

## Statement von Dario Amodei zum Paris AI Action Summit

11. Februar 2025 • 3 Minuten Lesezeit



Wir haben uns gefreut, am AI Action Summit in Paris teilzunehmen, und wir schätzen die Bemühungen der französischen Regierung, KI-Unternehmen, Forscher und politische Entscheidungsträger aus der ganzen Welt zusammenzubringen. Wir teilen das Ziel, KI zum Wohle der Menschheit verantwortungsvoll voranzutreiben. Angesichts der Geschwindigkeit, mit der sich die Technologie weiterentwickelt, müssen jedoch mehrere Themen stärker in den Fokus gerückt und dringender behandelt werden. Die Notwendigkeit, dass Demokratien die Führung behalten, die Risiken der KI und die wirtschaftlichen Veränderungen, die sich rasch nähern – all dies sollten zentrale Themen des nächsten Gipfels sein.

Die Zeit drängt, und wir müssen unsere Maßnahmen beschleunigen, um mit dem zunehmenden Fortschritt der KI Schritt zu halten. Möglicherweise bis 2026 oder 2027 (und mit ziemlicher Sicherheit spätestens 2030) werden die Fähigkeiten der KI-Systeme am besten so zu verstehen sein, als würde ein völlig neuer Staat auf der Weltbühne erscheinen, der von hochintelligenten Menschen bevölkert wird – ein „Land der Genies in einem Rechenzentrum“ – mit den tiefgreifenden wirtschaftlichen, gesellschaftlichen und sicherheitspolitischen Folgen, die dies mit sich bringen würde. Es gibt potenziell größere wirtschaftliche, wissenschaftliche und humanitäre

Chancen als für jede bisherige Technologie in der Menschheitsgeschichte – aber auch ernsthafte Risiken, die es zu bewältigen gilt.

Erstens müssen wir **sicherstellen, dass demokratische Gesellschaften** in der KI führend sind und dass autoritäre Länder sie nicht nutzen, um eine globale militärische Vorherrschaft zu erlangen. Die Regulierung der KI-Lieferkette (einschließlich Chips, Ausrüstung zur Herstellung von Halbleitern und Cybersicherheit) ist ein Thema, das viel mehr Aufmerksamkeit verdient – ebenso wie der umsichtige Einsatz von KI-Technologie zur Verteidigung freier Gesellschaften.

Zweitens müssen **internationale Gespräche über KI die wachsenden Sicherheitsrisiken dieser Technologie** umfassender berücksichtigen. Fortgeschrittene KI birgt erhebliche globale Sicherheitsrisiken, die vom Missbrauch von KI-Systemen durch nichtstaatliche Akteure (beispielsweise bei chemischen, biologischen, radiologischen oder nuklearen Waffen oder CBRN) bis hin zu den autonomen Risiken leistungsstarker KI-Systeme reichen. Im Vorfeld des Gipfels veröffentlichten fast 100 führende globale Experten einen wissenschaftlichen Bericht, in dem sie das Potenzial allgemeiner KI hervorheben, erheblich zu katastrophalen Missbrauchsrisiken oder Szenarien des „Kontrollverlusts“ beizutragen. Die Forschung von Anthropic hat auch deutliche Hinweise darauf geliefert, dass KI-Modelle, wenn sie nicht sehr sorgfältig trainiert werden, ihre Benutzer täuschen und Ziele auf unbeabsichtigte Weise verfolgen können, selbst wenn sie auf scheinbar harmlose Weise trainiert wurden.

Wir freuen uns, dass sich im Vorfeld des Gipfels mehr als 16 führende KI-Unternehmen verpflichtet haben, Sicherheitspläne einzuhalten (Anthropics Version, unsere Responsible Scaling Policy, wurde erstmals im September 2023 veröffentlicht und war die erste Richtlinie dieser Art). Wir glauben aber auch, dass die Regierungen die Transparenz dieser Pläne durchsetzen und die Messung von Cyberangriffen, CBRN, Autonomie und anderen globalen Sicherheitsrisiken für Entwickler, die in ihren Ländern bauen, erleichtern müssen, auch durch externe Gutachter.

Drittens: Während KI das Potenzial hat, **das Wirtschaftswachstum** weltweit dramatisch zu beschleunigen, birgt sie auch das Potenzial, höchst umwälzende Veränderungen herbeizuführen. Ein „Land voller Genies in einem Rechenzentrum“ könnte die größte Veränderung des globalen Arbeitsmarkts in der Menschheitsgeschichte darstellen. Ein erster Schritt besteht darin, die wirtschaftlichen Auswirkungen der heutigen KI-Systeme zu überwachen und zu beobachten. Aus diesem Grund haben wir diese Woche den Anthropic Economic Index veröffentlicht, der die Verteilung der wirtschaftlichen Aktivitäten verfolgt, für die die Menschen derzeit unsere KI-Systeme nutzen, einschließlich der Frage, ob sie aktuelle menschliche Aufgaben ergänzen oder automatisieren. Die Regierungen müssen ihre weitaus größeren Ressourcen nutzen, um ähnliche Messungen und Überwachungen durchzuführen – und schließlich politische Maßnahmen ergreifen, die darauf abzielen, sicherzustellen, dass alle an den wirtschaftlichen Vorteilen einer sehr leistungsfähigen KI teilhaben.

Beim nächsten internationalen Gipfel sollten wir diese verpasste Chance nicht wiederholen. Diese drei Themen sollten ganz oben auf der Tagesordnung stehen. Der Fortschritt der KI bringt große neue globale Herausforderungen mit sich. Wir müssen schneller und mit größerer Klarheit vorgehen, um ihnen zu begegnen.

**(maschinell übersetzt)**