ChatGPT als nützlicher Cybersecurity-Co-Pilot

Das KI-Modell kann bösartige Aktivitäten in XDR-Telemetriedaten leichter filtern, Spam-Filter verbessern und die Analyse von „Living Off the Land Binaries" – kurz „LOLBins" – vereinfachen.

Sophos hat einen neuen Report veröffentlicht. Thema ist das GPT-3-Sprachmodell, das hinter dem bekannten ChatGPT-Framework steht, und wie die Cybersecurity-Branche das Modell für die Abwehr von Angreifern nutzen kann. Der aktuelle Report „GPT for You and Me: Applying AI Language Processing to Cyber Defenses" beschreibt Projekte, die von Sophos X-Ops entwickelt wurden und die umfangreichen Sprachmodelle von GPT-3 nutzen. Ziel ist die vereinfachte Suche nach bösartigen Aktivitäten in Datensätzen von Sicherheitssoftware, das genauere und schnellere Filtern von Spam sowie die schnellere Analyse von Binär-Attacken (LOLBin).

„Seit der Vorstellung von ChatGPT durch OpenAI im November 2022 hat sich die Sicherheitsbranche weitgehend auf die potenziellen Risiken konzentriert, die diese neue Technologie mit sich bringen könnte. Kann die KI den Möchtegern-Angreifern beim Schreiben von Malware oder Cyberkriminellen beim Verfassen überzeugenderer Phishing-E-Mails helfen? Vielleicht, aber wir bei Sophos sehen KI seit jeher als Verbündeten und nicht als Feind für die Verteidigung, was sie zu einer Eckpfeilertechnologie für Sophos macht, und das gilt auch für GPT-3. Die Sicherheitsbranche sollte nicht nur auf die potenziellen Risiken der Technologie achten, sondern auch auf die möglichen Chancen", sagt Sean Gallagher, Principal Threat Researcher bei Sophos.

Die Forscher von Sophos X-Ops arbeiten an drei Prototyp-Projekten, die das Potenzial von GPT-3 als Assistent für Cybersecurity-Verteidiger demonstrieren. Alle drei Projekte nutzen eine Technik namens „few-shot learning", um das KI-Modell mit nur wenigen Datenproben zu trainieren und so die Notwendigkeit zu verringern, eine große Menge an vorklassifizierten Daten zu sammeln.

Die erste Anwendung, die Sophos mit der „few-shot learning"-Methode getestet hat, war ein Natural Language Query Interface zum Durchsuchen bösartiger Aktivitäten in der Telemetrie von Sicherheitssoftware. Sophos hat das Modell insbesondere mit seiner Endpoint Detection and Response-Lösung geprüft. Mit dieser Schnittstelle können Verteidiger die Telemetrie mit einfachen englischen Befehlen filtern, ohne SQL oder die zugrunde liegende Struktur einer Datenbank verstehen zu müssen.

Als nächstes testete Sophos einen neuen Spam-Filter mit ChatGPT und stellte fest, dass der Filter mit GPT-3 im Vergleich zu anderen maschinellen Lernmodellen für die Spam-Filterung deutlich genauer war. Schließlich konnten die Forscher von Sophos ein Programm erstellen, das das Reverse-Engineering der Befehlszeilen von LOLBins vereinfacht. Ein solches Reverse-Engineering ist bekanntermaßen schwierig, aber auch entscheidend, um das Verhalten von LOLBins zu verstehen und diese Art von Angriffen in Zukunft zu unterbinden.

„Eine der wachsenden Sorgen in SOCs (Security Operation Center) ist die schiere Menge an ‚Lärm', die hereinkommt. Es gibt einfach zu viele Meldungen und Erkennungen, die sortiert werden müssen, und viele Unternehmen haben mit begrenzten Ressourcen zu kämpfen. Wir haben bewiesen, dass wir mit GPT-3 bestimmte arbeitsintensive Prozesse vereinfachen und den Verteidigern wertvolle Zeit zurückgeben können. Wir arbeiten bereits daran, einige der oben genannten Prototypen in unsere Produkte zu integrieren und haben die Ergebnisse

unserer Bemühungen auf unserem GitHub für diejenigen bereitgestellt, die GPT-3 in ihren eigenen Analyseumgebungen testen möchten. Wir glauben, dass GPT-3 in Zukunft sehr wohl ein Co-Pilot für Sicherheitsexperten werden kann", so Gallagher.

Social Media von Sophos für die Presse
Wir haben speziell für Sie als Journalist*in unsere Social-Media-Kanäle angepasst und aufgebaut. Hier tauschen wir uns gerne mit Ihnen aus. Wir bieten Ihnen Statements, Beiträge und Meinungen zu aktuellen Themen und natürlich den direkten Kontakt zu den Sophos Security-Spezialisten.
Folgen Sie uns auf und

LinkedIn: https://www.linkedin.com/groups/9054356/
Twitter: @sophos_info

Pressekontakt:

Sophos
Jörg Schindler, PR-Manager Central & Eastern Europe
joerg.schindler@sophos.com, +49-721-25516-263

TC Communications
Arno Lücht, +49-8081-954619
Thilo Christ, +49-8081-954617
Ulrike Masztalerz, +49-30-55248198
Ariane Wendt +49-172-4536839
sophos@tc-communications.de