Innovationswettbewerb „Sicherheit mit und für KI“ Baden-Württemberg (2023)

Projektsteckbrief

KIPHI: Intelligenter KI-Stack zur Erkennung von Phishing-Angriffen

Baden-Württemberg
Ministerium für Wirtschaft,
Arbeit und Tourismus

Worum geht es: 91% aller Cyberattacken beginnen mit einer Phishing-E-Mail. Vor allem Weiterentwicklungen im Bereich KI bringen dabei die Bedrohungen durch Phishing-Kampagnen auf eine neue, gefährliche Ebene. Hier setzt das Projekt KIPHI der Inlyse GmbH an. Während herkömmliche Spam-Filter beim Herausfiltern von Phishing-Emails versagen, sollen mit Hilfe einer KI-basierten Lösung fortschrittliche Phishing-Kampagnen identifiziert und geblockt werden. Hierzu sollen neben der Überwachung verdächtiger Aktivitäten auch das Nutzerverhalten miteinzogen werden.

Projektkonsortium: inlyse GmbH

Innovationswettbewerb „Sicherheit mit und für KI“ Baden-Württemberg

In einer zunehmend digital vernetzten und von Künstlicher Intelligenz (KI) beeinflussten Welt wird die Sicherheit und die Vertrauenswürdigkeit von Produkten und Dienstleistungen immer wichtiger.

Um die Entwicklung sicherer und vertrauenswürdiger KI-Produkte und KI-Dienstleistungen zu forcieren, hat das Wirtschaftsministerium Baden-Württemberg den Innovationswettbewerb „Sicherheit mit und für KI“ ausgeschrieben. Mit dem Wettbewerb werden Unternehmen in Baden-Württemberg in der Entwicklung innovativer Sicherheitslösungen gefördert, bei der KI-Technologien zum Einsatz kommen oder die dazu dienen KI-Anwendungen sicherer zu machen. Sicherheit umfasst hierbei die drei Dimensionen Security (Cybersicherheit), Safety (Betriebssicherheit) und Privacy (Datenschutz).

Der Innovationswettbewerb „Sicherheit mit und für KI“ ist eine Maßnahme im Rahmen des Aktionsprogramms „KI für den Mittelstand“ des Ministeriums.

KI für mehr Schutz vor Phishing und Spear-Phishing

Während herkömmliche Spamfilter bereits heute einen Großteil der unerwünschten E-Mails herausfiltern können, versagen sie oft bei Phishing und insbesondere gezielten Spear-Phishing-Angriffen auf ganz bestimmte Personen oder Personengruppen. Insgesamt sind bereits heute Phishing und Spear-Phishing-Angriffe das Einfallstor von 91% aller Cyberattacken. Eine Entwicklung, die sich durch Innovationen im Bereich KI weiter verschärft. Mit effektiven Natural Language Processing Tools (NLP) wie ChatGPT, die die Verarbeitung und Analyse von natürlicher Sprache durch Maschinen ermöglichen, gelingt es Angreifern immer leichter, überzeugende und grammatikalisch korrekte Texte zu verfassen.

Mittels maschinellen Lernmodellen und Tools aus dem Bereich des NLP soll im Rahmen des Projekt KIPHI ein KI-basiertes Filtersystem entwickelt werden, welches Schutz gegen 99% aller Spam-Nachrichten bietet. Ein besonderes Augenmerk liegt dabei auf dem Schutz vor besonders gefährlichen und immer weiter verbreiteten Spear-Phishing-Angriffe. Insgesamt soll die intuitiv bedienbare KI-Anwendung mittels eines Plug-in leicht in die bestehende IT-Infrastruktur integriert werden können und zunächst für deutsch- und englischsprachige E-Mails zu Verfügung stehen.

Entwicklung eines KI-basierten Filtersystems für ein neues Sicherheitslevel

Das entwickelte KI-basierte Filtersystem gegen Phishing und insbesondere Spear-Phishing-Attacken umfasst insgesamt 5 Filterstufen:

1) Vorprüfung: Identifikation und Aussortierung generische Spam-Mails.
2) Kopfzeilenanalyse: Erkennung von Spam-Mails an Hand von Metadaten.
3) Inhaltsanalyse: Textbasierte Identifizierung von Phishing-Mails.
4) Kontextuelle Inhaltsanalyse: Erkennung von Phishing- und Spear-Phishing-Nachrichten durch sprachlich-kontextuelle Analysen.
5) Verhaltensanalyse: Profilbasierte Erkennung von Phishing-Nachrichten anhand historischer Kommunikationsmuster.

Die Entwicklung umfasst zum einen CatBoost-Algorithmen, also eine Open-Source-Implementierung, welche auf Grundlage heterogener Datenquellen Voraussagen von hoher Qualität treffen kann. Diese kommen in Stufe 2 der Kopfzeilenanalyse und Stufe 3 der Inhaltsanalyse zum Einsatz. Für die Kontextuelle Inhaltsanalyse in Stufe 4 werden dann zum anderen Long Short-Term Memory Networks (LSTM) verwendet. Diese Form von neuronalem Netzwerk ist in der Lage, Informationen länger zu speichern und komplexe Beziehungen in diesen Sequenzdaten zu erfassen. Dadurch können durch den Einsatz von LSTM Texte besser erfasst und zuverlässige Vorhersagen getroffen werden, als dies bei anderen neuronalen Netzwerken der Fall ist. Abschließend wird die E-Mail mit Hilfe eines Klassifikationsensemble – also der Kombination verschiedener Modelle –abschließend bewertet.

Insgesamt soll mit dem KI-basierten Filtersystem ein bisher unerreichtes Sicherheitslevel gegen Phishing und Spearphishing – Angriffe erreicht werden. Perspektivisch kann dieses dann auch als Grundlage für weitere KI-basierte IT-Sicherheitssysteme dienen.

Kontakt

Inlyse GmbH
Christian Boll
Haid-und Neu-Straße 18
76131 Karlsruhe
E-Mail: info@inlyse.com

Gefördert durch

Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg
Postfach 10 01 41
Schlossplatz 4 (Neues Schloss)
70001 Stuttgart
Tel: 0711 123-2869
Fax: 0711 123-2871
pressestelle@wm.bwl.de
www.wm.baden-wuerttemberg.de

Weitere Informationen

https://www.inlyse.com

Quellenhinweis

Weitere Informationen zum Innovationswettbewerb finden Sie unter:

www.wirtschaft-digital-bw.de

Baden-Württemberg
Ministerium für Wirtschaft,
Arbeit und Tourismus

Initiative Wirtschaft 4.0 BW