

Liebe Mitglieder des Netzwerkes der ehemaligen SSAB, liebe Interessierte

Mit immer potenteren Tools, insbesondere der KI-Firma Anthropic, stellen sich weltweit dringliche Sicherheits- und damit auch Regulierungsfragen mit geopolitischer Brisanz.

Alarm mit Claude Mythos Preview

Das im April 2026 von Anthropic entwickelte [Claude Mythos Preview](#) kann bisher unbemerkte Sicherheitslücken in der Software entdecken. In der Hand von Hackern würde es sehr gefährlich. Anthropic stellte es deshalb nur rund 40 Tech-Firmen zur Verfügung, damit diese rasch wichtige Betriebssysteme überprüfen, Sicherheitslücken schliessen und sich so gegen künftige Cyberangriffe wappnen können. Der Anthropic-CEO Dario Amodei warnte, dass Anthropic in der dynamischen KI-Entwicklung höchstens einen mehrmonatigen Vorsprung habe, den es zu nutzen gilt. Andere Labs, insbesondere auch chinesische, könnten innerhalb eines Jahres ähnlich leistungsfähige Tools entwickeln. Tatsächlich stellte OpenAI, Konkurrent von Anthropic, kurz darauf mit [GPT-5.4-Cyber](#) ein vergleichbares Tool zur Verfügung, das aus Sicherheitsgründen ebenfalls nur einem eingeschränkten Kreis zugänglich ist.

KI-Regulierung: (verschobenes) Thema am Gipfeltreffen USA-China

Die KI-Entwicklung verändert Paradigmen. Erstmals erhalten nicht-staatliche Organisationen, Hacker und Kriminelle mit diesen Tools eine Wunderwaffe für verheerende Cyberattacken. Gegen diese wären auch die beiden Grossmächte, USA und China, nicht gefeit. Nur mit weltweit wirksamen Standards und Sicherheitskontrollen vor der Freigabe sehr potenter KI-Tools liesse sich diese Gefahr bannen. Tatsächlich waren Regulierungsfragen ein Thema am Gipfeltreffen der USA mit China in Beijing Mitte Mai 2026. US-Finanzminister Scott Bessent informierte am 14. Mai auf [CNBC](#), dass vereinbart wurde, gemeinsam ein entsprechendes Protokoll zu erarbeiten. Die USA seien derzeit im Lead und könnten deshalb vorangehen.

Claude Mythos 5 und Claude Fable 5

Die EU reagierte rasch und forderte [Zugang zu Claude Mythos Preview](#), insbesondere auch mit Blick auf die europäischen Banken und weltweiten Konsequenzen bei Angriffen. Auch [US-Kongressabgeordnete](#) verlangten einen erweiterten Zugang für Unternehmen. Mitte Mai [lockerte Anthropic die Regeln](#) und präsentierte am 9. Juni zwei neue Modelle: [Claude Mythos 5 und Claude Fable 5](#). Beide basieren auf Claude Mythos Preview. Das erstere steht, in verbesserter Version, weiterhin nur dem klar definierten Nutzungskreis zur Verfügung. Das zweite ist offen zugänglich, aber mit Sicherheitsvorkehrungen ausgestattet. Gefährliche Anfragen werden nicht beantwortet oder an ein älteres KI-Modell weitergeleitet.

Verdict der Trump-Regierung

Kaum stand Claude Fable 5 zur Verfügung, musste Anthropic gegenüber Nutzenden die [äusserst strengen Sicherheitsvorkehrungen rechtfertigen](#). Die KI-Firma hatte das Tool [in rund 1'000 Stunden](#) überprüft, um zu verhindern, dass die Vorkehrungen mit einem Trick, einem sogenannten [Jailbreak](#), umgangen werden könnten. Trotzdem wurde die Trump-Regierung alarmiert, es sei Firmen ein Jailbreak gelungen. Nur drei Tage nach der Veröffentlichung [verordnete die Trump-Regierung](#), dass die beiden Tools aus Gründen der nationalen Sicherheit sowohl für das Ausland wie auch für ausländische Personen innerhalb der USA gesperrt werden sollten. Betroffen waren damit auch Mitbegründer und Mitarbeitende von Anthropic. Sofort [schaltete Anthropic die Tools ab](#).

Reaktionen und ein Eigengoal für die USA

Die Reaktionen waren heftig. In den USA verlangten [Cyberexperten](#) in einem [offenen Brief](#) die Freigabe der Tools. Sie argumentierten, wie [Anthropic](#), dass die entdeckten Schwachstellen von geringerer Bedeutung seien und sich auch bei anderen Tools finden, zum Beispiel jenen von OpenAI. Am [G7-Gipfel](#) in Évian vom 15.-17. Juni brachte die EU in Anwesenheit der Tech-CEO's das Thema auf. Schliesslich wurde ein Programm für vertrauenswürdige Partner diskutiert. Dieses soll allen demokratischen Staaten Zugang ermöglichen.

Die USA wollen mit ihrer KI global dominieren, aber mit ihrem unüberlegten «Schnellschuss» reduzieren sie die Risiken nicht, unterminieren jedoch das Vertrauen in ihre KI-Unternehmen weltweit, womit Milliardeninvestitionen auf dem Spiel stehen. Offenbar braucht es globale Leitplanken – trotz oder gerade wegen des KI-Wettlaufs USA-China.

Mit freundlichen Grüssen

Für das Netzwerk der ehemaligen SSAB: Hanna Muralt Müller

23.6.2026

Falls Sie diese E-Mail nicht mehr erhalten möchten, melden Sie sich bitte bei mir: info@muralt-mueller.ch.