Spuerkeess hat umstrittenes KI-Modell „Mythos“ auf dem Radar

Analyse. Das KI-Unternehmen Anthropic stellte zu Beginn des Jahres „Claude Mythos“ vor. In den falschen Händen kann das Modell zur Cyberwaffe für Kriminelle werden

Von Melanie Ptok

Luxemburgs Finanzwelt ist alarmiert. Ein neues KI-Modell macht die Runde, das für die Institute gefährlich werden könnte. Die Künstliche Intelligenz, die den luxemburgischen Banken aktuell Kopfzerbrechen bereitet, kommt von der US-Firma Anthropic. Der Name der KI: Claude Mythos.

Bei Mythos handelt es sich um ein KI-Modell mit ungewöhnlich starken Fähigkeiten bei der Erkennung und Ausnutzung von Sicherheitslücken. Laut Medienberichten birgt es eine „enorme Zerstörungskraft“. Das System ist darauf ausgelegt, Sicherheitslücken in Systemsoftwares zu erkennen, und kann diese ausnutzen, indem es Angriffsprogramme schreibt – sogenannte Exploits. Diese gehören zu den beliebtesten Werkzeugen von Betrügern und Cyberkriminellen.

Ein Sprecher der Spuerkeess teilt auf Anfrage mit, dass man die Entwicklungen rund um das KI-Modell „sehr aufmerksam“ verfolge. „Modelle wie ‚Mythos‘ stellen einen realen Fortschritt dar, der jedoch in der Kontinuität der jüngsten Entwicklungen im Bereich der Identifikation und Nutzung von Sicherheitsschwachstellen zu sehen ist“, so der Sprecher weiter.

Intern wurden bereits Expertenteams zusammengestellt. Wie man mit Mythos und derlei Modellen künftig umgehen will, steht aktuell auf der Sitzungsagenda, wie das „Luxemburger Wort“ erfuhr.

Die Spuerkeess beobachtet zudem auch andere Modelle mit vergleichbaren Potenzialen sehr genau. „Auf dieser Grundlage haben wir einen angemessenen Maßnahmenplan definiert, der sowohl die potenziellen Risiken für die Bank als auch die sich daraus ergebenden Chancen berücksichtigt.“

Zu diesem Plan zählen beispielsweise eine kontinuierliche Optimierung der Sicherheitsprozesse und ein regelmäßiger Austausch mit den zuständigen Aufsichtsbehörden, „um eine Vorgehensweise sicherzustellen, die den aktuellen Anforderungen im Finanzsektor entspricht“.

„Das Finanzministerium wird gegebenenfalls Maßnahmen zu einer koordinierten und verhältnismäßigen Reaktion auf diese potenzielle Bedrohung unterstützen“, sagt Finanzminister Gilles Roth. Foto: Anouk Antony/LW-Archiv

Zugang vorerst eingeschränkt

Als das US-Unternehmen Anthropic das neue KI-Modell im April vorstellte, war es sich der möglichen Konsequenzen bereits bewusst und hat den Algorithmus nicht für das breite Publikum freigegeben. Das Unternehmen erklärte vor wenigen Wochen, den Zugang zu Mythos einzuschränken, und warnte sogar selbst davor, dass es potenziell für Cyberangriffe genutzt werden könnte.

Deshalb lancierte Anthropic gemeinsam mit anderen großen Unternehmen, wie Apple, Microsoft, Google, Cisco und JPMorgan, „Project Glasswing“. „Um unsere eigenen Analysen weiter zu vertiefen“ sei die Spuerkeess in Bezug auf Mythos „im Austausch mit Akteuren, die am Projekt ‚Glasswing‘ beteiligt sind“. Bei dem Projekt geht es auch darum, die eigene Cyberabwehr zu testen.

Das neue KI-Modell „Mythos“ bringt manche eine Bank am Boulevard Royal zum Nachdenken. Foto: Marc Wilwert

Noch ist das KI-Modell Claude Mythos gar nicht freigegeben, schon zieht es die Aufmerksamkeit der weltweiten Finanzbranche auf sich. Foto: Shutterstock

> *Wir verfolgen die Entwicklungen um Mythos bei Spuerkeess sehr aufmerksam.*
>
> Spuerkeess

In den USA haben viele der größten Banken Zugang zu Mythos, von JPMorgan über Morgan Stanley bis hin zu Goldman Sachs. Auch sie haben teilweise interne Teams zusammengestellt, die sich mit der neuen KI beschäftigen. Die US-Institute arbeiten auch mit Geheimdiensten zusammen, wie Bloomberg berichtet. Der Medienkonzern zitiert JPMorgan-Chef Jamie Dimon mit den Worten: „Ich glaube, wir haben mittlerweile Hunderte Mitarbeiter, die sich vollzeitlich damit beschäftigen.“

Maßnahmen noch unklar

Auch in Europa versuchen Finanzinstitute, Zugang zu dem KI-Modell zu erhalten. Dies in der Hoffnung, den Kriminellen und deren Hackerversuchen zuvorzukommen und zu sehen, wie sich die Geldhäuser auf „Mythos“ vorbereiten können.

Jedoch: „Weder die CSSF noch eine andere europäische Finanzaufsichtsbehörde hat einen Zugang zu Mythos“, sagt eine Sprecherin der Finanzaufsichtsbehörde CSSF. Auch wenn die Cybersicherheit seit längerer Zeit eine Priorität der Aufsichtsbehörden auf europäischer und internationaler Ebene sei. „Mit Mythos wird diese noch einmal verstärkt.“

Anfragen bei anderen Banken gehen derweil ins Leere. Das ist verständlich: Das Thema ist komplex und niemand weiß so genau, was die Auswirkungen und Bedrohungen dieses mächtigen KI-Modells sein werden.

Auch Luxemburgs Finanzminister Gilles Roth (CSV) beobachtet die Entwicklungen rund um Mythos, wie er in seiner Antwort auf eine parlamentarische Frage des CSV-Abgeordneten Laurent Mosar schreibt. Roth verweist auf die EU-Kommission, die derzeit Gespräche rund um das KI-Modell führt. Denn „andere KI-Modelle, die ‚Mythos‘ ähneln, könnten grundsätzlich in Zukunft ähnliche Risiken bergen“.

Luxemburg werde weiterhin am Austausch auf EU-Ebene zu diesem Thema beitragen und „gegebenenfalls geeignete Maßnahmen zur Festlegung einer koordinierten und verhältnismäßigen Reaktion auf diese Art potenzieller Bedrohung unterstützen“. Wie diese Maßnahmen aussehen, bleibt offen.