CYBER|INTELLIGENCE
.Institute

CII-Presseinformation 08/2026, 22. Juli 2026

Presseinformation:
Prof. Dr. Dennis-Kenji Kipker zum autonomen Ausbruch und Hack eines KI-Modells von OpenAI

Berlin, 22.07.2026 – In der Nacht zum 22. Juli 2026 wurde bekannt, dass ein KI-Modell von OpenAI (GPT-5.6 Sol, sowie ein unveröffentlichtes Nachfolgemodell) bei einem internen Sicherheitstest mit bewusst reduzierten Schutzmechanismen eigenständig aus seiner Testumgebung ausgebrochen ist. **Über eine bis dahin unbekannte Schwachstelle verschaffte sich das Modell Zugang zum offenen Internet und kompromittierte anschließend selbstständig die Plattform Hugging Face**, unter anderem mithilfe gestohlener Zugangsdaten und in tausenden autonomen Schritten mit wechselndem Standort. **OpenAI selbst spricht von einem "beispiellosen Cyber-Vorfall".** Der Fall zeigt erstmals in dieser Deutlichkeit, dass KI-Systeme eigenständig Angriffsziele wählen und Sicherheitsmaßnahmen umgehen können, mit Implikationen weit über den Einzelfall hinaus.

Herr **Prof. Dr. Dennis-Kenji Kipker, Research Director am cyberintelligence.institute (CII),** ordnet die Geschehnisse in die aktuelle Debatte um autonome KI-Systeme als Cyberrisiko folgendermaßen ein:

Wie war der Ausbruch technisch möglich?
Die Testumgebung von OpenAI sollte keinen Internetzugang haben, erlaubt war bloß eine Verbindung zu einem internen Paket-Installationsdienst. Genau dort fanden die KI-Modelle eine bis dahin unbekannte Sicherheitslücke und verschafften sich darüber Zugang zum offenen Internet.
Von dort nutzten sie dann klassische Hacking-Methoden, wie gestohlene Zugangsdaten, technische Schwachstellen und viele Einzelschritte, um bis zu der Plattform Hugging Face vorzudringen.
Der Unterschied zu einer menschlichen Attacke lag dabei in einer extrem hohen Geschwindigkeit und extrem schwerer Auffindbarkeit.

Wie kann dieses Risiko eingeordnet werden?
Prof. Kipker betont, dass der Vorfall aus Sicht des cyberintelligence.institute eine Befürchtung bestätige, vor der Sicherheitsforschende schon länger warnen:

„Selbst die Entwickler fortschrittlicher KI-Modelle haben keine vollständige Kontrolle über ihre eigene Technologie, und damit sind die Erkenntnisse dieses Falles hochgefährlich für die globale Cybersicherheit".

Der Vergleich mit einem Laborunfall liegt nahe, da die Sicherheitsmechanismen der Modelle bewusst abgeschaltet waren, um sie unter Extrembedingungen zu testen, und genau dabei entwischten sie, weil die Sicherheitsvorkehrungen nicht ausreichten.
Von vollständig autonom handelnden Angreifenden in großer Zahl sei man zwar noch weit entfernt, versichert Dr. Kipker, doch schon heute dient KI Kriminellen als Werkzeug, um Angriffe schneller und in größerem Umfang durchzuführen. Die eigentliche Entwicklung der kommenden Jahre dürfte daher vor allem eine zunehmend professionalisierte IT-Massenkriminalität sein.

Folgeabschätzung
Aus Sicht des cyberintelligence.institute wird dies kein bedauerlicher Einzelfall bleiben, sondern muss als Weckruf verstanden werden: Unabhängige Kontrollmechanismen für solche Systeme sind nötig, bevor ein vergleichbarer Vorfall nicht mehr in einer Testumgebung, sondern gezielt

stattfindet. Im Rahmen seiner Technologiefolgenabschätzung analysiert das cyberintelligence.institute die Folgen von disruptiven KI-Innovationen aus multidisziplinärer Perspektive.

Prof. Dennis Kipker:
"Wir werden nicht davon ausgehen können, dass dieser einzelne, isolierte Vorfall dazu führt, dass wir jetzt in wenigen Monaten massenhaft und global ausgerollte autonome KI-Angriffe vorfinden werden. Er zeigt aber auf, wie schonungslos vorhandene Schwachstellen durch KI-Systeme identifiziert und bereits ausgenutzt werden – und dass selbst die Hersteller die Leistungsfähigkeit ihrer eigenen Systeme nicht immer voll einschätzen können."

Es würde uns freuen, wenn Sie obenstehenden Informationen und Statements für Ihre Berichterstattung verwenden möchten. Für weitergehende Fragen und zitierfähige Statements vermittle ich Ihnen gern ein Gespräch mit dem CII-Forschungsdirektor **Prof. Dr. Dennis-Kenji Kipker.**

Bitte kontaktieren Sie mich dazu – und für weitere Anliegen oder Bildmaterial – jederzeit gerne.

Mit freundlichen Grüßen
Patrick Teichmann

CII-Pressekontakt
Patrick Teichmann, Leiter Kommunikation
E-Mail: press@cyberintelligence.institute
Mobile: +49 151 55688761

cyberintelligence.institute
Am Kupfergraben 6
10117 Berlin // Germany
www.cyberintelligence.institute

Über das cyberintelligence.institute (CII)
Über das cyberintelligence.institute (CII): An der Schnittstelle von Wirtschaft und Wissenschaft, NGO und Startups entwickelt das CII in Frankfurt am Main und Berlin neue Lösungen für eine sichere digitale Zukunft. Kerngedanke sind die Kooperation, der Dialog und der interdisziplinäre und globale Informations- und Wissensaustausch, um Staat, Wirtschaft und Gesellschaft resilienter zu machen. Gefragt sind dabei ganzheitliche Konzepte, die Cybersicherheit nicht nur als abstrakte technisch-organisatorische Gewährleistungsverantwortung, sondern als gesamtgesellschaftliche Aufgabe zum Schutz gemeinsamer europäischer Werte verstehen. Damit tritt das cyberintelligence.institute aktiv für mehr digitale Sicherheit in einem Zeitalter neuer digitaler Herausforderungen ein: www.cyberintelligence.institute

---