications for software users. If crypto updates go awry, the consequences for overall software security can be severe. In this context, Krause refers to the concept of "crypto agility": "This recurring update process for cryptographic implementations ideally begins with something called 'crypto agility.' It means that when developers are designing a software, they already keep in mind that they may need to replace or update the cryptographic implementation at some point in the future." Thinking ahead in this way is meant to facilitate updating the software later on with state-of-the-art cryptographic methods. However, executing crypto updates requires highly specialized knowledge that many software developers do not possess.

Crypto Libraries Require Maintenance

Cryptographic implementations tend to come from publicly accessible, free crypto libraries that are maintained by specialized developer communities. These open-source projects, which benefit developers around the world, are usually supported by just a handful of individuals who contribute their time on a volunteer basis. While reusing existing algorithms and functions makes for efficient programming, it also introduces unique security risks where cryptography is concerned. If crypto libraries are not properly maintained and bugs go unfixed, those vulnerabilities can proliferate across a wide range of app-

lications. In the context of the "supply chain"—that is, a kind of dependency of software projects from other resources—this creates what's known as a "single point of failure." If a crypto library is not reliably maintained, it can jeopardize the functionality of all products that rely on it within the supply chain.

Crypto Updates and Their Challenges

Conducting a qualitative interview study with 21 participants, Alexander Krause and his colleagues have explored the challenges that software developers, who usually aren't crypto experts themselves, face when updating cryptographic implementations. Their goal was to find answers to four narrowly defined research questions: How do developers learn about a recommended crypto update? What goals do they pursue with the update? What processes do they follow when planning and executing a crypto update? And finally, what experiences did they gain when carrying out those updates? "There's already a lot of research on updating software projects in general," says Krause. "But here, we wanted to explore whether expert populations with highly specialized knowledge have unique requirements, too."

Heterogeneous Results: Crypto Updates Depend on Context

One of the key findings of the interview study is that the information flow around recommended crypto updates is inconsistent and sometimes incomplete. Updates were primarily triggered by information that developers received through sources like blogs, social media, and GitHub. However, depending on their institutional affiliation, some developer groups are more likely to receive information about updates than their colleagues. "If you work for a large company, there are often agreements. They often receive advance notice of vulnerabilities and can be the first to patch them—for example as part of a disclosure process. This information is passed on through private mailing lists that only a few people have access to," Krause summarizes. "It was a big takeaway for us that it is difficult to get into these communities."

The interview study also revealed that there rarely are structured processes to manage crypto updates in companies or projects. Prioritization of such updates sometimes depended on resources such as team size. Decision-making processes and responsibilities around crypto updates were also unclear at times. "Who decides who's responsible for a crypto update? This varied a lot," Krause says. "Sometimes there actually were leaders assigned to it. In other cases, it was, 'You just discovered yourself that there is this vulnerability, so it's your job to fix it.'" As one of their key research contributions, the researchers have outlined such an update process, consolidating the heterogeneous statements that the participants had made.

Other study results turned out to be both more positive and predictable for the research team, such as for example the motivations behind implementing cryptographic updates. “We were positively surprised overall that many developers are intrinsically motivated to ensure their software is future-proof,” Krause explains. In addition, preventive updates were performed to gain a security edge over future threats. Feedback was also fairly consistent regarding the perception that crypto updates are onerous and complex. Krause summarizes: “All our participants had very individual backgrounds and very individual projects, but overall, what makes updating crypto difficult is that you need the knowledge to do it—and at the end of the day many don’t have that.”

Networking is Key: A Gap Between Research and Practice

The question of how this knowledge gap could be closed in the interest of IT security continues to occupy Krause. “The biggest challenge that we see—and this extends beyond our paper to crypto research more broadly—is translating new research findings into a format that actually reaches developers.” While gaining access to the relevant mailing lists is often difficult, the responses from the interview study have shown that software developers rarely use academic publication databases to stay informed about new developments. “In our study, those with a higher academic degree—a master’s or PhD—had an advantage here, because they bring the necessary skillset,” Krause explains. Ultimately, obtaining relevant information still largely depends on the personal initiative of individual developers. In this respect, there is a clear gap between research and practice that needs to be bridged. The CISPA researchers have already made their findings available to all developers who participated in the interview study.

Funding acknowledgements on page 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

Researcher: Alexander Krause
Author: Eva Michely

Publication date
August 11, 2025

© Chiara Schwarz

CISPA researcher Sarath Sivaprasad, together with Hui-Po Wang and Mario Fritz from CISPA and other colleagues from HIPS, has developed an AI system that can automatically detect abnormalities in zebrafish embryo development. The approach combines a large-scale, high-resolution image dataset with a transformerbased machine learning model to identify toxicity effects and fertility outcomes with high accuracy and efficiency. This advancement could significantly speed up drug screening processes. The paper "Automated Detection of Abnormalities in Zebrafish Development" will be presented at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2025.

AI Accelerates Drug Discovery Through the Automatic Analysis of Zebrafish Embryos

Sarath Sivaprasad

Anomaly detection has been a focus of Sivaprasad's research for quite some time. "In machine learning anomaly detection is the process of identifying data points, events, or patterns that deviate significantly from the expected behavior," he explains. "During training, the system learns what 'normal' looks like, and at inference each sample is scored by how much it deviates from that notion of normal. Unlike traditional classification, which assigns inputs to specific categories (e.g., cat, dog, or car), anomaly detection focuses on distinguishing between 'A' and 'not A.'" In this latest publication, a similar concept is applied to the biological sciences. "In this case, we applied a version of anomaly detection to observe the development of zebrafish embryos," the researcher explains.

Zebrafish: A Tiny Powerhouse in Drug Discovery

"Zebrafish are an excellent model organism for biomedical research," Sivaprasad says. "This is due to their transparent bodies and genetic similarities to humans." Their rapid development and responsiveness to chemicals make them ideal for high-throughput toxicity screening—an important methodology used in drug discovery. "However, analyzing their development still relies heavily on expert manual inspection—a time-consuming and subjective process." The challenge here lies in accurately detecting subtle developmental abnormalities that emerge over time in image sequences. "Existing datasets lack both the temporal span and the scale required to train large-scale models," Sivaprasad adds.

A Breakthrough Dataset and Model

To overcome this bottleneck, Sivaprasad's colleagues at HIPS first compiled one of the most comprehensive image datasets of zebrafish embryonic development to date, comprising more than 185,000 microscopic images. "They placed zebrafish embryos in wells, monitored them under the microscope, and captured their development continuously," he explains. The dataset covers two critical experimental tasks:

- Fertility classification: 130,368 images over 8-hour sequences to determine egg viability.

- Toxicity assessment: 55,296 images over 48 hours to detect toxic compound effects.

Images for fertility detection were annotated with sequence-level labels, and developmental anomalies had fine-grained temporal annotations, creating a valuable benchmark for developing and testing automated tools.

Transformer-Based AI Boosts Accuracy

The second step was to train a model on this dataset. Sivaprasad trained a new transformer-based neural network that can interpret both the structure of each image and how embryos change over time in the sequences. The AI achieved 98% accuracy in identifying whether an embryo was fertilized and 92% accuracy in detecting developmental abnormalities caused by exposure to a toxic compound (3,4-dichloroaniline). Importantly, the model mimics how human experts analyze developmental progression over time, enabling early predictions of toxicity.

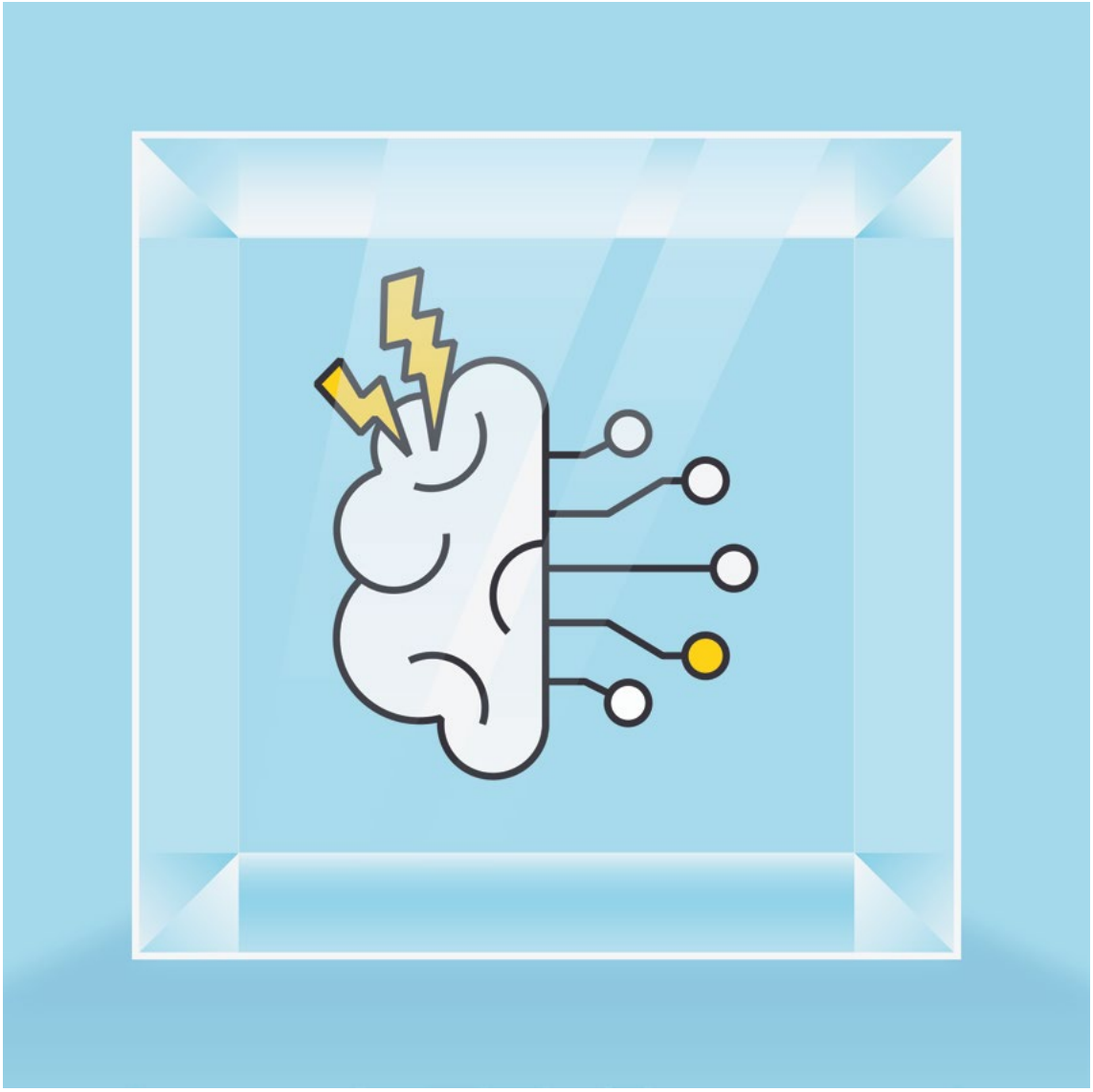
»Right now, we evaluate only one chemical to understand how anomalies develop. Our goal is to scale this up to an entire library of chemicals.«

A Platform for Future Innovation

This dataset and model lay the groundwork for future research into early-stage developmental toxicity, improving both sensitivity and prediction speed. “Right now, we evaluate only one chemical to understand how anomalies develop. Our goal is to scale this up to an entire library of chemicals,” Sivaprasad says. The complete dataset will be made freely available on GitHub, allowing other researchers to use and expand upon it at no cost. The aim is to empower both the biomedical and AI research communities to create more advanced, efficient, and ethical toxicity-screening methods. “It’s a valuable resource for the machine learning community to benchmark their methods—and for biomedical research to better understand the effects of different drugs,” Sivaprasad adds.

Funding acknowledgements on page 78.

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, 2025, Daejeon, Korea, Conference: Medical Image Computing and Computer Assisted Intervention



© Chiara Schwarz

Liberate AI, an interdisciplinary project uniting researchers from the medical domain, computer science, and trustworthy AI, aims to develop an AI model capable of supporting doctors in the treatment of ischemic stroke. Serving as a digital assistance system, it is intended to predict the long-term outcome of patients after mechanical thrombectomy as well as potential complications. The AI model will be trained in a privacy-preserving fashion on medical data residing at different sites across Germany. Further challenges addressed in Liberate AI are the AI's explainability as well as its ability to make differentiated predictions for patient subgroups. Funded by the Helmholtz Association with 250,000 euros, Liberate AI is a joint project between DZNE, the University Hospital Bonn, and CISPA.

From Black Box to Glass Box: AI Explainability in Stroke Treatment

Jilles Vreeken

Ischemic stroke occurs when blood clots become lodged in brain vessels, obstructing blood flow and, hence, oxygen supply. One possible treatment in this situation is mechanical thrombectomy, a minimally invasive procedure that uses a special catheter to remove vessel blockage. If mechanical thrombectomy is the most promising option for any given patient, however, depends on a number of case-specific factors. To assist doctors in making this time-critical decision, the researchers in Liberate AI seek to train an AI model on medical data stored in the German Stroke Register as well as associated MRI and CT scans held at various hospitals across Germany. To achieve this end, they leverage Swarm Learning, an AI technology developed by DZNE in cooperation with Hewlett Packard Enterprise. Swarm Learning effectively allows the AI to learn in a decentralized fashion, traveling to all the data repositories in the network to collect knowledge without the data itself leaving the sites where it is stored.

Toward a Glass Box: Explainability is Key

While Swarm Learning is at the heart of Liberate AI, there are further technological challenges that go beyond the actual training of the AI model. The first of these challenges concerns explainability as one of the key characteristics the AI model needs to exhibit. In contrast to deep-learning-based applications, which tend to operate on a black-box basis, the AI model developed in Liberate AI will have to make its reasoning transparent to the medical doctor using it. As CISPA researcher Professor Dr. Jilles Vreeken, an expert on trustworthy information processing, explains: “We want to develop a glass-box AI that can predict as well as a black-box one. Because if you are a medical doctor and the AI says ‘yes’ or ‘no,’ your first question is ‘Why should I trust you?’. This means we need to use explainable AI, which is the branch of AI research in which we develop AI models where we can understand based on which evidence they are saying what they are saying. This is the form of AI that can really support experts, because medical doctors will be able to tell whether the AI’s prediction is based on accidental evidence or on actual biomarkers.” While Vreeken and his research group aim at designing a transparent AI model, explainability poses specific technological challenges in the context of Swarm Learning. “We have to keep in mind that

while we can develop this glass-box AI, it will still need to be able to learn in a swarm learning environment and to predict as reliably as a black-box one. It is not trivial to make that happen," he says. In the project, the researchers will have to strike the balance between the AI model's degree of transparency and its ability to be swarmed successfully.

In Search of Subpopulations and Causal Conclusions

The second major challenge addressed by the CISPA researchers concerns the identification of those patient populations that respond positively or negatively to mechanical thrombectomy in terms of quality of life over time. Ideally, the AI model will be able to automatically identify these statistical subgroups based on certain patterns that it extracts from the accumulated medical data it has digested. "The question is, can we develop a transparent box AI that can find conditions under which people have exceptional survival behavior? For example, this could depend on the size of the blood clot, high or low blood pressure, genetic factors, or the intake of blood thinners—you can think of certain conditions that will select some patients but not all of them," Vreeken explains. These subgroups, he points out, can still be identified even if the training of an explainable glass-box AI should prove impossible in the Swarm Learning environment. "The beauty of our transparent box AI is that we can use it on top of a black-box AI, namely we can ask the question: For which people does the black-box AI tend to predict very well? This means that if we end up using a black-box AI because it's more accurate than any transparent model we can develop, we are still able to determine the subgroups for which we should or should not ask its opinion."

Liberate AI: Combining Domain Expertise with Machine Learning

Ultimately, what the CISPA researchers would like to design is a transparent AI system capable of giving causal guarantees for its predictions. So that if, for instance, it predicted that high blood pressure will minimize the efficacy of the treatment, it could also give the reasons for this. "This is very difficult to do," Vreeken says, "because you need a randomized control trial to determine if high blood pressure is the only factor or a confounder, something that seems to be relevant but that isn't. So, the ultimate AI that we would like to develop is a glass-box AI capable of saying: Based on all available stroke data, there's a clear difference between otherwise similar patients that cannot be explained in any other way except for blood pressure."

Even if the tripartite challenge of offering explainability, subgroup identification, and causal guarantees should prove too ambitious in the end, Vreeken is certain that Liberate AI will be making a significant contribution to

the applicability of AI in healthcare. The interdisciplinarity of the project team in particular opens up new possibilities for the treatment of acute stroke, as he highlights: “If you ask a domain expert what they want, they want a better machine X, but maybe they need something else that they don’t know is possible. The opposite problem is that computer scientists often develop new machines where the domain expert might say, this solves a problem that we do not have. I am very happy that in this project we have an excellent constellation of people with computer science expertise, people with pure medical domain expertise, and people in the middle. In Liberate AI, we will not be developing a machine that nobody is waiting for, but rather the machine people don’t even know they need.”

Funding acknowledgements on page 78.

»I am very happy that in this project we have an excellent constellation of people with computer science expertise, people with pure medical domain expertise, and people in the middle.«

© Janine Paulus

Passwords remain "the go-to authentication tool" in everyday life, says CISPA researcher Alexander Ponticello. At the same time, passwords are often a security weak spot: too short, too simple, and reused far too often. Blind and low-vision people face an additional hurdle: Systems need to work together sensibly for authentication processes to run smoothly. A new qualitative study with 33 U.S. participants shows how this group manages passwords—and where improvements are needed. Alexander Ponticello presented his paper "How Blind and Low-Vision Users Manage Their Passwords" at the renowned Conference on Computer and Communications Security (CCS) 2025.

How Blind and Low-Vision Users Manage Their Passwords

Alexander Ponticello

Passwords are still the default tool for online security—but they're also a constant source of problems. Many people today have hundreds of accounts for which they must manage passwords of varying complexity. Password managers can help: They create strong passwords, store them, and autofill login credentials—problem solved, right? Unfortunately, this isn't the case, because password managers are far from being used consistently by everyone. Previous studies show that the main reasons are the fear of complicated setup, lack of trust, and lack of knowledge about existing tools. Older user groups also tend to be generally hesitant about digital tools. Alexander Ponticello's new study expands research on password management and password manager use to a group that has received little attention so far: blind and low-vision users.

Widespread Use of Password Managers in the Community

Password managers can be an important tool for blind and low-vision people to manage their login credentials. "In fact, all 33 respondents in our study used password managers—sometimes consciously, sometimes unconsciously, simply because their browser or device offered to manage them." These included third-party programs such as LastPass or 1Password, as well as browser-integrated password managers like the one built into Google Chrome and system-integrated password managers such as Apple Passwords. "Those who intentionally chose a password manager usually relied on recommendations from acquaintances or advice in relevant forums. Accessibility played at least as important a role as system security," Ponticello explains.

Real Accessibility Only if Systems Work Together

Depending on the degree of impairment, blind and low-vision users rely primarily on screen readers to use their devices in everyday life. "Our first intuition was that it must be a big problem that screen readers read passwords aloud in public. However, this proved to be less of a problem, as almost all study participants told us that they use headphones," says the researcher. In addition, the speech output usually runs so fast that bystanders can hardly understand anything. However, for blind and low-vision people to use password managers smoothly, screen readers, password managers, apps, and websites must work together accordingly. "If one of these parties

fails, the whole system breaks down," says Ponticello. Unfortunately, there are still programs where accessibility seems to be an afterthought. At the latest when updates need to be installed, some users have experienced that programs no longer work properly. The result: Users feel they cannot reliably depend on the systems.

Security Versus Everyday Life: Compromises Are Common

Many of the users surveyed therefore combine password managers with backup strategies. Some even keep password lists in Braille—safely stored, but still analog. "That's not inherently insecure," the researcher explains. "But you have to be aware of who might have access to that list." Other study participants said they intentionally create simpler passwords so they can enter them without a tool if necessary. „That contradicts security best practices," he says, "but above all it shows that systems need to become more reliable."

What (Still) Needs to Be Done—And How To Do it Better

According to Ponticello, one problem is how password managers generate passwords: Random passwords with special characters are often hard for blind people to find on the keyboard. A better alternative would be passphrases that string whole words together. "Unfortunately, screen readers then read those passwords letter by letter instead of recognizing the words. The integration hasn't been thought through to the end," the researcher says. App stores could also help by clearly labeling a tool's accessibility and introducing special review categories for affected users where blind and low-vision people can get information directly. "But the most important thing is: We need accessibility by design—correct labels for buttons, a sensible focus order, and consistent screen reader flows."

Outlook

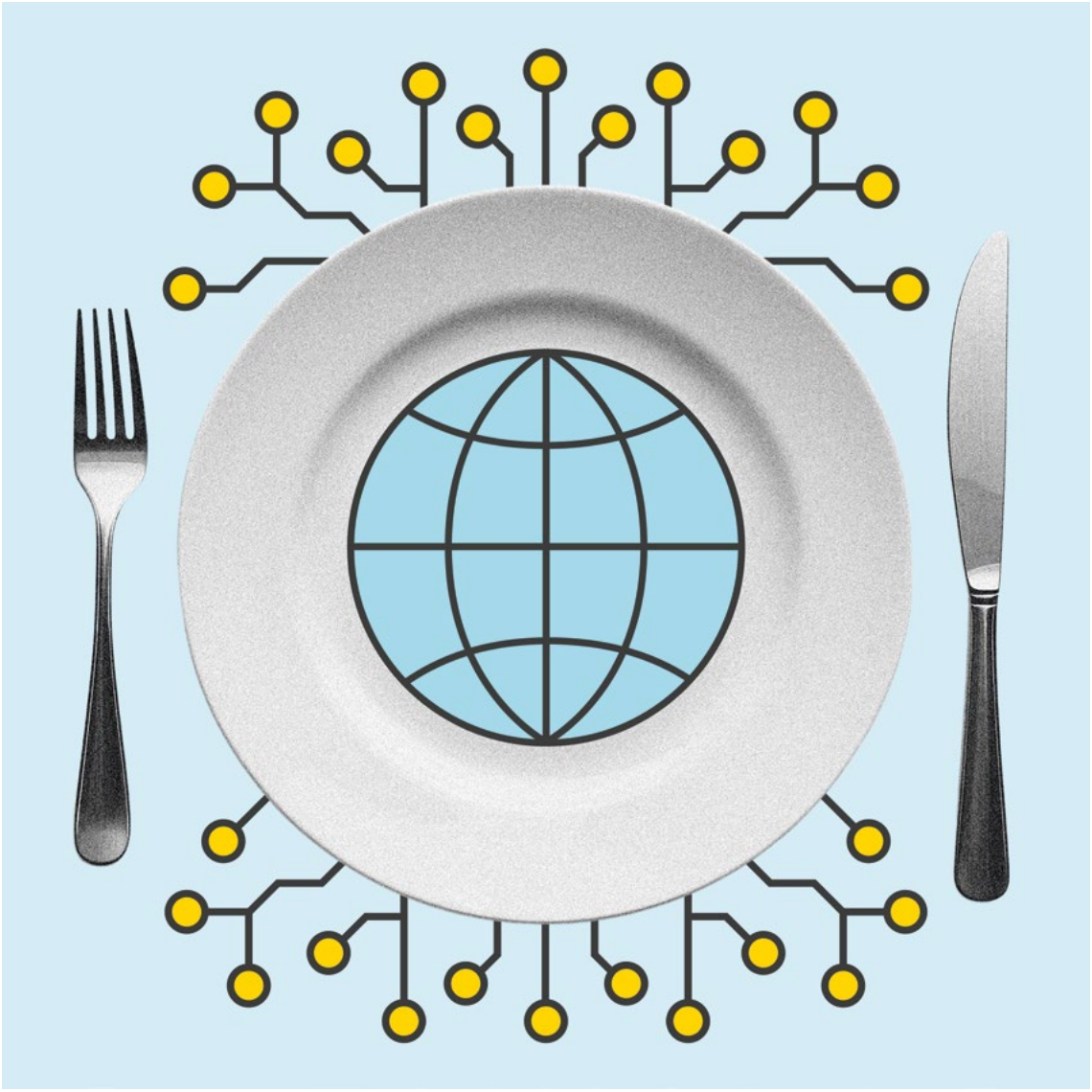
Conducting a similar study with German users could be Ponticello's next step. So far, legislation in the U.S. has been stricter than in the EU. Laws such as the "Americans with Disabilities Act" have long enforced strict accessibility standards for websites and digital services there. The EU is following suit with the "European Accessibility Act" (EAA). In Germany, this led to the "Accessibility Strengthening Act," which has been required to be applied since June 28, 2025. "I'm curious to see what effects this will have in the future." Ponticello's study shows: Accessibility is not a luxury but a basic prerequisite for digital security. Many hurdles—from lack of labeling to fragile integrations—can be solved if platforms, developers, and lawmakers take them seriously. "We need to adapt the systems, not the people," the researcher says. "Only then can passwords be used securely by everyone."

»Unfortunately, screen readers then read those passwords letter by letter instead of recognizing the words. The integration hasn’t been thought through to the end.«

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)

Researcher: ***Alexander Ponticello***
Author: ***Annabelle Theobald***

Publication date
October 23, 2025



© Chiara Schwarz

CISPA researcher Tejumade Àfọ̀njá co-authored a new international study that uses food as a starting point to reveal significant cultural blind spots in today's AI systems. The study also introduces a new participatory research approach to create more inclusive datasets and evaluate biases in AI models. The paper "The World Wide Recipe: A Community-Centred Framework for Fine-Grained Data Collection and Regional Bias Operationalisation" was presented at the ACM Conference on Fairness, Accountability, and Transparency (FAccT '25) in Athens in June 2025 and won the Best Paper Honorable Mention award.

World Wide Dishes:Using Food to Uncover AI's Cultural Blind Spots

Ṭejúmádé Àfọ̀njá

"Food is an important gateway to culture," explains CISPA researcher Ṭejúmádé Àfọ̀njá, a PhD student on the team of CISPA-Faculty Dr. Mario Fritz. "We wanted to explore how generative AI represents people's food cultures in generated images." Behind it was the desire to explore possible cultural biases in AI models. "The lead project coordinator of our paper, Siobhan Mackenzie Hall, had found in previous studies that many models are biased in one way or another," Àfọ̀njá continues. "We asked ourselves which lens we could use to look at this problem. Food turned out to be a good one, because it's a universal language." Specifically, the team investigated how certain dishes are depicted in AI-generated images. To do this, a new reference dataset was developed in a first step, which was then used to test existing models in a second step.

A New Dataset Featuring Dishes from Around the World

The author team chose a community-oriented research approach and called it the World Wide Recipe. People from all over the world were invited to contribute their knowledge. "We wanted to give people a say in how their cultures are represented in AI systems," says Àfọ̀njá. As a first case study, they built the World Wide Dishes (WWD) dataset. It is a collection of 765 dishes from 106 countries, described in 131 local languages. Each entry was contributed directly by community participants all over the world, who shared the cultural, linguistic, and culinary context behind each dish and provided images. "We compared WWD with existing datasets whose data had been collected from the internet," explains Àfọ̀njá. "More than half of the dishes in the dataset don't appear there, which gives it its unique character." The dataset and all code have been released under an open license to encourage transparency and collaboration.

Misrepresentation in Existing Models

In a second step, Àfọ̀njá and her colleagues used WWD to compare the images of the dishes included in it with AI-generated images of those dishes. The comparative analysis was again carried out by members of the communities. "We found that many of the models were

stereotypical in their outputs. For example, when we prompted the models to generate an image of a Nigerian dish like Amala, the results were often unappealing or inaccurate," explains Àfọ̀njá. "In contrast, when we asked for something like a hot dog from the United States, the generated images were much closer to the real thing." This applied to all models tested: DALL·E 2, DALL·E 3, and Stable Diffusion. "The image quality was generally poor, and there were clear misrepresentations of the culture," she continues. "The reason is that many models are trained on internet data, and if foods from certain regions aren't represented online, those regions will be overlooked."

»It's not enough to design a model in Silicon Valley or Germany and expect it to work everywhere. Collecting more data is key—but it has to be done in collaboration with communities, not just extracting data from them.«

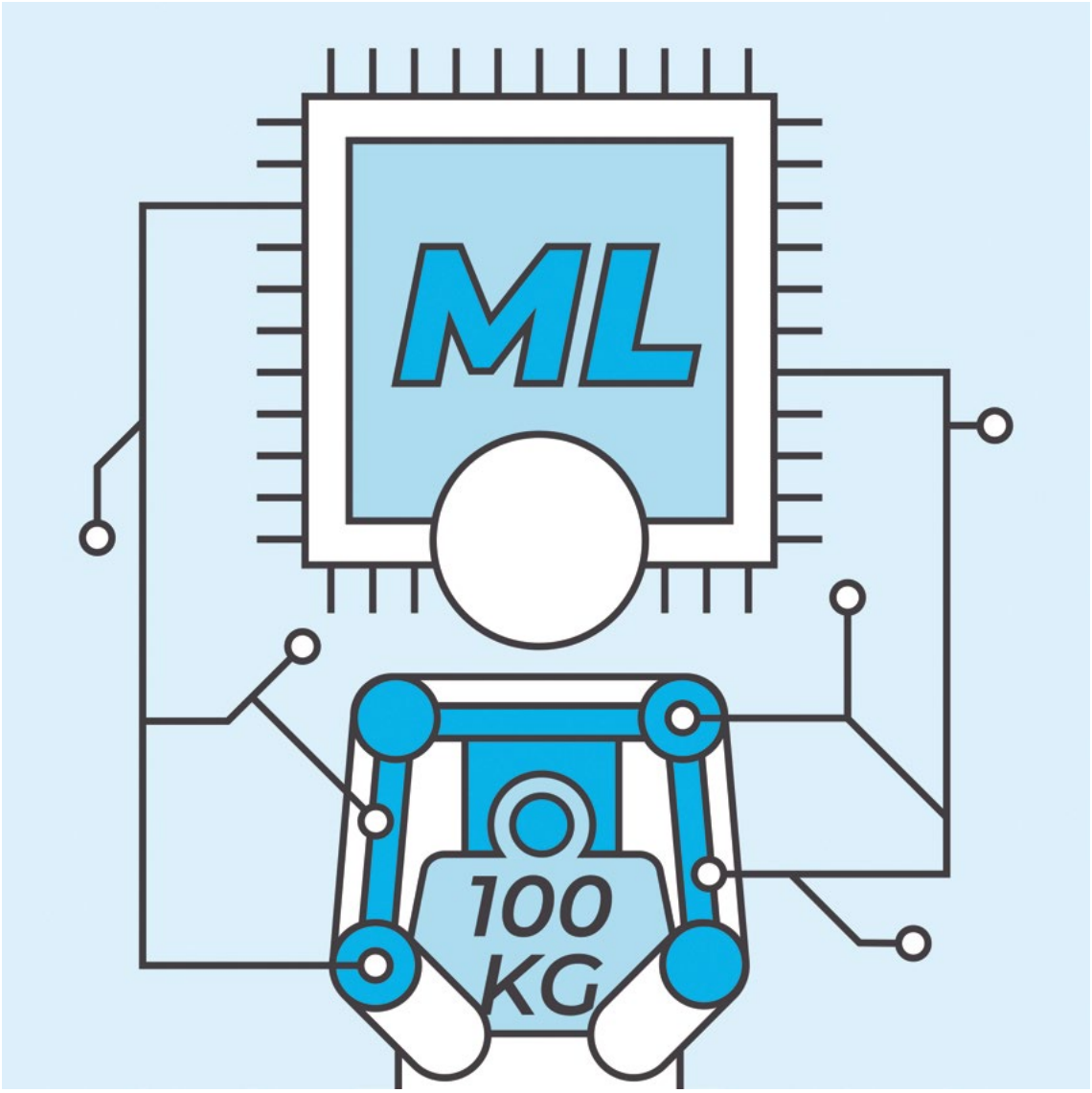
A Global Tool Requires Global Input

Àfọ̀njá and her colleagues conclude from this finding that the companies behind the models need to invest more in long-tail training and data collection for large language models. “Our argument is that the companies must prioritize all regions if they want to build models that truly represent global culture,” says Àfọ̀njá. “It’s not enough to design a model in Silicon Valley or Germany and expect it to work everywhere. Collecting more data is key—but it has to be done in collaboration with communities, not just extracting data from them.” An important keyword here is ownership of the data. “When you collect data from communities, the question is always: Who owns it—the community or the organization that funded the data collection?” says the CISPA researcher.

Data Collection and the Fight Against Cultural Bias

Àfọ̀njá would love to scale World Wide Dishes, but it’s very expensive. Until now, the whole project was entirely volunteer-driven. “None of the contributors were paid,” she explains. “With proper funding, we could pay community contributors to collect even more local data—asking families for recipes that aren’t online, for instance. That kind of data is invaluable but costly to obtain.” Because the method of data collection was so important for the project, another paper was produced as a follow-up product. “We published a paper called ‘The Human Labour of Data Work,’ which documents how we collected the dataset and the challenges involved. It focuses on the human effort, cultural trust, and lessons for anyone building similar datasets in the future.” Anyone who listens to Àfọ̀njá knows that this issue is close to her heart and that she will continue to campaign for AI models to lose their cultural bias.

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



Exoskeletons are often described as a technology of the future, yet they already ease physical strain for workers in logistics and manufacturing today. Still, wearable assistive devices do not always meet the complex demands of real-world application. CISPA researcher Julian Rodemann, together with colleagues from LMU Munich and the Biodesign Lab at Harvard University, is working to change that, using a machine-learning approach that not only identifies optimal assistance settings but also explains why it recommends a particular configuration. He presented his paper "Explaining Bayesian Optimization by Shapley Values Facilitates Human-AI Collaboration For Exosuit Personalization" at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD).

Explainable AI Makes Exoskeletons Understandable—and Ready for Everyday Use

Julian Rodemann

"Exoskeletons are already being used—but mostly in clearly defined work environments. Typically, they support repetitive movements in logistics or production and are preconfigured precisely for those tasks," says Julian Rodemann. Soft exosuits—the lighter variant—work similarly: They are usually configured for specific tasks, such as repetitive lifting or sorting. They work less well when users need to alternate between different activities. According to Philipp Arens, a doctoral researcher at the Harvard John A. Paulson School of Engineering and Applied Sciences, this is the core challenge: "A key challenge lies in making exosuits work seamlessly across a variety of users and tasks. From a design perspective, devices can be made more lightweight and less obtrusive—but the real challenge is determining how much and at what moment in time assistance should be provided. That varies individually. This is why user feedback within the assistance loop becomes so important."

Why Optimal Settings Are So Hard to Find

To determine which settings are optimal for which person, the researchers rely on computational support. "Real-world test sessions often tend to be lengthy," explains Rodemann. "Participants perform different movements while my colleagues at the Biodesign Lab at Harvard University continuously record physiological data. As a result, the number of testable combinations is heavily limited." The team therefore uses Bayesian optimization—a machine-learning technique that identifies optimal configurations efficiently by targeted sampling and progressively reducing uncertainty.

Why the AI Tests Not Only What Works Best—But Also What It Doesn't Know Yet

In each iteration, the algorithm does not simply select the setting that appears to be the best. It must also explore unknown regions of the parameter space—even if a given configuration is temporarily less comfortable for the user. "We distinguish between exploitation, meaning the use of existing knowledge, and exploration, the targeted testing needed to close knowledge gaps," Rodemann explains. This balance is crucial for adaptive exosuits. "Many optimization methods operate as black boxes. Users don't know what configuration is being proposed—or why. If we

can clearly and granularly explain what the system plans to do next and for what reason, trust increases. Users can also judge for themselves which regions of the parameter space do not make sense—and we avoid unnecessary tests," says Arens.

»Human-in-the-loop approaches are very powerful, but they are labor-intensive and can be demanding for participants. A promising next step is to develop groupspecific starting points—'warm starts'—based on typical user profiles.«

ShapleyBO: A Method That Reveals Why the AI Makes Certain Decisions

To achieve this transparency, Rodemann and colleagues developed ShapleyBO—a method that makes optimization decisions interpretable. “Until now, humans only saw the outcome—for example, that ‘Assistance Level 7 when bending and Level 3 when lifting’ has been selected. With ShapleyBO, we show which parameters contributed to the recommendation. We also explain whether the suggestion results from optimization or from deliberate exploration of new settings. This allows users to judge whether the suggestion truly makes sense in their situation and intervene if necessary,” says Rodemann. The previously abstract balance of exploration and exploitation thus becomes visible to users.

More Studies Needed to Enable Effective Knowledge Sharing Between Humans and AI

“The algorithm knows patterns from many user datasets, while the human knows their immediate situation best. The question is how to combine both efficiently,” says Rodemann. The method is still under development and has so far been tested using simulation data from a real soft exosuit. Next, user studies will investigate how people interact with explainable optimization suggestions. “Human-in-the-loop approaches are very powerful, but they are labor-intensive and can be demanding for participants. A promising next step is to develop group-specific starting points—‘warm starts’—based on typical user profiles. This could accelerate optimization or, in some cases, even eliminate the need for it entirely,” Philipp Arens explains. Rodemann sees a fundamental advancement in this: “Our approach has the potential not only to improve the personalization of exosuits but also to strengthen trust in AI-based assistance systems overall.”

Funding acknowledgements on page 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

FUNDING ACKNOWLEDGEMENTS

LLM-Based Web Application Scanner Recognizes Tasks and Workflows **14**

The research described in this article was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project titled “Semantische Modelle und Agenten für die Verwundbarkeitsprüfung von Web-Anwendungen” with project number 452850842.

Open-Source Fuzzer with Evolutionary Algorithm Produces Customized Test Inputs **38**

The research described in this article was funded by the European Union (ERC S3, 101093186). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

A New Method Can Detect Whether Copyright-protected Images Were Used to Train AI Models **46**

The research described in this article was supported by the German Research Foundation (DFG) within the framework of the Weave Programme under the project titled “Protecting Creativity: On the Way to Safe Generative Models” with number 545047250. This research was also supported by the Polish National Science Centre (NCN) grant no. 2023/51/I/ST6/02854 and 2020/39/O/ST6/01478 and by the Warsaw University of Technology within the Excellence Initiative Research University (IDUB) programme.

How Agile is Your Crypto? Interview Study Explores Cryptographic Update Processes **54**

The research described in this article was funded by VolkswagenStiftung Niedersächsisches Vorab (ZN3695). Any findings and opinions expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

The research described in this article was partially funded by Image-Tox (ZT-I-PF-4-037), supported by the impulse and networking fund of the Helmholtz Association. The views and opinions expressed are those of the authors only and do not necessarily reflect those of the funding agencies, which can neither be held responsible for them. It was also partially funded by ELSA – European Lighthouse on Secure and Safe AI (grant agreement No. 101070617). The project on which this report is based was funded by the German Federal Ministry of Education and Research (funding code 16KIS2012).

The research described in this article is funded by the Helmholtz Association and through the Helmholtz Impulse and Networking Fund.

The research described in this article was supported by the Federal Statistical Office of Germany within the co-operation project “Machine Learning in Official Statistics.” JR further acknowledges support by the Bavarian Academy of Sciences (BAS) through the Bavarian Institute for Digital Transformation (bidt) and by the LMU mentoring program of the Faculty for Mathematics, Informatics, and Statistics. YS is supported by the DAAD program Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

ABOUT CISPA

The CISPA Helmholtz Center for Information Security is a national Big Science institution within the Helmholtz Association of German Research Centers. CISPA researchers explore all aspects of information security. They conduct cutting-edge foundational research as well as application-oriented research, addressing the most pressing challenges in cybersecurity, artifical intelligence and privacy. Research results achieved at CISPA find their way into industrial applications and products that are available worldwide. CISPA thus contributes to German as well as European competitiveness.

CISPA offers a world-class research environment as well as extensive resources to a large number of researchers. It strongly supports the undergraduate and graduate education of cybersecurity students and seeks to become an elite training ground for the next generation of cybersecurity experts and leading scientists in this field. CISPA is located in Saarbrücken and St. Ingbert. The center's proximity to France and Luxembourg puts it in an ideal position for cross-border cooperation with other research institutions.

Our research currently focuses on the following six research areas:

Algorithmic Foundations and Cryptography

Trustworthy Information Processing

Reliable Security Guarantees

Threat Detection and Defenses

Secure Connected and Mobile Systems

Empirical and Behavioral Security

IMPRINT

CISPA – Helmholtz Center for Information Security gGmbH
Stuhlsatzenhaus 5
66123 Saarbrücken, Germany

Publisher

Sebastian Klöckner

Editor-in-Chief

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

Editors

Stephanie Bremerich,
Alexandra Goweiler,
Janine Paulus,
Chiara Schwarz

Illustration

Alexandra Goweiler,
Janine Paulus,
Chiara Schwarz,

Design

Tobias Ebelshäuser,
David Rohner

Photography

February 2026

Information as of

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

Contact Corporate Communications

Digital Fingerprint: CSS Opens New Possibilities for User Tracking

LLM-Based Web Application Scanner Recognizes Tasks and Workflows

The Underestimated Risk: Why Website Owners Often Neglect Security Updates in WordPress

Security is Just a Side Quest: Insights From the Video Game Industry

The Power of Words: How Wording Influences Consent Behavior in App Permission Requests

Unequal Internet: Differences Between Websites from Industrialized and Emerging Countries

Cybersecurity Practices of People with Low Socioeconomic Status in Pakistan

Open-Source Fuzzer with Evolutionary Algorithm Produces Customized Test Inputs

Fuzzing Reloaded: Targeted Manipulation for Enhanced Security on the Web

A New Method Can Detect Whether Copyright-protected Images Were Used to Train AI Models

C++ Coroutines: Prone to Code-reuse Attack despite CFI

How Agile is Your Crypto? Interview Study Explores Cryptographic Update Processes

AI Accelerates Drug Discovery Through the Automatic Analysis of Zebrafish Embryos

From Black Box to Glass Box: AI Explainability in Stroke Treatment

How Blind and Low-Vision Users Manage Their Passwords

World Wide Dishes: Using Food to Uncover AI's Cultural Blind Spots

Explainable AI Makes Exoskeletons Understandable—and Ready for Everyday Use

CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY