ChatGPT und Co.-KI als Gamechanger für Gut und Böse?

ChatGPT und Co.: KI als Gamechanger für Gut und Böse?

Seit Veröffentlichung des smarten Chatbots von OpenAI überschlagen sich die Meldungen, was man damit alles machen kann – aber auch die Warnungen, gerade in Sachen IT-Sicherheit.

Sicherheitsforscher von Check Point hatten kürzlich den ersten durch ChatGPT erstellten Angriffscode im Darkweb entdeckt (siehe auch iX 2/2023, Seite 20). Nun gibt es weitere Erkenntnisse: In Untergrundforen tauschten sich russische Kriminelle darüber aus, wie sie die von OpenAI für Russland vorgesehenen Beschränkungen – Kontrolle von IP-Adressen, Zahlkarten und Telefonnummern – umgehen können. Die Forscher stießen auf Nachfragen, wie man mit gestohlenen Zahlkarten an einen leistungsfähigeren OpenAI-Account gelangt, außerdem auf zahlreiche halblegale russische Online-SMS-Dienste, die Anleitungen zum Registrieren bei ChatGPT geben.

Auch Sicherheitsexperten von WithSecure loteten die kriminelle KI-Kompetenz aus und ließen das Sprachverarbeitungsmodell GPT-3 in verschiedenen Bereichen wie Fake News, Social Media und Phishingmails agieren. Voraussetzung für die in der Regel gut lesbaren und glaubwürdigen Texte ist laut Ergebnisbericht (siehe ix.de/zh1v) ein präzises Briefing der KI. Wie bei echten Kriminellen wird eine maßgeschneiderte Spear-Phishing-Mail umso glaubwürdiger, je mehr Details über den Mitarbeiter, das Unternehmen und den Kontext bekannt sind.

Einen Gamechanger nicht nur für die dunkle Seite sieht das Sicherheitsunternehmen Sophos: Auch für die Jagd nach Cyberkriminellen entfalten die neuen KI-Tools Potenzial. So

ließ sich GPT-3 per „Few-Shot Learning“ mit nur wenigen kommentierten Beispielen für Erkennung trainieren und wies nicht die Schwächen herkömmlicher maschineller Lernmodelle der Überanpassung und dadurch fehlender Verallgemeinerungen auf (Details siehe ix.de/zh1v). Einsetzen lässt sich das Werkzeug laut Sophos vor allem in der Spamerkennung und beim Reverse Engineering von Befehlszeilen, bei dem es eine Befehlszeile in eine verständliche Beschreibung übersetzen kann. In den Augen der Sophos-Forscher ist GPT-3 „ein Meilenstein für die Cybersicherheit“. (ur@ix.de)