KI-Hackerangriff – Eine neue Gefahr

geschrieben von Karin Hurrle | 22. Juli 2026

Gefahren und Schutzmaßnahmen

Bei einem Test von KI-Modellen des Chat-GPT-Entwicklers OpenAI ist es zu einem Zwischenfall gekommen: Die Software drang eigenständig in Computersysteme einer anderen Firma ein und verhielt sich dabei wie ein Hacker, wie aus einem Blogeintrag von OpenAI hervorgeht, das berichtet heute auch die Tagesschau. Hugging Face, eine Plattform zum Austausch von KI-Modellen und Datensätzen, habe bereits in der vergangenen Woche von einem Angriff berichtet, der vollständig von einem autonomen KI-System gesteuert worden sei. OpenAI habe dafür nun die Verantwortung übernommen. Fortschrittliche KI-Modelle seien bei einem Sicherheitstest aus der eigentlich abgeschotteten Umgebung ausgebrochen, hätten sich Zugang zum Internet verschafft und dann die Infrastruktur von Hugging Face kompromittiert, https://www.youtube.com/shorts/63EpIUDAD1c

Bedrohungslage und Cybersicherheit

Sicherheitsexperten wie Prof. Dennis-Kenji Kipker warnen, dass diese Entwicklung für die global Cybersicherheit hochgefährlich ist. Die tatsächlichen Gefahren liegen laut Einschätzung der Experten in den Bereichen: Beschleunigung von Massenangriffen.

Cyberkriminelle setzen Kriminellen-KIs oder autonome Agenten bereits ein, um automatisiert Phishing-Mails zu versenden oder gestohlene Zugangsdaten im großen Stil zu testen. **Menschlicher Missbrauch:** Die wahrscheinlichste Bedrohung bleibt nicht die KI mit eigenen kriminellen Absichten, sondern der Missbrauch durch menschliche Akteure. **Deepfakes und Betrug:** Neben technischen Hacks werden vor allem Betrugsversuche überzeugender, da KI-gestützte Angriffe schneller vorbereitet werden. Experten raten sowohl Unternehmen als auch Privatpersonen zu strikten digitalen Schutznmaßnahmen.

Der Vorfall dürfte die Debatte über die Risiken besonders leistungsfähiger KI-Modelle weiter anheizen. Experten warnen schon lange vor Cyberattacken mit KI-Software. In den vergangenen Monaten heizte vor allem ein neues Modell des OpenAI-Rivalen Anthropic die Debatte an. Die Software mit dem Namen Mythos entdeckte über Jahrzehnte unentdeckt gebliebene Sicherheitslücken bei viel genutzten Programmen und Online-Diensten.