HEIDELBERGER AKADEMIE DER WISSENSCHAFTEN
Akademie der Wissenschaften des Landes Baden-Württemberg

Leitlinie zur Nutzung von Künstlicher Intelligenz in Forschung und Verwaltung

Präambel

Künstliche Intelligenz (KI) und große Sprachmodelle (Large Language Models, LLMs) verändern Forschungsprozesse und administrative Abläufe in wissenschaftlichen Einrichtungen grundlegend. Die Heidelberger Akademie der Wissenschaften erkennt das transformative Potenzial dieser Technologien. Sie fördert den verantwortungsvollen Umgang mit diesen Technologien, der sowohl Chancen nutzt als auch mögliche Risiken minimiert.

Diese Leitlinie richtet sich an alle Beschäftigten der Akademie und schafft einen verbindlichen Rahmen für den Umgang mit KI-Systemen in den Bereichen Forschung und Verwaltung. Sie orientiert sich an den Grundsätzen akademischer Integrität, Datensicherheit, Transparenz und menschlicher Verantwortung.

Die Leitlinie versteht sich als lebendes Dokument. Es wird regelmäßig überprüft und an technologische Entwicklungen und rechtliche Anforderungen angepasst. Einbezogen sind die geltenden rechtlichen Rahmenbedingungen, insbesondere das Bundesdatenschutzgesetz (BDSG), die Datenschutz-Grundverordnung (DSGVO), das Urheberrechtsgesetz (UrhG) und die KI-Verordnung der Europäischen Union (Verordnung (EU) 2024/1689).

1 Geltungsbereich

Diese Leitlinie gilt für alle Beschäftigten der Akademie, die im Rahmen ihrer dienstlichen Tätigkeit KI-Systeme einsetzen. Sie erfasst den Einsatz von KI-Tools in:

- der wissenschaftlichen Forschung (u.a. Datenanalyse, Textgenerierung, Literaturrecherche, Publikation),
- der wissenschaftsunterstützenden Verwaltung (u.a. Korrespondenz, Berichtswesen, Dokumentenbearbeitung, interne Kommunikation, Organisation von Begutachtungsverfahren).

Für KI-Systeme, die im Rahmen drittmittelfinanzierter Projekte eingesetzt werden, gelten ergänzend die Vorgaben der jeweiligen Förderinstitution. Projektleitende sind dafür verantwortlich, ihr Team über die projektspezifischen Anforderungen zu informieren.

2 Grundprinzipien

KI-Systeme sind Hilfsmittel; sie ersetzen weder das wissenschaftliche Urteilsvermögen noch die Entscheidungsverantwortung der handelnden Personen. Alle Ergebnisse aus KI-Systemen sind vor ihrer Verwendung sorgfältig zu prüfen. Die inhaltliche Verantwortung verbleibt stets bei den Nutzenden.

Der Einsatz von KI-Systemen ist transparent zu gestalten. In allen Publikationen, Berichten und formalen Dokumenten ist jede Nutzung von KI-Werkzeugen offenzulegen, die über eine sprachliche Überarbeitung bzw. Korrektur hinausgeht. Die Kennzeichnungspflicht besteht ebenso bei Abbildungen und Graphiken, die mithilfe von KI-Systemen generiert wurden. Art, Umfang und Zweck des KI-Einsatzes sind nachvollziehbar zu dokumentieren. Der Einsatz von KI-Agenten ist gesondert zu prüfen, da diesen ggf. weitreichende Zugriffe auf IT-Infrastrukturen eingeräumt werden muss. Dies ist mit dem Referat Wissenschaft und Digitale Infrastruktur jeweils im Vorfeld abzuklären.

Der Einsatz von KI darf die Grundsätze guter wissenschaftlicher Praxis nicht unterlaufen. Insbesondere ist die Verwendung von KI-generierten Inhalten als eigene intellektuelle Leistung ohne entsprechende Kennzeichnung unzulässig. Die Leitlinie der Akademie und der Deutschen Forschungsgemeinschaft (DFG) zur guten wissenschaftlichen Praxis gelten in Verbindung mit dieser Leitlinie.

Beschäftigte sind sensibilisiert dafür, dass KI-Systeme systematische Verzerrungen (Bias) enthalten und diskriminierende Inhalte erzeugen können. Ergebnisse sind vor Verwendung kritisch zu prüfen, insbesondere wenn sie Grundlage für wissenschaftliche Bewertungen sind.

3 KI in der Forschung

Generative KI-Tools dürfen unterstützend in der Forschung eingesetzt werden:

- Literaturrecherche, Zusammenfassung und Erschließung wissenschaftlicher Quellen;
- Analyse und Strukturierung großer Text- und Datenmengen (Text- und Data Mining, zu Forschungszwecken ggf. auch geschütztes, anonymisiertes Material);
- Sprachliche Überarbeitung von Textentwürfen (Formulierung, Übersetzung);
- Unterstützung bei der Erstellung von Abbildungen, u.a. Visualisierungen von Forschungsdaten;
- Unterstützung bei der Anforderungsanalyse, Konzeption und dem Design von Anwendungen und Softwarearchitekturen

- Programmierung, Code-Optimierung, Refactoring und Code-Dokumentation, Entwicklung und Durchführung von Tests, Validierung von Software, Debugging, Überprüfung von Code auf Fehler, Sicherheitslücken und potenzielle Schwachstellen
- Vorbereitung von Berichten und Präsentationen.

Bei der Einreichung wissenschaftlicher Publikationen ist außerdem die Nutzung von KI entsprechend den Richtlinien der jeweiligen Fachzeitschrift oder Förderinstitution offenzulegen. KI-Systeme dürfen nicht als Autorinnen oder Autoren ausgewiesen werden; die inhaltliche Verantwortung verbleibt bei den Forschenden.

4 KI in der Verwaltung

KI-Systeme können in der Verwaltung eingesetzt werden für:

- Textformulierung, -überarbeitung und Übersetzung in der Korrespondenz;
- Erstellung von Berichten, Protokollen und Präsentationen;
- Unterstützung bei internen Recherchen und Zusammenfassungen;
- Automatisierung standardisierter, nicht vertraulicher Verwaltungsprozesse.
- zur Bildgenerierung und Textüberarbeitung in der Presse- und Öffentlichkeitsarbeit

5 Unzulässige Anwendungen

KI-Systeme dürfen nicht zur vollständigen Erstellung von Forschungsergebnissen und wissenschaftlicher Arbeiten oder zur autonomen Durchführung von Peer-Reviews verwendet werden. Sollten mit KI-Anwendungen neue Datensätze generiert werden, die auf Informationen basieren, die nicht als Forschungsergebnisse im engeren Sinne gelten, müssen diese entsprechend gekennzeichnet werden. KI-generierte Hypothesen dürfen nicht ungeprüft als eigene Leistungen ausgegeben werden.

Personenbezogene Daten, vertrauliche Projektinformationen, unveröffentlichte Ergebnisse, interne Dokumente sowie urheberrechtlich geschütztes Material dürfen nicht in KI-Systeme eingegeben werden, die diese Daten dauerhaft speichern, an Dritte weitergeben oder zu Trainingszwecken verwenden. Die für die Verarbeitung notwendige temporäre Speicherung im Rahmen eines Auftragsverarbeitungsvertrags (AVV) ist hiervon ausgenommen. Dies schließt Zugangsdaten, Projektdetails und strategische Informationen ein. Bei Unsicherheit über die Vertraulichkeit von Informationen oder die Zuverlässigkeit des Anbieters ist im Zweifel von einer Eingabe abzusehen. Für die Klassifikation von Informationen gelten die Vertraulichkeitsstufen gemäß Anlage 1 dieser Richtlinie.

Die Systeme dürfen nicht zur Umgehung von Urheberrechten oder zur Manipulation von Daten eingesetzt werden. Auch die Bewertung wissenschaftlicher Leistungen durch KI mit Folgen für die verantwortlichen Personen (u.a. Peer Review, vertrauliche Antragsunterlagen, Originalgutachten) ist nicht gestattet.

Nicht zulässig ist der Einsatz von KI-Systemen für Personalentscheidungen (Einstellung, Beurteilung, Kündigung) ohne menschliche Überprüfung sowie für Entscheidungen, die rechtlich erhebliche Auswirkungen auf Personen haben. Der Einsatz von KI zur Überwachung oder Verhaltensanalyse von Beschäftigten ist untersagt.

Der Einsatz von KI-Systemen ist in allen Fällen verboten, in denen die DSGVO verletzt würde. Verboten sind alle KI-Praktiken gemäß Artikel 5 der Verordnung (EU) 2024/1689 (u.a. unterschwellige Beeinflussung, biometrische Kategorisierung, Emotionserkennung am Arbeitsplatz).

5 Zugelassene KI-Systeme

Beschäftigte nutzen vorrangig KI-Systeme, die von der Akademie datenschutzkonform bereitgestellt oder geprüft und für den jeweiligen Einsatzzweck freigegeben worden sind. Diese umfassen lokal installierte, vollständig DSGVO-konforme Instanzen sowie spezifische cloudbasierte LLMs, die nachweislich den Datenschutzanforderungen entsprechen. Forschungsprojekte, die eine eigenständige IT betreiben, sind an die Vorgaben dieser Richtlinie gebunden und dazu angehalten, das Referat Wissenschaft und digitale Infrastruktur bei der Einführung von KI-Anwendungen zu informieren. Systemtypen nach Schutzniveau sind in Anlage 1 zusammengefast.

7 Verantwortung im Umgang mit KI

Die Akademie fördert die KI-Kompetenz aller Beschäftigten durch Informationsangebote und Schulungen. Beschäftigte, die KI-Systeme im dienstlichen Kontext nutzen, sind verpflichtet, vor der erstmaligen Nutzung eine geeignete Schulung zu absolvieren.

Verstöße gegen diese Richtlinie sind unmittelbar der/dem zuständigen Vorgesetzten und, soweit personenbezogene Daten betroffen sind, der/dem Datenschutzbeauftragten zu melden. Je nach Schwere des Verstoßes können arbeitsrechtliche Konsequenzen bis hin zur Kündigung folgen. Bei Datenschutzverletzungen sind die gesetzlich vorgeschriebenen Meldepflichten zu beachten.

8 Inkrafttreten

Diese Leitlinie tritt mit Beschluss des Vorstands in Kraft. Sie ersetzt alle bisherigen internen Regelungen zum Einsatz von KI-Systemen.

Beschlossen durch den Vorstand am 16.11.2026

Nachweis der Quellen:

Für die Materialrecherche, den sprachlichen Erstentwurf und die Strukturierung dieses Textes wurde das KI-Tool Claude Sonnet 4.6 (Anthropic) verwendet.

Weitere Quellen:

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW), KI-Richtlinie Version 1, in Kraft gesetzt am 17. Juni 2025; edoc.bbaw.de/frontdoor/index/index/docId/5105.

DFG, Leitlinien zur Sicherung guter wissenschaftlicher Praxis: https://zenodo.org/records/14281892

DFG, Leitlinie zu KI in der Begutachtung: https://www.dfg.de/de/grundlagen-themen/digitale-themen/ki/begutachtung#388878.

DFG, Stellungnahme des Präsidiums zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG (September 2023); dfg.de/resource/blob/289674/ki-230921-stellungnahme-praesidium.pdf.

EU-KI-Verordnung (AI Act), besonders Transparency obligations (Art. 50): https://ai-act-service-desk.ec.europa.eu/de/ai-act/article-50.

EU Icons for labelling AI-generated content: https://digital-strategy.ec.europa.eu/en/policies/eu-icons-labelling-ai-generated-contentture.

Universität Freiburg, Policy zum Umgang mit generativer KI in der Forschung (2026); uni-freiburg.de/forschung/qualitaetssicherung/gute-wissenschaftliche-praxis/policy-zum-umgang-mit-generativer-ki-in-der-forschung/.

Universität Freiburg, Nutzungsbedingungen Open WebUI; openwebui.uni-freiburg.de/nutzungsbedingungen/.

Universität Mannheim, Leitlinie für den Einsatz von KI | Universität Mannheim: https://www.uni-mannheim.de/universitaet/profil/digitalisierung/kuenstliche-intelligenz/ki-richtlinie/.

Universität Stuttgart, Whitelist für KI-Tools; https://www.beschaeftigte.uni-stuttgart.de/uni-services/it-kommunikation/nutzung-von-ki-systemen/whitelist-ki-tools/.

Universität Stuttgart, Richtlinie zur Nutzung von KI und zum Einsatz von KI-Systemen; beschaeftigte.uni-stuttgart.de/document/dokument_it/richtlinie-KI-nutzung-beschaeftigte.pdf.

Universität Tübingen, Leitlinie zum Umgang mit generativen KI-Tools (wird kontinuierlich weiterentwickelt); uni-tuebingen.de/lehrende/generative-ki-in-lehre-und-forschung/leitlinie-der-universitaet-tuebingen/.

Anhang 1: Vertraulichkeitsklassen für Informationen im Kontext des Einsatzes von KI-Systemen

Kürzel	Klassifizierung	Farbcode	Beschränkung	● Beispiele Verwaltung	▲ Beispiele Forschung	Bereichsübergreifend	Schadenspotenzial	Personenbezogene Daten	Regeln für KI-Systeme (Vertrag, Hosting, Training, Accountbindung)	Beispiele für KI-Tools
C1	Öffentlich	weiss	Darf ohne Einschränkung geteilt werden, auch mit der Öffentlichkeit.	• Marketing- und PR-Material (Flyer, Broschüren, Pressemitteilungen, Abbildungen) • Öffentliche Informationen auf der Website der Einrichtung	• Öffentlich zugängliche Forschungs- und Publikationsinformationen (z. B. Abstracts, Vorträge, Konferenzbeiträge) • vollständig anonymisierte Testdatensätze • entsprechend lizensiertes Material, z.B. CC BY-NV 4.0	• Übersetzungen • sprachliche Überarbeitungen • Erstellung von graphischen Darstellungen (auszeichnungspflichtig)	Keine Auswirkungen auf Betroffene, Prozesse, Systeme oder Einrichtungen; keine Folgen für die Akademie.	Personenbezogene Daten, die von den Betroffenen frei zugänglich gemacht wurden.	Nutzung ohne Einschränkung erlaubt. Hosting in der Cloud (EU und Nicht-EU) sowie On-Premise zulässig. Training durch den Anbieter akzeptabel. Nutzung eines privaten Accounts nach eigenem Ermessen.	ChatGPT Free / Plus; MS Copilot mit privatem Account
C2	Intern	grün	Darf nur innerhalb der Akademie und gegebenenfalls mit Partnern geteilt werden.	• Interne Rundschreiben und organisatorische Dokumente der Fachbereiche ohne vertrauliche Personenbezüge	• Erste konzeptionelle Forschungsansätze (ohne sensible Details)	• Urheberrechtlich geschütztes Material	Sehr geringe bis geringe Auswirkungen; minimale bis kaum spürbare Folgen für die Akademie.	Daten, deren unsachgemäße Handhabung keine besondere Beeinträchtigung erwarten lässt, die aber nicht frei zugänglich gemacht wurden.	Nutzung erlaubt, sofern die Eingaben nicht für Trainingszwecke verwendet werden. Hosting On-Premise oder in der EU-Cloud zulässig; Nicht-EU-Clouds nur mit vertraglicher Absicherung (AVV). Nutzung eines privaten Accounts nicht empfohlen.	GWDG Open AI mit Account; bereitgestellte Zugänge, bspw. eines KI-Aggregators
C3	Vertraulich	gelb	Darf nur innerhalb der Akademie und gegebenenfalls mit Partnern nach dem Need-to-know-Prinzip geteilt werden.	• Personaldaten (z. B. Gehaltsinformationen, Bewerbungen) • Interne Protokolle von Gremien und Kommissionen mit sensiblen Themen • vertauliche Antrags- oder Begutachtungsunterlagen	• Noch unveröffentlichte Forschungsdaten • Training eigener Modelle zu Forschungszwecken		Mäßige bis erhebliche Auswirkungen; wahrnehmbare bis ernsthafte Folgen für die Akademie.	Daten, deren unsachgemäße Handhabung die Betroffenen in gesellschaftlicher Stellung oder wirtschaftlichen Verhältnissen beeinträchtigen könnte („Ansehen“).	Nutzung nur mit vertraglicher Zusicherung, dass Eingaben nicht für Trainingszwecke verwendet werden. Hosting nur On-Premises oder in der EU-Cloud mit starker vertraglicher Absicherung (kein Zugriff Dritter, Verschlüsselung bei Übertragung und Speicherung, kein Zugriff Dritter, keine menschliche Einsicht). Nicht-EU-Clouds nicht zulässig. Zugang intern auf Akademiesangehörige beschränkt. Nutzung eines privaten Accounts nicht gestattet.	bereitgestellte Zugänge mit AAV bei kommerziellen Anbietern; Akademie-eigenes KI-Tool mit hochschulnah gehosteten Modellen
C4	Streng vertraulich	rot	Darf nur innerhalb einer explizit definierten Gruppe geteilt werden.		• Medizinische oder psychologische Daten von Proband*innen und Patient*innen • Details zu laufenden Drittmittelprojekten mit strengem Patentschutz, Verschwiegenheitsvereinbarungen oder hohen Reputationsrisiken • Geschäftsgeheimnisse in kooperativen Forschungsprojekten mit Unternehmen (z. B. geheime technische Daten, Prototypen, neue Verfahren)		Große bis sehr große Auswirkungen; schwerwiegende bis katastrophale Folgen für die Akademie.	Daten, deren unsachgemäße Handhabung Gesundheit, Leben oder Freiheit beeinträchtigen oder die gesellschaftliche Stellung bzw. wirtschaftlichen Verhältnisse erheblich beeinträchtigen könnte („Existenz“).	Nur auf isolierten On-Premises-Systemen mit vollständiger Kontrolle. Cloud-Nutzung (auch EU) in der Regel nicht erlaubt. Eingaben dürfen unter keinen Umständen für Training verwendet werden; vertragliche Absicherung zwingend. Interne Nachvollziehbarkeit erwünscht oder notwendig, z. B. um Missbrauch zu verhindern.	Lokale Open-Source-Modelle wie LLaMA, Mistral (momentan nicht verfügbar)

Ursprüngliche Quelle: Universität Stuttgart, Vertraulichkeitsklassen für Informationen im Kontext des Einsatzes von KI-Systemen; https://www.beschaeftigte.uni-stuttgart.de/document/dokument_it/vertraulichkeitsklassen-KI-systeme.pdf. Diese Übersicht wurde am 28. Oktober 2026 mit Unterstützung von Claude Sonnet 5.5 (Anthropic) erstellt und von DSB/Ref. Wiss.&Dig.Infrastruktur redaktionell geprüft.