idw - Informationsdienst Wissenschaft
Nachrichten, Termine, Experten

Press release

CISPA Helmholtz Center for Information Security
Eva Michely
02/21/2025
http://idw-online.de/en/news847895

Research results
Information technology
transregional, national

# LLM-basierter Scanner für Webanwendungen erkennt Tasks und Workflows

**Ein automatisierter Scanner für Webanwendungen kann Tasks und Workflows in Webanwendungen selbstständig erkennen und ausführen. Das Tool namens YuraScanner nutzt das Weltwissen in Large Language Models (LLMs), um sich wie menschliche Nutzende durch Webanwendungen zu navigieren. Es kann Tasks auf kohärente Weise bearbeiten und dabei die korrekte Abfolge von Arbeitsschritten ausführen, wie sie beispielsweise ein Online-Shop erfordert. YuraScanner wurde auf 20 Webanwendungen getestet und hat dabei zwölf bislang unbekannte Cross-Site-Scripting-Schwachstellen aufgedeckt. Die Methode hinter YuraScanner sowie das Tool selbst wurde am CISPA Helmholtz-Zentrum für Informationssicherheit entwickelt.**

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

## Mithilfe von LLMs navigiert YuraScanner durch Webanwendungen

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

## Automatisiertes Scannen auf tieferliegenden Ebenen der Webanwendung

Unter der Anleitung eines LLM arbeitet YuraScanner task-orientiert, wodurch er in die tieferen Ebenen der zu testenden Webanwendung vordringt. Er kann die in der Webanwendung vorgesehenen Tasks nicht nur erkennen, sondern sie auch gezielt ausführen. Dabei beachtet er die Abfolge von Schritten, die zum Erledigen des jeweiligen Tasks erforderlich ist.

Stafeev erläutert: „Normalerweise unterscheiden Testing-Tools nicht zwischen verschiedenen Arten von Buttons, sondern klicken einfach auf alles, was verfügbar ist. Der größte Nachteil dabei ist, dass bei einem sehr spezifischen mehrschrittigen Workflow, wie zum Beispiel in einem Online-Shop, bei dem man einen Artikel in den Warenkorb legen, zur Kasse gehen und ein Formular ausfüllen muss, die Wahrscheinlichkeit sehr gering ist, dass ein einfacher Webcrawler das erfolgreich erledigen kann.“ Mit YuraScanner haben Stafeev und seine Kollegen gezeigt, dass LLMs für Web-Sicherheitsscans eingesetzt werden können und damit den Weg für weitere Forschung auf diesem Gebiet geebnet. Ihre Forschung zum YuraScanner stellen sie auf dem Network and Distributed System Security Symposium (NDSS) 2025 vor, das vom 24. bis 28. Februar 2025 in San Diego, Kalifornien, stattfindet.

Um weitere Forschung anzuregen, ist der Quellcode von YuraScanner auf GitHub veröffentlicht worden: https://github.com/pixelindigo/yurascanner/tree/ndss25

contact for scientific information:

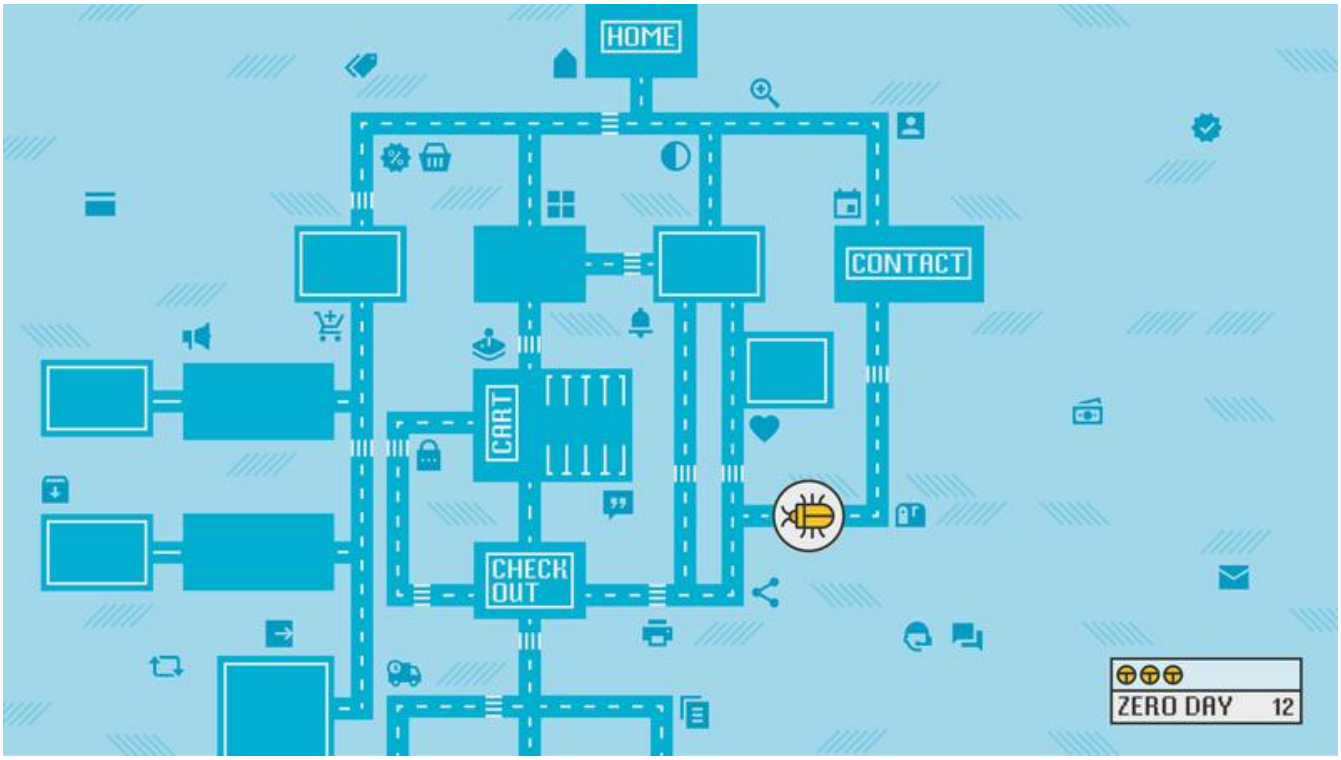
Aleksei Stafeev und Dr. Giancarlo Pellegrino
CISPA Helmholtz-Zentrum für Informationssicherheit
Stuhlsatzenhaus 5
66123 Saarbrücken
aleksei.stafeev@cispa.de
pellegrino@cispa.de

Original publication:

Aleksei Stafeev, Tim Recktenwald, Gianluca De Stefano, Soheil Khodayari, and Giancarlo Pellegrino. 2024. YuraScanner: Leveraging LLMs for Task-driven Web App Scanning. DOI: https://doi.org/10.14722/ndss.2025.240388

URL for press release: https://github.com/pixelindigo/yurascanner/tree/ndss25 - um weitere Forschung anzuregen, ist der Quellcode von YuraScanner auf GitHub veröffentlicht worden

URL for press release: https://cispa.de/en/tldr-episode-37 - In der neuen Folge des “CISPA TL;DR”-Podcasts geht es ebenfalls um YuraScanner




YuraScanner untersucht tieferliegende Ebenen der Webanwendung
CISPA