Deutsches AI Security Institut (DE-AISI)

Positionspapier

bitkom

Deutsches AI Security Institut auf Weltklasseniveau

Die zentrale Herausforderung

Die rasanten Fortschritte im Bereich der Künstlichen Intelligenz beeinflussen unmittelbar die nationale Sicherheit und Souveränität Deutschlands. Nicht zuletzt die neuartigen Cyberfähigkeiten des KI-Modells »Mythos« von Anthropic demonstrieren dies noch einmal eindrücklich.[1] Deutschland ist hierauf jedoch nicht angemessen vorbereitet – unter anderem, weil der Bundesregierung ein hinreichend differenziertes Lagebild über die Fähigkeiten und Auswirkungen von Frontier-AI-Modellen fehlt. Deutschland hat sich mit der Seoul Declaration 2024[2] ausdrücklich dazu bekannt, den Aufbau bzw. die Erweiterung von AI Safety/Security Instituten und vergleichbaren Einrichtungen zu unterstützen. Die Bundesregierung plant diese Zusage jetzt einzulösen, kritische Ausgestaltungsfragen zum Aufbau eigener staatlicher Kapazitäten sind jedoch bislang noch offen.

Handlungsempfehlung auf einen Blick

Deutschland sollte kurzfristig ein eigenes **deutsches AI Security Institut (DE-AISI) auf Weltklasseniveau nach Vorbild des UK AI Security Instituts (UK AISI)** aufbauen. Folgende Ausgestaltungsaspekte sind hierbei zentral für den Erfolg:

- Das DE-AISI sollte ein Mandat ausschließlich **zur Forschung und Lagebilderstellung zu Fähigkeiten, Sicherheitsmaßnahmen und Auswirkungen von Frontier-AI-Modellen** erhalten.
- Das DE-AISI sollte explizit **kein Regulierungsmandat erhalten** und sein Forschungsmandat muss **klar von bestehenden Forschungseinrichtungen abgegrenzt werden**.
- Das DE-AISI sollte sich thematisch auf **systemische Auswirkungen von Frontier-AI-Modellen auf Deutschlands nationale Sicherheit und Souveränität fokussieren** – insbesondere auf Cyberoffensivfähigkeiten, Missbrauch zur Entwicklung von Massenvernichtungswaffen, Kontrollverlustrisiken und Massenmanipulation.
- Das DE-AISI sollte sich **nicht mit Fragen des Arbeits-, Verbraucher- und Datenschutzes oder der KI-Ethik** befassen. Hierfür sind andere Organe zuständig.
- **Die Gewinnung internationaler technischer Spitzentalente für das DE-AISI muss absolute Priorität haben.** Zentrale Strukturfragen des DE-AISI sind maßgeblich mit Blick auf dieses Ziel zu beantworten.
- Das DE-AISI braucht eine angemessene Finanzierung mit einem **Anfangsbudget von mindestens 100 Millionen Euro für die ersten zwei Jahre** sowie einer langfristig gesicherten **Anschlussfinanzierung von mindestens 75 Millionen Euro jährlich**, auf dem Niveau des UK AISI. Die Freigabe der Mittel sollte jedoch an die erfolgreiche Gewinnung von Spitzentalenten geknüpft werden.

Konzeption

- **Kernaufgaben:**
  - **Technische Evaluierung von Frontier-AI-Modellen:** Systematische Tests der Fähigkeiten aktueller und kommender Modelle unter kontrollierten Bedingungen, mit dem Ziel, reproduzierbare Ergebnisse zu liefern, die als Entscheidungsgrundlage für die Bundesregierung dienen. Hierzu zählen insbesondere Fähigkeitsevaluierungen, Red-Teaming, Missbrauchsanalysen sowie Untersuchungen sicherheitsrelevanter emergenter Fähigkeiten von Frontier-AI.
  - **Lagebilderstellung:** Regelmäßige ressortübergreifend nutzbare Lagebilder und Berichte über den Stand der Frontier-AI-Entwicklung, gerichtet an das Bundesministerium für Digitales und Staatsmodernisierung (BMDS), das Bundeskanzleramt (BKAmt), das Bundesministerium des Inneren (BMI), das Bundesministerium der Verteidigung (BMVg), den Nationalen Sicherheitsrat, sowie weitere relevante Ressorts. Diese sollen bestehende strategische Analysen ergänzen und insbesondere Auswirkungen auf nationale Sicherheit, Souveränität und kritische Infrastrukturen bewerten. Ziel ist dabei nicht die operative Gefahrenabwehr, sondern die frühzeitige Identifikation technologischer Entwicklungen und sicherheitsrelevanter Fähigkeiten fortgeschrittener KI-Systeme.
  - **Erforschung von Sicherheitsmaßnahmen:** Erforschung, Evaluierung und Entwicklung technischer Ansätze zur Verbesserung der Sicherheit und Kontrollierbarkeit fortgeschrittener KI-Systeme sowie zur Reduktion kritischer Missbrauchs- und Kontrollverlustrisiken.
  - **Internationale Kooperation im Netzwerk der AI Safety bzw. Security Institute:** Enge Abstimmung mit den DE-AISI-Pendants[3] anderer Länder und dem EU AI Office. Das DE-AISI sollte dabei an die etablierten Evaluierungs-Arbeitsstränge der KI-Sicherheitsinstitute des Vereinigten Königreichs und der USA sowie an die G7-Bemühungen anknüpfen und aktiv zu internationalen Standardsetzungsprozessen (u. a. ISO) und gemeinsamen Sicherheitsbenchmarks beitragen, um global interoperable, wissenschaftsbasierte und innovationsfreundliche Ansätze sicherzustellen.

- **Umsetzung:**
  - Ein Kabinettsbeschluss sollte kurzfristig die Gründung eines DE-AISI in Form einer bundeseigenen GmbH ermöglichen.
  - Ergänzend hierzu braucht es eine gesetzliche Grundlage, um eine gezielte Ausnahme vom Besserstellungsverbot zu schaffen und mittelfristig den Zugang zu Verschlusssachen zu eröffnen.
  - Für die Anbindung des DE-AISI bietet sich das BMDS an. Relevante Ministerien und staatliche Organe, wie das BKAmt, das BMI, und das BMVg, sollten sinnvoll eingebunden werden. Entscheidend ist jedoch eine **klare Federführung**, um **strategische Kohärenz** sicherzustellen, **Abstimmungsaufwände** zu begrenzen und **schnelle Entscheidungen** des Instituts zu gewährleisten.
  - Das DE-AISI sollte zielgerichtet mit **der deutschen Cybersicherheitsstrategie verzahnt werden und Doppelstrukturen zu bestehenden Behörden** vermeiden – insbesondere gegenüber dem Bundesamt für Sicherheit in der Informationstechnik (BSI) und der Bundesnetzagentur (BNetzA).

- **Finanzierung:** Die Größendimension der Finanzierung sollte sich an der des UK AISI orientieren. Das heißt, für den Aufbau sollten **mindestens 100 Millionen Euro für die ersten zwei Jahre reserviert werden, und daran anschließend sollte eine langfristige Finanzierungsgarantie von mindestens 75 Millionen Euro jährlich vorgesehen werden**. Dies ist erforderlich, um wettbewerbsfähige Gehälter zu zahlen und die sichere technische Infrastruktur aufzubauen, die für belastbare Modellevaluierungen auf Augenhöhe mit Frontier-AI-Anbietern erforderlich ist. Die Freigabe der Mittel sollte jedoch stufenweise und streng an die Verfügbarkeit von Spitzenpersonal mit entsprechender technischer und wissenschaftlicher Reputation gekoppelt sein.

- **Erfolgsbedingungen:** Ein DE-AISI sollte sich möglichst stark an dem britischen Vorbild des UK AI Security Instituts und seinem einzigartigen Ansatz orientieren. Das heißt, es darf nicht wie eine klassische Forschungseinrichtung oder Behörde funktionieren, sondern muss als eine **Art Hochleistungs-Technologie-Startup innerhalb der Regierung** wirken. Dafür müssen folgende Anforderungen erfüllt sein:

  - **Technische Qualifikation der Leitung, des Gründungsteams und der Mitarbeitenden auf internationalem Spitzenniveau:** Das technische Qualifikationsniveau der Mitarbeitenden wird der entscheidende Erfolgsfaktor für das DE-AISI sein. Nur mit internationalem Spitzenpersonal kann das DE-AISI Modellevaluierungen auf Augenhöhe mit den Frontier-AI-Anbietern durchführen und fundierte sicherheitskritische Erkenntnisse für die Bundesregierung gewinnen. Um Personal mit dem erforderlichen technischen Qualifikationsniveau gewinnen zu können, sind vor allem die Reputation und Kompetenz der technischen Leitung sowie des Gründungsteams maßgeblich – dies legen die Erfahrungen des UK AI Security Institute eindrücklich nahe. Daher muss der erste Schritt die Gewinnung einer technischen Leitungsperson mit internationaler Spitzenreputation sein, die ein Gründungsteam mit entsprechender Spitzenexpertise aufbauen kann.

  - **Forschungsexzellenz als Glaubwürdigkeitswährung**: Wie das UK AISI muss auch das DE-AISI eine klare Spitzenforschungsambition haben. Das UK AISI hat zahlreiche Beiträge auf den weltweit führenden KI-Konferenzen sowie eine Studie in *Science* veröffentlicht. Diese Verbindung aus akademischer Sichtbarkeit und operativer Exzellenz bewegt die großen Frontier-AI-Anbieter dazu, dem Institut freiwillig Zugang zu unveröffentlichten Modellen zu gewähren, weil sie selbst von dessen Kompetenz profitieren. Gleichzeitig ist diese Forschungsexzellenz notwendig, um Spitzentalente anzuziehen.

  - **Missionsdisziplin:** Das DE-AISI sollte sich auf die oben beschriebenen Kernaufgaben konzentrieren, um politischer Vereinnahmung zu widerstehen und eine hohe technische Reputation zu wahren. Angrenzende Themen wie KI-Wettbewerbsfähigkeit oder den Ausbau von Rechenzentren sollten andere Regierungseinheiten bearbeiten, die technische Expertise vom DE-AISI dann einbeziehen können.

  - **Flexibilität und Agilität:** Das DE-AISI muss schnell, flexibel und unbürokratisch agieren, dies beinhaltet beispielsweise die Möglichkeit, Personal sehr zügig einzustellen und Rechenressourcen zügig freigeschaltet zu bekommen.

  - **Wettbewerbsfähige Gehälter:** Hauptanziehungsfaktor für Spitzentalente werden Mission, Agilität und technische Exzellenz des DE-AISI sein. Dennoch muss das DE-AISI außertarifliche Vergütungsmöglichkeiten anbieten können.

- **Politische Unterstützung auf höchster politischer Ebene:** Das DE-AISI muss dauerhafte politische Unterstützung von höchster politischer Ebene erhalten. Im Vereinigten Königreich hat sich der damalige Premierminister persönlich für die Rekrutierung der ersten Spitzentalente eingesetzt. Ein vergleichbares Engagement ist auch in Deutschland erforderlich.
- **Sicherer Infrastrukturzugang als Voraussetzung für Modellzugang:** Das DE-AISI muss Zugang zu sicherer und hochleistungsfähiger technischer Infrastruktur für die Testung von Frontier-AI-Modellen erhalten.

- **Vorbild für Staatsmodernisierung:** Ein nach diesen Prinzipien aufgebautes DE-AISI könnte weit über den KI-Bereich hinaus Wirkung entfalten. Gelingt es, eine technisch exzellente, agile und international anschlussfähige Institution innerhalb des deutschen Staats aufzubauen, könnte das Institut zum Vorbild für Staatsmodernisierung werden.

Appendix

Vorbild UK AISI

Das Vereinigte Königreich hat eindrucksvoll gezeigt, wie sich in kürzester Zeit ein weltweit führendes AISI aufbauen lässt. Das UK AISI wurde 2023 gegründet, verfügt über ein Jahresbudget von umgerechnet über 76 Millionen Euro und hat bevorzugten und umfangreichen Zugang zu staatlichen Rechenressourcen.[4] Dank unbürokratischer Strukturen und wettbewerbsfähiger Gehälter konnte es bereits in den ersten elf Wochen ein Team internationaler Spitzenforscherinnen und -forscher und Expertinnen und Experten aus führenden KI-Unternehmen gewinnen.[5] Heute beschäftigt das Institut über 250 Mitarbeitende, trägt mit anerkannten Beiträgen zur internationalen Spitzenforschung bei und stellt der britischen Regierung einzigartige Expertise zu Frontier-AI-Entwicklungen zur Verfügung.[6] Damit hat das Vereinigte Königreich seine Führungsrolle im Bereich Künstlicher Intelligenz nachhaltig gestärkt. Zudem ist das UK AISI als eine Art Hochleistungs-Technologie-Startup innerhalb des britischen Forschungs- und Innovationsministeriums (DSIT) konzipiert. Diese Arbeitsweise hat Strahlkraft weit über das Institut hinaus entfaltet: Das AISI wird inzwischen regelmäßig von anderen Ressorts innerhalb der britischen Regierung um Rat und Unterstützung gebeten.

---

[1] UK AI Security Institute. (2026). ↗Our evaluation of Claude Mythos Preview's cyber capabilities. AISI Work Blog
[2] Außenministerium der Republik Korea. (2024, 21. Mai). ↗Seoul-Erklärung für sichere, innovative und inklusive KI der Teilnehmenden an der Sitzung der Staats- und Regierungschefs des AI Seoul Summit, 21. Mai 2024. Ministry News
[3] Diese Länder haben bereits ein AISI aufgebaut: Vereinigte Staaten von Amerika, Vereinigtes Königreich, Kanada, Frankreich, Japan, Südkorea, Singapur, China, Australien.
[4] ↗The AI Security Institute
[5] Frontier AI Taskforce. *↗First progress report. GOV.UK.*
[6] The AI Security Institute. *↗AISI Blog*

Bitkom vertritt mehr als 2.300 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie generieren in Deutschland gut 200 Milliarden Euro Umsatz mit digitalen Technologien und Lösungen und beschäftigen mehr als 2 Millionen Menschen. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 700 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig, kreieren Content, bieten Plattformen an oder sind in anderer Weise Teil der digitalen Wirtschaft. 82 Prozent der im Bitkom engagierten Unternehmen haben ihren Hauptsitz in Deutschland, weitere 8 Prozent kommen aus dem restlichen Europa und 7 Prozent aus den USA. 3 Prozent stammen aus anderen Regionen der Welt. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem leistungsfähigen und souveränen Digitalstandort zu machen.

Herausgeber
Bitkom e.V.
Albrechtstr. 10 | 10117 Berlin

Ansprechpartner
Lucy Czachowski | Bereichsleiterin KI & Cloud – Resilienz & Infrastruktur
T +49 30 27576-320 | l.czachowski@bitkom.org

Janis Hecker | Bereichsleiter AI – Regulierung & Strategie
T +49 30 27576-239 | j.hecker@bitkom.org

Verantwortliches Bitkom-Gremium
AK Artificial Intelligence