AI Governance Briefing

Frontier AI als strategische Infrastruktur - Was der Fall Anthropic über die Zukunft der AI Governance verrät

Autorin: Dr. Silvia Moyses-Scheingruber
SHER DeepAI Research
München, 02. Juli 2026

Kategorien:
AI Governance | Frontier AI | AI-Risikomanagement | AI-Strategie

Die Entwicklungen rund um die Anthropic-Modelle Fable 5 und Mythos 5 im Juni 2026 markieren einen möglichen Wendepunkt in der Diskussion über AI Governance[1]. Nachdem die US-Regierung den Zugang zu den Modellen für Nutzer außerhalb der USA aufgrund nationaler Sicherheitsbedenken vorübergehend eingeschränkt hatte, wurden die Exportbeschränkungen Anfang Juli nach der Einführung zusätzlicher Sicherheitsmaßnahmen wieder aufgehoben[2]. Unabhängig davon, wie der konkrete Fall regulatorisch oder geopolitisch bewertet wird, zeigt er deutlich: Hochleistungsfähige KI-Systeme werden zunehmend nicht nur als digitale Werkzeuge, sondern als strategische Infrastruktur verstanden. Der Fall verdeutlicht, dass der Zugang zu leistungsfähigen Frontier-Modellen zunehmend zu einer Governance- und Strategiefrage wird. Der Fall könnte sich damit als wichtiger Referenzpunkt für eine neue Dimension der AI Governance erweisen: die Frage, wer Zugang zu besonders leistungsfähigen KI-Fähigkeiten erhält und unter welchen Bedingungen dieser Zugang gewährt oder eingeschränkt werden kann.

Unter AI Governance werden die Strukturen, Prozesse und Verantwortlichkeiten verstanden, mit denen Organisationen die Entwicklung, Bereitstellung und Nutzung von KI-Systemen steuern, überwachen und kontrollieren (IBM, 2026; OECD, 2019; ISO/IEC, 2023; NIST, 2023). In den vergangenen Jahren standen beim Thema AI Governance vor allem Themen wie Transparenz, Fairness, Risikomanagement und regulatorische Compliance im Vordergrund. Diese Aspekte bleiben auch künftig von zentraler

[1] Siehe Veröffentlichung von Atrophic vom 16.06.2026
https://www.anthropic.com/news/fable-mythos-access
[2] Siehe mediale Berichterstattung z,B. Tagesschau
https://www.tagesschau.de/wirtschaft/unternehmen/usa-anthropic-ki-100.html (aufgerufen am 02.07.2026)

Bedeutung, da sie die Grundlage für einen verantwortungsvollen und vertrauenswürdigen Einsatz von KI bilden.

Mit dem Aufkommen leistungsfähiger Frontier-Modelle[3] erweitert sich jedoch der Gegenstandsbereich der AI Governance erheblich. Während sich AI Governance bislang vor allem mit der Frage beschäftigte, wie KI-Systeme sicher und verantwortungsvoll entwickelt und genutzt werden können, rückt zunehmend in den Vordergrund, wer Zugang zu besonders leistungsfähigen KI-Fähigkeiten erhält, wie die damit verbundenen Risiken bewertet werden und unter welchen Bedingungen ihre Nutzung gesteuert oder eingeschränkt werden kann.

Damit geht AI Governance zunehmend über klassische Compliance- und Risikomanagement Ansätze hinaus. Sie umfasst heute immer mehr Überlegungen zur technologischen Souveränität, zu Abhängigkeiten von einzelnen Anbietern sowie zu den wirtschaftlichen und sicherheitspolitischen Auswirkungen leistungsfähiger KI-Systeme. AI Governance entwickelt sich damit von einem vorwiegend operativen Steuerungsinstrument zu einem strategischen Rahmenwerk für den Umgang mit KI als kritischer Technologie und potenzieller Infrastruktur der Zukunft.

Warum stellt Frontier AI neue Herausforderungen an die AI Governance?

Die wachsende Bedeutung von AI Governance hängt eng mit einer neuen Generation von KI-Systemen zusammen, die häufig als *Frontier AI* bezeichnet werden. Darunter werden jene KI-Systeme bezeichnet, die zum jeweiligen Zeitpunkt die Leistungsgrenze des technisch Machbaren repräsentieren. Sie zeichnen sich nicht nur durch ihre hohe Leistungsfähigkeit aus, sondern vor allem durch ihre Vielseitigkeit und die Breite ihrer Einsatzmöglichkeiten. Anders als spezialisierte KI-Anwendungen können Frontier-Modelle in einer Vielzahl von Bereichen eingesetzt werden und komplexe Aufgaben übernehmen, die bislang menschliche Expertise erforderten. Aufgrund ihres Potenzials, Wissenschaft, Wirtschaft, Cybersicherheit und nationale Sicherheit maßgeblich zu beeinflussen, rücken Frontier-Modelle zunehmend in den Fokus von Governance-, Sicherheits- und Risikodiskussionen und werden verstärkt als strategische Technologien betrachtet (NIST, 2023, UK AI Safety Institute, 2025; Stanford AI Index Report, 2026).

[3] Als Frontier AI werden KI-Systeme bezeichnet, die zum jeweiligen Zeitpunkt die Leistungsgrenze des technisch Machbaren repräsentieren, beispielsweise die jeweils leistungsfähigsten Versionen von Claude, GPT oder Gemini. Diese sehr leistungsfähigen KI-Systeme können potenziell weitreichende Auswirkungen auf Wirtschaft, Gesellschaft und Sicherheit haben.

Je leistungsfähiger diese Systeme werden, desto größer wird die Sorge, dass ihre Fähigkeiten missbraucht oder für Zwecke eingesetzt werden könnten, die erhebliche Schäden verursachen. Diskutiert werden beispielsweise Risiken im Bereich der Cybersicherheit, die Unterstützung bei der Entwicklung biologischer oder chemischer Gefahrstoffe, sowie die automatisierte Erstellung und Verbreitung von Desinformation. Gleichzeitig besteht die Herausforderung darin, dass viele dieser Risiken schwer zu messen sind und ihre tatsächlichen Auswirkungen nur begrenzt vorhersehbar sind.

Gerade diese Kombination aus hoher Leistungsfähigkeit, breiter Einsetzbarkeit und potenziell weitreichenden Folgen unterscheidet Frontier-Modelle von herkömmlichen KI-Anwendungen und macht sie zu einem besonderen Gegenstand von Governance- und Sicherheitsdebatten.

AI Governance wird zur Führungsaufgabe

Leistungsfähige Frontier-Modelle entwickeln sich zunehmend zur technologischen Grundlage von Produkten, Dienstleistungen und Geschäftsprozessen. Ihre Auswahl ist daher nicht mehr ausschließlich eine technische, sondern vor allem eine strategische Entscheidung. Neben Leistungsfähigkeit, Sicherheit und Kosten gewinnen die langfristige Verfügbarkeit von Modellen, mögliche Nutzungsbeschränkungen sowie regulatorische und geopolitische Entwicklungen zunehmend an Bedeutung. Mit der Entscheidung für ein Frontier-Modell entstehen zugleich Abhängigkeiten von dessen Anbieter, die sich auf Innovationsfähigkeit, Planungssicherheit und unternehmerische Handlungsoptionen auswirken können.

Für Unternehmen, die KI-Systeme entwickeln, einsetzen oder betreiben, hat diese Entwicklung weitreichende Konsequenzen. AI Governance kann künftig nicht mehr allein als Frage von Compliance, Datenschutz oder Risikomanagement verstanden werden. Mit der wachsenden strategischen Bedeutung von Frontier AI entwickelt sie sich zu einer Führungsaufgabe. Entscheidungen über Frontier AI betreffen heute nicht mehr ausschließlich Technologie, sondern ebenso Wettbewerbsfähigkeit, Resilienz, technologische Souveränität und die langfristige Unternehmensstrategie. Dies erfordert eine enge Zusammenarbeit zwischen Unternehmensführung, Technologieverantwortlichen sowie den Bereichen Risiko, Recht und Compliance, um Chancen und Risiken leistungsfähiger KI-Systeme ganzheitlich bewerten und steuern zu können.

Literatur

Anthropic (2025) *Responsible Scaling Policy (RSP)*. San Francisco: Anthropic.

IBM (2026) *What is AI Governance?* Available at: https://www.ibm.com/de-de/think/topics/ai-governance (Accessed: 17 June 2026

ISO/IEC (2023) *ISO/IEC 42001:2023 – Artificial Intelligence Management Systems*. Geneva: International Organization for Standardization.

National Institute of Standards and Technology (NIST) (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD: NIST.

OECD (2019) *OECD Principles on Artificial Intelligence*. Available at: https://www.oecd.org/en/topics/ai-principles.html (Accessed: 17 June 2026).

Stanford Institute for Human-Centered Artificial Intelligence (HAI) (2026) *AI Index Report 2026*. Stanford, CA: Stanford University. Available at: https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf
(Accessed: 17 June 2026).

UK AI Safety Institute (2025) *Frontier AI: Capabilities and Risks – Discussion Paper*. Available at: https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/frontier-ai-capabilities-and-risks-discussion-paper (Accessed: 16 June 2026).

Über SHER DeepAI

SHER DeepAI entwickelt eine modulare Trustworthy-AI-Plattform mit den Schwerpunkten Explainable und Efficient AI. Die Plattform unterstützt Unternehmen, Entwickler und AI-Teams dabei, vertrauenswürdige, transparente, effiziente und regulatorisch konforme KI-Systeme zu entwickeln und zu betreiben.

Zu den zentralen Capabilities gehören Explainability, Logging und Traceability, Data Governance, Risk Management, Model und Operational Monitoring, Compliance Monitoring, AI Security, sowie AI Optimization. Damit ermöglicht SHER DeepAI die Operationalisierung von AI Governance und unterstützt bei der Entwicklung vertrauenswürdiger, effizienter und nachhaltiger KI-Systeme über den gesamten KI-Lebenszyklus.

SHER DeepAI ist Ihr Partner für die Entwicklung vertrauenswürdiger, erklärbarer und effizienter KI-Systeme.