Die dunkle Seite der Macht: KI in der Cyberkriminalität

Sophos X-Ops veröffentlicht zwei Forschungsberichte, die zum einen den aktuellen Einsatz von KI für Attacken untersucht und zum anderen die Einstellung von Cyberkriminellen zu Künstlicher Intelligenz anhand der Untersuchung von Dark-Web-Foren analysiert.

„Wir haben im Dark Web zahlreiche Posts über die möglichen negativen Auswirkungen von KI auf die Gesellschaft und die ethischen Implikationen ihres Einsatzes gefunden. Mit anderen Worten: Zumindest im Moment scheint es, dass Cyberkriminelle die gleichen Debatten in Sachen Künstliche Intelligenz führen wie der Rest von uns", so Christopher Budd, Direktor X-Ops-Forschung bei Sophos.

Sophos hat heute zwei Berichte über den Einsatz von KI in der Cyberkriminalität veröffentlicht. Der Report „The Dark Side of AI: Large-Scale Scam Campaigns Made Possible by Generative AI" untersucht anhand eines konkreten Fallbeispiels, wie Betrüger in Zukunft Technologien wie ChatGPT nutzen könnten, um mit minimalen technischen Fähigkeiten Betrugsattacken in großem Umfang durchzuführen. Der zweite Bericht „Cybercriminals Can't Agree on GPTs" stellt die Untersuchung verschiedener Dark-Web-Foren vor. Die Ergebnisse zeigen, dass einige Cyberkriminelle trotz des Potenzials der neuen Technologie skeptisch gegenüber dem aktuellen Einsatz von Chatbots und anderen KI-Technologien sind.

Die dunkle Seite der KI

Mit einer einfachen E-Commerce-Vorlage und großen Sprachmodell-Tools wie GPT-4 konnte Sophos X-Ops eine voll funktionsfähige Website mit KI-generierten Bildern, Audio- und Produktbeschreibungen sowie eine gefälschte Facebook-Anmeldeseite und eine gefälschte Checkout-Seite erstellen, um persönliche Anmelde- und Kreditkartendaten abzuschöpfen. Für die Erstellung und den Betrieb der Website waren nur minimale technische Kenntnisse erforderlich. Mit demselben Tool war es zudem möglich, innerhalb von Minuten mit nur einem Knopfdruck Hunderte ähnlicher Websites zu erstellen.

„Es ist selbstverständlich und wird entsprechend von uns erwartet, dass Kriminelle zur Automatisierung ihrer Machenschaften auf neue Technologien zurückgreifen, um so effektiv wie möglich agieren zu können", so Ben Gelman, Senior Data Scientist bei Sophos. „Die ursprüngliche Erstellung von Spam-E-Mails war ein entscheidender Schritt in der Betrugshistorie, da er das Ausmaß des kriminellen Spielfelds maßgeblich veränderte. Die aktuelle Entwicklung in Sachen Künstlicher Intelligenz hat ein ähnliches Potential: Sobald es eine KI-Technologie gibt, die vollständige, automatisierte Bedrohungen erzeugen kann, wird sie zum Einsatz kommen. Die aktuell zu beobachtende Integration generativer KI-Elemente in klassische Betrügereien, etwa durch KI-generierte Texte oder Fotos, um Opfer anzulocken, bildet nur den Anfang."

Über die Absichten des aktuellen Forschungsprojekts sagt Gelman: „Wir führen das aktuelle Projekt unter anderem deshalb durch, um den Kriminellen einen Schritt voraus zu sein. Durch die Schaffung eines Systems zur groß angelegten Erstellung betrügerischer Websites, das fortschrittlicher ist als die von Kriminellen derzeit verwendeten Tools, haben wir die einzigartige Möglichkeit, die Bedrohung zu analysieren und uns darauf vorzubereiten, bevor sie sich ausbreitet."

Cyberkriminelle stellen Potential von GPTs & Co. in Frage

Im zweiten Teil seiner KI-Forschungsoffensive untersuchte Sophos X-Ops die Einstellung von Angreifern gegenüber dem Einsatz von KI-Technologien und analysierte dazu vier prominente

Dark-Web-Foren hinsichtlich der auf große Sprachmodelle (Large Language Models, LLM) bezogene Diskussionen. Während der Einsatz von KI durch Cyberkriminelle offenbar noch in den Kinderschuhen steckt, diskutieren Bedrohungsakteure auf diesen Plattformen bereits intensiv über das Potenzial für Social Engineering. Ein Beispiel dafür ist die aktuelle „Pig Buchtering"-Welle mit Romance Scams.

Darüber hinaus stellte Sophos fest, dass sich die meisten Beiträge auf zum Verkauf stehende kompromittierte ChatGPT-Konten und „Jailbreaks" bezogen – Möglichkeiten also zur Umgehung der in LLMs integrierten Schutzmaßnahmen, sodass Cyberkriminelle die Tools für böswillige Zwecke missbrauchen können. Das Forscherteam fand außerdem zehn ChatGPT-Anwendungen, von denen die Entwickler behaupteten, dass sie für Cyberangriffe und die Entwicklung von Malware verwendet werden könnten. Allerdings wurde die Effektivität solcher Tools von Teilen der Dark-Web-Gemeinde stark angezweifelt und zum Teil sogar als Versuch gewertet, mit nutzlosen Programmen zu betrügen.

„Wir haben zwar einige Cyberkriminelle gesehen, die versuchten, mithilfe von LLMs Malware zu erstellen oder Tools anzugreifen, aber die Ergebnisse waren rudimentär und stießen bei anderen Benutzern oft auf Skepsis. Wir haben sogar zahlreiche Posts über die möglichen negativen Auswirkungen von KI auf die Gesellschaft und die ethischen Implikationen ihres Einsatzes gefunden. Mit anderen Worten: Zumindest im Moment scheint es, dass Cyberkriminelle die gleichen Debatten in Sachen Künstliche Intelligenz führen wie der Rest von uns", so Christopher Budd, Direktor X-Ops-Forschung bei Sophos.

Weitere Informationen zu KI-generierten Betrugswebsites und der Einstellung von Bedrohungsakteuren zu LLMs sind in den kompletten, englischen Berichten „The Dark Side of AI: Large-Scale Scam Campaigns Made Possible by Generative AI" und „Cybercriminals Can't Agree on GPTs" zu finden

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/
X/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