KI: Der Vatikan denkt schärfer als das Bundeskanzleramt

Posted on 15. Juni 2026 by Axel Fersen

Screenshot: Frontier Risiko: Neue Modelle lösen Panik bei OpenAI aus! auf Youtube

Es ist ein bemerkenswerter Befund: Während in Washington, London, Paris, Tokio und Singapur staatliche Institutionen entstanden sind, die sich systematisch mit den Risiken hochentwickelter Künstlicher Intelligenz befassen, und während die katholische Kirche mit der Note *Antiqua et nova* vom 28. Januar 2025 eine philosophisch wie ethisch dichte, hundertsiebzehn Abschnitte umfassende Analyse der KI-Risiken vorgelegt hat, liest sich die Antwort der deutschen Bundesregierung wie ein Werbeprospekt: Deutschland werde „KI-Nation“ — so der Koalitionsvertrag der schwarz-roten Regierung vom April 2025 —, KI sei eine Schlüsseltechnologie, die Hightech Agenda Deutschland vom Juli 2025 setze auf „massive Investitionen“ und der Bundesminister für Digitalisierung verkünde aus Dubai, Deutschland nehme die Technologie „at full speed“ an. Von Risiken ist kaum die Rede. Von einem deutschen AI Safety Institute, von systematischer Risikoforschung: nichts.

Dass ausgerechnet der Vatikan in dieser Frage gegenwärtig schärfer denkt als das Bundeskanzleramt, ist mehr als eine Pointe. Es ist ein Alarmsignal. Wer Antiqua et nova liest, findet dort eine nüchterne Aufzählung jener Gefahrenfelder, die im internationalen Fachdiskurs seit Jahren benannt werden:

- autonome Waffensysteme, deren Verbot der Vatikan unmissverständlich fordert;
- die Konzentration enormer Macht in den Händen weniger Konzerne;
- Desinformation und Deepfakes;
- algorithmische Diskriminierung und Social Scoring;
- die ökologische Last der Rechenzentren;
- der Verlust menschlicher Arbeit durch Substitution statt Begleitung;
- und nicht zuletzt — Abschnitt 101 der Note ist hier von beklemmender Klarheit — das von KI-Forschern selbst beschriebene „existenzielle Risiko" einer Technologie, „die in der Lage ist, auf eine Weise zu handeln, die das Überleben der Menschheit oder ganzer Regionen bedrohen könnte".

Das ist nicht römische Apokalyptik, sondern die nüchterne Wiedergabe dessen, was Nobelpreisträger *Geoffrey Hinton*, Turing-Preisträger *Yoshua Bengio*, *Stuart Russell* von Berkeley, *Demis Hassabis* von Google DeepMind, *Sam Altma*n von OpenAI und — bemerkenswerterweise — auch *Dario Amodei*, der CEO von Anthropic, mit zunehmender Dringlichkeit formulieren. Hinton, der sein Lebenswerk bei Google im Frühjahr 2023 öffentlichkeitswirksam aufgab, schätzte das Risiko einer Übernahme der Kontrolle durch KI auf „zehn bis zwanzig Prozent"; bei seinem Nobelpreis-Bankett in Stockholm warnte er vor *„schrecklichen neuen Viren und grauenhaften tödlichen Waffen, die selbst entscheiden, wen sie töten oder verstümmeln*". Bengio leitet das Expertengremium hinter dem International AI Safety Report 2026. Amodei beschreibt in seinem im Januar 2026 veröffentlichten Memorandum, der Welt drohe binnen ein bis zwei Jahren ein „country of geniuses in a datacenter", dessen Risiken Regierungen und Öffentlichkeit „dramatisch unterschätzen".

Unter dem Titel „Antiqua et nova – und das deutsche Nichts" erschien der Beitrag zuerst als eine Veröffentlichung des Erhard-Eppler-Kreises. Danke für die Übernahmemöglichkeit.

Die ehrliche Antwort lautet: erstaunlich wenig

Diese Stimmen sind nicht randständig, sie sind das Zentrum der Disziplin. Wer mit den vier Turing-Preisträgern Hinton, Bengio, Russell und Hassabis nicht einverstanden ist, müsste argumentieren, warum man Pioniere der Disziplin in der Sache ihres eigenen Lebenswerks für inkompetent halten soll. Der Vatikan hat diese Stimmen ernst genommen und in eine theologisch begründete, aber gerade nicht abergläubische oder reaktionäre Stellungnahme eingearbeitet. Was hat die deutsche Bundesregierung getan?

Die ehrliche Antwort lautet: erstaunlich wenig. Deutschland verfügt zwar seit 2023 über ein Institut für KI-Sicherheit beim DLR in Sankt Augustin, dessen Auftrag jedoch primär auf zertifizierungsnahe technische Prüfverfahren zielt. Es gibt das DFKI, dessen CEO Antonio Krüger — als deutsches Mitglied des Expert Advisory Panel des AI Safety Reports — im Februar 2026 öffentlich erklärte: *„Andere wichtige KI-Nationen*

sind uns hier einen Schritt voraus. Wir brauchen ein deutsches AI Safety Institute, das die dynamische Entwicklung von KI und ihre Risiken im Blick hat und die Bundesregierung kontinuierlich berät“. Es gibt den Deutschen Ethikrat, dessen Stellungnahme *Mensch und Maschine* vom März 2023 in vielen Punkten klug formuliert ist, der jedoch durch die rasante Entwicklung der Frontier-Modelle der Jahre 2024 bis 2026 in entscheidenden Risikofragen — agentische Systeme, instrumentelle Konvergenz, Schutz vor Modellabschaltung — bereits beim Erscheinen veraltet war und seither nicht fortgeschrieben wurde. Und es gibt seit Mai 2025 ein nigelnagelneues Bundesministerium für Digitalisierung unter *Karsten Wildberger*, dem ehemaligen CEO der Mediamarkt-Saturn-Holding Ceconomy, dessen öffentliche Äußerungen sich bisher in Standardformeln zu Geschwindigkeit, Innovationsfreundlichkeit und Souveränität erschöpfen.

Was Deutschland — anders als Großbritannien, die USA, Frankreich, Japan, Kanada, Indien, Singapur, Südkorea und Australien — nicht hat, ist ein staatliches AI Safety Institute mit dem klaren Auftrag, Frontier-Modelle vor und nach ihrer Veröffentlichung systematisch zu evaluieren, die Bundesregierung in Fragen von KI und nationaler Sicherheit zu beraten und technische Forschung zu KI-gestützten Bedrohungen zu betreiben. Pikant: Deutschland hatte bereits 2024 in einer gemeinsamen Erklärung mit neun anderen Ländern und der EU empfohlen, dass Staaten eigene Institute für KI-Sicherheit gründen sollten — und ist heute der einzige Unterzeichner dieser Erklärung außer Italien, der nicht Teil des internationalen Network of AI Safety Institutes ist. Man hat unterschrieben, ohne zu liefern.

Die Risikoanalyse fehlt fast vollständig

Open Access

Der Vatikan dagegen liefert — wenigstens auf der Ebene der ethischen Reflexion. *Antiqua et nova* benennt die Macht weniger Unternehmen als ethisches Problem ersten Ranges; sie warnt vor dem „technokratischen Paradigma“, das alle Probleme allein mit technologischen Mitteln zu lösen vorgibt; sie spricht in Abschnitt 100 deutlich aus, dass autonome tödliche Waffensysteme „ein ernster Grund für ethische Bedenken“ seien, und ruft zu einem Verbot auf: „*Keine Maschine darf jemals die Wahl treffen können, einem Menschen das Leben zu nehmen*“. Im Koalitionsvertrag der Bundesregierung findet sich von alldem so gut wie nichts. KI taucht dort als Wirtschaftsfaktor auf, als Modernisierungshebel der Verwaltung, als Verteidigungsressource für die Bundeswehr und — bemerkenswert — als Werkzeug für die „retrograde biometrische Fernidentifizierung“ durch Sicherheitsbehörden. Die Risikoanalyse aber, die der Vatikan auf siebzig Seiten leistet, fehlt fast vollständig.

Diese Abwesenheit ist umso bemerkenswerter, als die KI-Risiken längst nicht mehr spekulativ sind. Anthropic selbst hat im November 2025 den Fall GTG-1002 öffentlich gemacht — einen Cyberangriff, bei dem ein Frontier-Modell agentisch gegen die Schutzmechanismen seines eigenen Herstellers eingesetzt wurde. Die

Forschungsliteratur zu *scheming*, zu *reward hacking*, zu *instrumental convergence* und zu agentischer Fehlauslegung füllt mittlerweile Hunderte begutachteter Papers. Dass die deutsche Bundesregierung diese empirisch gesicherten Befunde nicht zum Anlass nimmt, eine eigene staatliche Risikobeobachtungsstelle zu errichten, ist nicht Vorsicht. Es ist Versäumnis.

Geschwindigkeit ist das eine, Bremsfähigkeit das andere

Es ist auch ein politisches Problem für die Sozialdemokratie. Wer in der Tradition Egon Bahrs, Willy Brandts und Erhard Epplers eine Politik gemeinsamer Sicherheit vertritt, kann technologischen Risiken globalen Ausmaßes nicht mit Standortlogik begegnen. Die Erfahrung des Kalten Krieges lehrt, dass Sicherheit nicht entsteht, indem man der Gegenseite zuvorkommt, sondern indem man Strukturen schafft, in denen das Wettrüsten gehegt und kontrolliert wird. Auf das KI-Wettrüsten übertragen heißt das: Deutschland und Europa brauchen Institutionen, die nicht nur Geschwindigkeit organisieren, sondern Bremsfähigkeit. Die Vatikan-Note formuliert es in Abschnitt 99 fast wortgleich: Die Leichtigkeit, mit der autonome Waffen Krieg führbarer machen, beschleunige „ein destabilisierendes Wettrüsten mit verheerenden Folgen für die Menschenrechte“.

Was wäre zu tun? Ein deutsches AI Safety Institute mit dem expliziten Auftrag, Frontier-Modelle vor und nach ihrer Veröffentlichung zu evaluieren — finanziell ausgestattet auf dem Niveau des britischen AISI. Eine ressortübergreifende Stabsstelle im Bundeskanzleramt, die KI-bezogene Sicherheitsfragen koordiniert. Eine Beauftragung des Deutschen Ethikrats mit einer Fortschreibung von *Mensch und Maschine*, die den Stand 2026 abbildet. Eine Bundestagskommission „KI und nationale Sicherheit“ mit Anhörungsrechten gegenüber Frontier-Lab-Vertretern. Und eine sozialdemokratische wie christdemokratische Wiederaneignung jenes Themas, das gegenwärtig zwischen Beschleunigungsrhetorik und kulturpessimistischen Reflexen zerrieben wird.

Quellen

Dikasterium für die Glaubenslehre / Dikasterium für die Kultur und die Bildung (2025): *Antiqua et nova. Note über das Verhältnis von künstlicher Intelligenz und menschlicher Intelligenz*, 28. Januar 2025. vatican.va
Deutscher Ethikrat (2023): *Mensch und Maschine — Herausforderungen durch Künstliche Intelligenz*, Stellungnahme vom 20. März 2023. ethikrat.org
Bundesregierung (2025): *Hightech Agenda Deutschland*, Kabinettsbeschluss vom 30. Juli 2025. bundesregierung.de
DFKI (2026): International AI Safety Report 2026 erschienen — Statement Antonio Krüger, 3. Februar 2026. dfki.de
Fox, Philip / Leicht, Anton (2025): Warum Deutschland ein AI Security Institute braucht. Tagesspiegel

Background, 28. August 2025. tagesspiegel.de
Europäische Kommission (2024): First meeting of the International Network of AI Safety Institutes, San Francisco, 20.–21. November 2024. digital-strategy.ec.europa.eu
Center for AI Safety (2023): Statement on AI Risk, 30. Mai 2023. Unterzeichner u. a. Hinton, Bengio, Hassabis, Altman, Amodei, Russell. safe.ai
Deutsches Ärzteblatt (2024): Nobelpreisträger Hinton warnt vor Gefahren von Künstlicher Intelligenz. 11. Dezember 2024. aerzteblatt.de
Amodei, Dario (2024): *Machines of Loving Grace*. darioamodei.com — sowie Axios (2026): Anthropic CEO's grave warning, 26. Januar 2026. axios.com
Euronews (2026): Germany is embracing AI „at full speed" — Interview Karsten Wildberger beim World Governments Summit, Dubai, 5. Februar 2026. euronews.com
Kuhn, Johannes (2026): Nach Mythos — Ist die Zeit reif für ein deutsches AI Safety Institut? Tagesspiegel Background, 21. Mai 2026. tagesspiegel.de

- E-Mail
- teilen
- teilen
- teilen
- teilen

Entdecke mehr von bruchstücke

Melde dich für ein Abonnement an, um die neuesten Beiträge per E-Mail zu erhalten.

Gib deine E-Mail-Adresse ein …

Abonnieren