BDI

POSITIONSPAPIER | DIGITALPOLITIK | KÜNSTLICHE INTELLIGENZ

Gründung und Aufbau eines deutschen KI-Sicherheitsinstituts

Leitlinien für Erfolg und Wettbewerbsfähigkeit des DE-AISI

August 2026

Executive Summary

Die rasanten Fortschritte im Bereich hochleistungsfähiger KI-Modelle haben in kurzer Zeit eine neue Dimension technologischer Wirkmacht erreicht und haben auch Auswirkungen auf die nationale Sicherheit. Im vergangenen Jahr wurde die Cybersicherheitslage in Deutschland durch das Bundesamt für Sicherheit in der Informationstechnik (BSI) weiterhin als „angespannt“ mit „besorgniserregendem Zustand“ der Angriffsflächen beschrieben[1]. Gespiegelt wird diese Einschätzung in der deutschen Wirtschaft: Im vergangenen Jahr entstand durch Cyberangriffe ein Schaden von 202,4 Milliarden Euro[2] – ein Wert, der durch den Einsatz Künstlicher Intelligenz (KI) in den nächsten Jahren weiter steigen dürfte.

Um dieser Bedrohungslage zu begegnen, haben bereits mehrere Staaten – etwa Japan, Singapur, die USA und Kanada sowie jüngst Südkorea – hoch spezialisierte Institute gegründet. Diese Institute sind Teil eines entstehenden internationalen Netzwerks zur Koordination von Sicherheitsstandards und Forschung. Besonders Großbritanniens **AI Security Institut (AISI)** hat zuletzt Aufsehen erregt: AISI UK bekam privilegierten Zugang zu Anthropics Claude Mythos, um die Cyberangriffsfähigkeiten des Modells zu testen. Solche Zugänge sind im Hinblick auf nationale Sicherheit und Souveränität zukunftsweisend und das Ergebnis der Anerkennung von Forschungsexzellenz und Netzwerkverankerung des britischen AISI. Gleichzeitig sind sie zwingend, um Wettbewerbsfähigkeit und innere Sicherheit zu garantieren. Ohne ein eigenes solches Institut bzw. entsprechend anschlussfähige Institutionalisierung konnte Deutschland bislang nicht auf Augenhöhe Teil des International Networks for Advanced AI Measurement, Evaluation and Science sein. Daher begrüßen wir das Vorhaben der Bundesregierung, auch in Deutschland ein AISI zu gründen. Ein agil ausgestaltetes Institut bietet für Deutschland die große Chance, Talente im KI-Bereich anzuziehen und vorhandene Expertise und Infrastrukturen fokussiert einzubinden, wodurch sich Deutschland als führender Wissensstandort profilieren und bei nationalen Sicherheitsfragen zukunftsfest aufstellen kann. Die gewonnenen Erkenntnisse zahlen auf ein sicheres Innovationsökosystem ein, in dem KI perspektivisch auch in sensiblen Bereichen Anwendung finden wird und neue Märkte für Investoren und Unternehmen erschließt.

[1] BSI 2025: Die Lage der IT-Sicherheit in Deutschland 2025
[2] Bitkom 2025: Wirtschaftsschutz 2025 – Lagebild der deutschen Wirtschaft

Dr. Michael Dose | Senior Referent Innovation, Sicherheit und Technologie | T: +49 30 2028-1523 | m.dose@bdi.eu
Steven Heckler | Senior Referent Innovation, Sicherheit und Technologie | T: +49 30 2028-1523 | s.heckler@bdi.eu
Tabea Winzer | Praktikantin Innovation, Sicherheit und Technologie (bis 30.06.2026) | T: +49 30 2028-1604 | t.winzer@bdi.eu

Handlungs- und Gestaltungsempfehlungen für den Aufbau eines nationalen AI Security Institutes

Da der globale Wettbewerb auch bei Talentgewinnung nicht endet, braucht es attraktive Konditionen für den Aufbau des DE-AISI, um internationale Expertinnen und Experten mit Frontier AI Safety und Security Expertise zu gewinnen. Ebenfalls bedeutsam ist die Bereitstellung der Rahmenbedingungen, um die ihnen anvertrauten Aufgaben zu bewältigen. Um die vorgenannten Potenziale zu entfesseln, sehen wir folgende Erfolgsvoraussetzungen:

- **Klares Mandat**: Das Institut sollte sich konsequent auf die gravierendsten sicherheitsrelevanten Risiken fortgeschrittener KI-Systeme (sogenannter Frontier AI-Modelle) konzentrieren. Dazu zählen insbesondere die Analyse von Sicherheitsverstößen, Cyberoffensivfähigkeiten, Missbrauchsrisiken zur Entwicklung von Massenvernichtungswaffen, Kontrollverlustrisiken sowie Risiken der Massenmanipulation oder Manipulation wichtiger Entscheidungsträger. Nicht Teil des Mandats sollten hingegen Themen wie Datenschutz, Datenrecht oder allgemeine KI-Ethik sein, da sich hier bereits feste Strukturen auf nationaler Ebene etabliert haben. **Ziel muss ein klarer Fokus auf die schwerwiegendsten sicherheitsrelevanten Fragestellungen im Bereich KI sein**.

- **Technologische Spitzenfähigkeit in Evaluierung und Testing**: Das DE-AISI muss die Fähigkeit haben, modernste KI-Modelle unter realitätsnahen Bedingungen systematisch zu testen und zu evaluieren. Hierzu sind auch international anschlussfähige und robuste, reproduzierbare Methoden zu entwickeln, die eine objektive Bewertung von Fähigkeiten und Risiken ermöglichen. Nicht nur ist diese technische Exzellenz die Grundlage für belastbare politische Entscheidungen und wirtschaftliche Anwendungen; auch zieht das DE-AISI seine Forschungslegitimität aus der Produktion internationaler Forschungs- und Evaluationsstandards. Ebenjene **wissenschaftliche Anerkennung wird über Zugänge zu Modellen und Netzwerken entscheiden**.

- **Gewinnung internationaler Expertinnen und Experten**: Das Institut kann seine Rolle nur erfüllen, wenn es gelingt, führende internationale Fachkräfte nach Deutschland zu holen. Dafür sind wettbewerbsfähige Rahmenbedingungen und flexible Strukturen erforderlich. Die **Personalgewinnung ist keine Nebenfrage, sondern eine zentrale Erfolgsvoraussetzung**. Die Netzwerkwirkung und Forschungslegitimität des Instituts werden nicht zuletzt von der personellen Besetzung des Kernteams sowie vom klugen Andocken an vorhandene globale Talentpools wichtiger KI-Nationen wie USA und Großbritannien, Mittelmächten wie Kanada und Südkorea sowie führenden deutsch-europäischen Expertinnen und Experten für KI-Sicherheitsfragen abhängen.

- **Spitzenforschung durch Spitzenförderung**: Die Bereitstellung von ausreichenden finanziellen Mitteln wird entscheidend dafür sein, ob das DE-AISI global wettbewerbsfähig ist – als Arbeitgeber gleichwohl als Forschungsnetzwerk. Auch die Absicherung der Finanzierung über Haushaltsjahre hinweg ist dafür zu berücksichtigen. KI wird eine immer zentralere Rolle einnehmen. Dementsprechend muss eine institutionelle Auseinandersetzung in Deutschland mit KI-Sicherheitsfragen im Einklang mit den sich stellenweise exponentiell entwickelnden Anwendungsmöglichkeiten und -Szenarien im KI-Technologieumfeld zukunftsfähig („future-proof“) ausgerichtet werden. Um Souveränität und nationale Sicherheit hochzuhalten, darf nicht an Investitionen in die Cybersicherheitsarchitektur gespart werden.

- **Bevorzugter Zugang zu Rechenressourcen**: Der Trainings- und Testprozess von KI-Modellen erfordert enorme Rechenressourcen (GPUs, TPUs) und ist mit hohem Energieverbrauch verbunden. Da die Arbeit des DE-AISI der nationalen Sicherheit dienen wird, **sollte dem Institut nach britischem Vorbild ein bevorzugter Zugang zu über den Bund geförderten Rechenressourcen und im Aufbau befindlicher AI-Gigafatories garantiert werden**. Die Bereitstellung dieser Kapazitäten ist nötig, um den Frontier-Modell-Entwicklern auf Augenhöhe zu begegnen.

- **Anbieterneutraler Modellzugang und Unabhängigkeit der Bewertung**: Privilegierte Modellzugänge und Evaluationspartnerschaften des DE-AISI sollten anbieterneutral und pluralistisch ausgestaltet sein und die Unabhängigkeit der Bewertung sichern. Frontier-Entwickler sollten nicht die alleinigen Prüfer ihrer eigenen Modelle sein; die Bewertung muss frei von kommerziellen Eigeninteressen erfolgen.

- **Flexibilität und Agilität**: KI-Entwicklung und Forschung ist dynamisch und schnelllebig. Das Institut muss dies abbilden können – durch unbürokratische Verfahren sowie flexiblem und eigenständigem Personalmanagement und etwaiger Einbindung externen Sachverstands.

- **System- und Lieferkettenperspektive berücksichtigen**: Sicherheitsrelevante Fähigkeiten von Frontier-AI-Modellen entfalten ihre Wirkung häufig erst im konkreten Anwendungskontext, etwa innerhalb der Tool-Nutzung, Agentenarchitekturen, Produkteinbettung, industriellen Prozessen oder kritischen Infrastrukturen. Deshalb sollte das DE-AISI auch Methoden zur Transparenz (und Resilienz) von KI-Lieferketten, zur Bewertung von Modell-System-Interaktionen und zur anwendungsnahen Risikoübersetzung gerade mit Blick auf industrielle und aktuelle sowie künftige Physical-AI-Entwicklungen mit abdecken können. Ergänzend sollte das Institut die Absicherung von Modellgewichten vor Exfiltration sowie die Infrastruktur- und Lieferkettensicherheit fortgeschrittener KI-Entwickler nach dem Prinzip „Security by Design“ bewerten können, da der Schutz besonders leistungsfähiger Modelle unmittelbar die nationale Sicherheit betrifft.

- **Agentische KI als eigenes Evaluationsfeld**: Autonome, werkzeugnutzende und mehrstufig planende KI-Systeme („agentic AI“) verschieben das Risikoprofil grundlegend. Das DE-AISI sollte von Beginn an eine dedizierte Kompetenz zur Bewertung von Autonomiegraden, Werkzeug- und Systemzugriffen sowie Agent-zu-Agent-Interaktionen aufbauen.

- **Post-Deployment-Monitoring und Vorfallberichterstattung**: Sicherheitsrelevante Eigenschaften zeigen sich häufig erst im Betrieb. Ergänzend zur Vorabbewertung sollte ein strukturierter, vertraulicher Mechanismus zur Meldung und Auswertung sicherheitsrelevanter Vorfälle etabliert werden.

- **Offene Evaluationswerkzeuge und Benchmark-Commons**: Vom DE-AISI entwickelte Methoden, Benchmarks und Werkzeuge sollten – so weit sicherheitspolitisch vertretbar – als öffentliches Gut bereitgestellt werden, damit auch Start-ups, KMU und Wissenschaft ihre Modelle prüfen können und Sicherheitsbewertung nicht zur exklusiven Ressource weniger großer Anbieter wird.

- **Institutionelle Verankerung**: Zur Umsetzung dieser Voraussetzungen und Ziele kommt der konkreten institutionellen Ausgestaltung eine zentrale Bedeutung zu. Das **Institut muss bewusst außerhalb klassischer Behördenstrukturen angesiedelt werden**, um tarifliche Restriktionen zu vermeiden und im globalen Wettbewerb um Talente bestehen zu können.

Vorbilder hierfür sind etwa die Bundesagentur für Sprunginnovationen (SPRIND) oder GovTech Deutschland, die aufgrund ihrer Strukturen als GmbH und Verein nicht an die öffentlichen Gehaltsstrukturen gebunden sind. Auf diese Weise wird das AISI auch über Deutschlands Grenzen hinaus für internationale Expertinnen und Experten aus Forschung und Anwendung attraktiv. Eine enge Verzahnung des AISI mit weiteren etablierten Behörden, insb. zum BSI und der BNetzA, ist hingegen zwingend notwendig.

- **Forschungsauftrag ohne Aufsichtstätigkeit:** Im deutschen Kontext ist bereits heute eine Vielzahl von Instituten und Behörden mit unterschiedlichen Aspekten der KI-Sicherheit befasst. Angesichts dieser dichten Governance-Landschaft sollte ausdrücklich klargestellt werden, dass mit dem DE-AISI keine Verdoppelung bestehender Zuständigkeiten angestrebt wird. Sicherheitsinfrastrukturen sind grundsätzlich zu begrüßen; der Aufbau neuer Behörden oder behördenähnlicher Strukturen sollte jedoch nur auf Grundlage eines klar abgegrenzten Mandats und ausreichender Koordinierung erfolgen.

- **Transfer systematisch stärken nach Vorbild des UK AISI:** Das Institut sollte seine Erkenntnisse und Methoden so aufbereiten, dass die zuständigen Stellen, insbesondere das BSI, aber auch Normungsgremien und Prüforganisationen, sie sehr zeitnah in praxistaugliche Bewertungsverfahren, Leitfäden und Pilotierungen überführen können. Zusätzlich sollte das DE-AISI frühzeitig regelmäßige sowie anlassbezogene Austausch- und Konsultationsformate mit Forschung, Industrie, Behörden und Anwenderbranchen etablieren, soweit dies die Qualität seiner Evaluierungs- und Sicherheitsforschung stärkt. Transfer sollte dabei nicht erst nachgelagert erfolgen, sondern wo relevant by-design mitgedacht werden, um sicherheitsrelevante Erkenntnisse frühzeitig in die Praxis zu überführen. Dies ist besonders für Cyberrisiken in kritischen Infrastrukturen und hochregulierten Bereichen relevant, in denen oft nur die Industrie über das notwendige Anwendungs- und Domänenwissen verfügt, um reale Angriffspfade, Fehlermodi und Absicherungsmaßnahmen zu identifizieren. Gleichzeitig kann das DE-AISI so seine Brückenfunktion zwischen Forschung und Praxis stärken und einen Beitrag zur sicheren und vertrauenswürdigen Einführung von KI leisten.

- **Internationale Standards gestalten**: Angesichts globaler Technologiedynamiken muss das DE-AISI eng mit internationalen Partnern – etwa im erwähnten, wachsenden International Network of AI Safety Institutes – vernetzt sein. Es muss **aktiv zur Entwicklung gemeinsamer Standards beitragen** und europäische Harmonisierungsansätze für sichere und vertrauenswürdige KI einbringen. Damit stärkt Deutschland seine Position im internationalen Wettbewerb. Eine gestaltende Rolle bei globalen Standards ist für Souveränität sowie das Souveränitätsverständnis ebenso entscheidend wie die schlussendliche Umsetzung. Internationale Anschlussfähigkeit entsteht zudem nicht allein durch Beteiligung an Standardsetzung, sondern durch technisch belastbare, vergleichbare und prüfbare Methoden. Auf Basis entsprechender Einbeziehung seiner führenden Akteure der Qualitätsinfrastruktur, Normung sowie Prüf- und Konformitätsbewertungspraxis kann Deutschland nicht nur Standards mitaushandeln, sondern konkrete methodische, global prägende Beiträge liefern. Leitziel sollte dabei eine internationale Interoperabilität der Evaluierungen im Sinne einer gegenseitigen Anerkennung sein – „einmal geprüft, im Netzwerk anerkannt“ – um Doppelprüfungen zu vermeiden und Marktzutrittsbarrieren gerade für kleinere Anbieter zu senken.

Für den Wirtschaftsstandort Deutschland bietet ein leistungsfähiges DE-AISI, welches diese Anforderungen erfüllt, einen strategischen Vorteil: Es entsteht ein Umfeld, in dem sichere und robuste KI „Assessed in Germany“ bzw. „Trusted AI Made in Germany“ zu einem internationalen Qualitätsmerkmal

werden kann. Moderne KI kann sowohl Innovationstreiber als auch Risikoquelle sein – Deutschlands Forschungsexzellenz, Transferkompetenz und seine Netzwerkwirkung wird maßgeblich darüber entscheiden, welche dieser Möglichkeiten national schwergewichtiger ist.

Das DE-AISI dient der Stärkung des gesamten Innovationsökosystems. Es eröffnet eine neue Chance der Teilhabe am aufstrebenden KI-Markt, der längst nicht mehr allein auf Anwendung begrenzt ist. In einer zunehmend technologiegetriebenen Welt ist ein solches Institut damit ein zentraler Baustein, um wirtschaftliche Wettbewerbsfähigkeit, technologische Souveränität und geopolitische Handlungsfähigkeit sowie gesellschaftliche Sicherheit langfristig zu sichern. Die Schnelllebigkeit von KI und ihrer zugehörigen Entwicklungen verpflichtet zu zeitnaher Umsetzung des AISI-Vorhabens. Entscheidend ist, dass das Institut schnell ins operative und partnerschafsorientierte Arbeiten kommt. Das Beispiel Großbritanniens zeigt, dass ein Aufbau unter attraktiven Konditionen innerhalb weniger Monate möglich ist. Hier gilt es zeitnah nachzuziehen, denn Souveränität im digitalen Zeitalter beginnt mit Talenten, die es zu gewinnen gilt. Die Institutsbesetzung sollte nach Sicherheitsüberprüfung entsprechend ebenfalls ausländischen Expertinnen und Experten offenstehen bzw. möglichst unkompliziert durch vorhandenen deutsch-europäischen technischen Sachverstand ergänzt werden können.

Ein späterer Aufbau eines KI-Forschungsnetzwerks auf nationaler und europäischer Ebene zugunsten der Bereitstellung von Rechenkapazitäten und effizienterer Forschung ist ebenso angeraten. Das bestehende europäische AI Office, welches als Regulierungsbehörde des KI-Netzwerks etabliert wurde, ist nicht hinreichend geeignet oder ausgestattet, um den facettenreichen Bedrohungen durch Cyberangriffe und KI-modellierte Exploits zu begegnen. Mit zukunftsgerichtetem Blick ist es daher ebenso dringlich, europäische Souveränitätsbestrebungen durch eine geeignete Forschungseinrichtung zu unterfüttern. Damit würde man beginnende Fragmentierung in Europa auffangen und Mittel bündeln, sodass Europa selbst nicht um Talente buhlt.

Um rasch ein leistungsfähiges Institut etablieren zu können, ist es erforderlich, dass ab dem **Bundeshaushalt für 2027 jährlich mindestens 75 Millionen Euro für das DE-AISI** vorgesehen werden. Nur so können die internationale Wettbewerbsfähigkeit sowie die Verteidigung der Wirtschaft und Sicherung sensibler Daten sichergestellt werden.

Impressum

Bundesverband der Deutschen Industrie e.V. (BDI)
Breite Straße 29, 10178 Berlin
www.bdi.eu
T: +49 30 2028-0

Lobbyregisternummer: R000534

Redaktion
Dr. Michael Dose
Senior Referent Innovation, Sicherheit und Technologie
T: +49 30 2028-1560
m.dose@bdi.eu

Steven Heckler
Senior Referent Innovation, Sicherheit und Technologie
T: +49 30 2028-1523
s.heckler@bdi.eu

Tabea Sophie Winzer
Praktikantin Innovation, Sicherheit und Technologie (bis 30. Juni 2026)
T: +49 30 2028-1604
t.winzer@bdi.eu

BDI Dokumentennummer: D 2327