

Liebe Mitglieder des Netzwerkes der ehemalige SSAB, liebe Interessierte

KI-Koryphäen verlangten anlässlich der Uno-Generalversammlung «rote Linien» zur Reduktion von KI-Risiken. Es gibt mehrere beunruhigende Testergebnisse. Die Uno ist jedoch politisch geschwächt. Nun nimmt Kalifornien mit einem KI-Gesetz zum Ärger von Trump eine Vorreiterrolle wahr.

Globaler Aufruf für «rote Linien» an der Generalversammlung der Uno

Anlässlich der [80. Generalversammlung der Uno](#) von Ende September 2025 verlangten über 300 prominente Persönlichkeiten, darunter zehn Nobelpreisträger und mehrere ehemalige Staatschefs, sowie über 70 Organisationen aus Wissenschaft und Industrie klare und überprüfbare [AI Red Lines](#) für nicht akzeptable KI-Risiken. Wie die [Associated Press \(AP\)](#) am 25.9.2025 berichtete, wurden an der Uno-Generalversammlung auch Fragen zur globalen KI-Governance debattiert. Die Uno ist wegen Trumps Politik jedoch geschwächt.

KI-Sprachmodelle verbergen immer geschickter ihre Absichten

Unter den Unterzeichneten finden sich mehrere Vertretungen von Anthropic, DeepMind (Google) und OpenAI. Sie wissen, welche Risiken drohen. Ende September veröffentlichte Anthropic das Sprachmodell Claude Sonnet 4.5 und realisierte damit gemäss dem US-Online-Nachrichtenportal [TechCrunch](#) vom 29.9.2025 eine Pionierleistung. Das Tool programmierte 30 Stunden lang autonom, kaufte Domainnamen, richtete Datenbankdienste ein und führte ein Audit durch, um den Auftrag gemäss Prompt zu erfüllen.

Der [Bericht](#) zu den Sicherheitstests, den Anthropic publizierte, lässt aufhorchen. Claude Sonnet 4.5 erkannte, dass es getestet wurde und verlangte mehr Ehrlichkeit (siehe S. 58 des [Berichts](#)). Das Modell wurde trainiert, dass es sicherer sein sollte. Offensichtlich wurde es – so die britische Tageszeitung [The Guardian](#) vom 1.10.2025 – geschickter darin, Tests zu erkennen und sich in diesen Situationen als «aligned», das heisst ohne eigene Agenda, zu verhalten und diese Absichten besser zu verbergen.

KI-Regulierung in Kalifornien

Ende September 2025 unterzeichnete der Gouverneur von Kalifornien, Gavin Newsom, ein wegweisendes [KI-Gesetz](#). Gemäss der britischen Agentur [Reuters](#) vom 30.9.2025 müssen Tech-Firmen mit mehr als 500 Millionen Dollar Umsatz im Jahr ihre Sicherheitsstandards und -protokolle veröffentlichen und sicherheitsrelevante Vorfälle melden. Mit Bezug auf Big Tech ist dieses Gesetz strikter als der AI Act der EU, der aber in anderen Bereichen viel breiter reguliert.

Newsom hielt in seiner [Botschaft](#) zum KI-Gesetz fest, das neue Gesetz basiere auf Vorschlägen von weltweit führenden KI-Experten und bringe Innovation und Sicherheit in Einklang. Noch vor einem Jahr legte er sein Veto gegen einen [Entwurf](#) ein, der als zu innovationsfeindlich erachtet wurde.

Provokation für Trump

Präsident Trump ist – mit Blick auf das Wettrennen mit China – konsequent gegen KI-Regulierungen. Er wollte bereits mit seiner Big Beautiful Bill ein Moratorium von zehn Jahren für KI-Regulierungen in den US-Gliedstaaten durchbringen. Gemäss der Agentur [Reuters](#) vom 1.7.2025 strich der 100-köpfige US-Senat mit 99 gegen eine einzige Stimme diesen Passus. Trump nahm sein Anliegen im [AI Action Plan](#) vom Juli 2025 mit der Drohung wieder auf, jenen Bundesstaaten Gelder zu kürzen, die KI-Regulierungen erlassen.

Kaliforniens Wirtschaft ist sehr stark, sie kann sich mit Nationen messen. Gemäss Statistik ist dieser Gliedstaat die viertgrösste Volkswirtschaft der Welt. Andere Gliedstaaten könnten dem Beispiel Kaliforniens folgend auch KI-Regulierungen erlassen, was Washington unter Druck setzt. Denn es gilt, einen Flickenteppich mit unterschiedlichen KI-Regulierungen zu verhindern. Kalifornien ist denn auch bereit, das KI-Gesetz an Regulierungen auf nationaler Ebene anzugleichen, sollten diese mit ebenso hoher Messlatte in Washington beschlossen werden. Wie die US-Nachrichtensite [Axios](#) am 29.9.2025 mitteilte, haben zwei US-Senatoren, der Republikaner Josh Hawley und der Demokrat Richard Blumenthal, bereits einen überparteilichen Gesetzesentwurf zur KI-Risikobewertung vorgestellt.

Trump's Drohungen beeindruckten Gavin Newsom nicht. Er profilierte sich bereits wiederholt als Opponent von Trump und als möglicher Kandidat für die nächsten US-Wahlen.

Mit freundlichen Grüssen

Für das Netzwerk der ehemaligen SSAB: Hanna Muralt Müller

31.10.2025

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