



Leitlinien für den Einsatz generativer KI im TAB

1 Einleitung

Generative KI-Tools auf Basis von Large Language Models (z.B. GPT-5) oder Diffusionsmodellen (z.B. Stable Diffusion) werden seit den letzten Jahren verstärkt als Hilfsmittel in der Recherche, für das Erstellen von Texten und Bildern sowie von Audio- und Videomaterial eingesetzt.

Im TAB verfolgen wir das Ziel, künstliche Intelligenz verantwortungsvoll, sicher und reflektiert einzusetzen. Daher sind in den folgenden Leitlinien genauere Vereinbarungen für die Nutzung von generativer KI im gesamten Lebenszyklus der wissenschaftlichen Produkte des TAB, von der Konzeption bis zur Ergebnisvermittlung einschließlich der institutionellen Kommunikation, formuliert. Die Leitlinien sollen der Qualitätssicherung der Arbeit des TAB bei Nutzung von generativer KI dienen und umfassen auch die vom TAB in Auftrag gegebenen Arbeiten (z.B. Gutachten, Übersetzungen, Grafiken).

Diese Leitlinien sind im Rahmen weiterer, teilweise verbindlicher bzw. übergeordneter Leitlinien und rechtlicher Rahmenbedingungen, z.B. der Verordnung (EU) 2024/1689¹ (KI-Verordnung) zu verstehen. Für die Wissenschaft gibt es bereits Handlungsempfehlungen für die Verwendung generativer KI, beispielsweise von der Europäischen Kommission (EK 2024), von der Deutschen Forschungsgemeinschaft (DFG 2023), von der Leopoldina (2023), der Helmholtz-Gemeinschaft (2024), dem KIT (2025) sowie von Verlagen und Zeitschriften (Ganjavi et al. 2024). Da sich diese grundsätzlich auf die Einhaltung wissenschaftlicher Standards bei Verwendung von generativer KI beziehen, sind sie auch für die Arbeit des TAB von Bedeutung. Die Grundprinzipien wissenschaftlicher Integrität, wie sie im Kodex der DFG (2019) formuliert sind, gelten selbstverständlich auch für das wissenschaftliche Arbeiten mit generativer KI.

Darüber hinaus nutzen wir bevorzugt Tools, die den Ethikleitlinien für vertrauenswürdige KI der Europäischen Kommission (EK 2019) entsprechen. Die Nutzung von KI-Tools setzt ausreichende Kompetenzen voraus (Art. 4 KI-Verordnung), daher sollten die Nutzer/innen regelmäßig Weiterbildungen absolvieren, die mindestens das Bewusstsein für Dokumentationspflichten, Datenhygiene, Datenschutz gemäß Datenschutz-Grundverordnung² (DSGVO) und die Risikoabwägung bei der Anwendung generativer KI vermitteln.

Anwendungsfälle für die Nutzung generativer KI

Die Anwendungsfälle generativer KI in der TA sind vielfältig und reichen von KI als Hilfsmittel in der Recherche bis zum Verfassen von Texten. Für die Arbeit des TAB sind insbesondere folgende Anwendungsfälle relevant:

- Die explorative **Wissensrecherche** umfasst Anwendungsfälle, bei denen generative KI als Hilfsmittel eingesetzt wird, um Wissensfragen zu stellen, insbesondere:

¹ Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)

² Verordnung (EU) 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)

- Recherche in wissenschaftlichen Datenbanken mittels generativer KI,
- Recherche mit Tools, die auf allgemein verfügbare Onlineinhalte zurückgreifen bzw. mit diesen trainiert wurden.
- Qualitative **Inhaltsanalysen** umfassen alle Anwendungen, bei denen mittels generativer KI bestimmte Textkorpora (bzw. Datensets) analysiert werden, insbesondere:
 - Beantwortung von Fragen zu einem Datensatz auf der Basis von Prompts,
 - automatische Annotation und Kodierung von Texten,
 - Identifikation von Themen im Datenset (beispielsweise durch Topic Modeling),
 - automatisierte Bewertung von einzelnen Texten/Textkorpora nach vorgegebenen Kriterien.
- **Ideengenerierung** umfasst den Einsatz von generativer KI für kreative Aufgaben wie z.B.:
 - Erstellung von Personas zur Repräsentation bestimmter Personengruppen und
 - Erstellung von Zukunftsszenarien,
- Bei der **Datenauswertung** kann generative KI insbesondere bei unstrukturierten Daten Unterstützung leisten, indem z.B.
 - Themen oder Haltungen in Textkorpora analysiert werden oder
 - Programme für die automatisierte Auswertung von Daten erstellt oder angepasst werden.
- Weitere Anwendungsfälle, bei denen generative KI als Hilfsmittel für das **Programmieren** herangezogen wird, sind insbesondere:
 - Verfassen von Code inklusive Formeln/Skripte für MS Excel und andere Anwendungen sowie
 - Korrigieren von erstelltem Code.
- Die **Ergebniskommunikation** (Text-, Bild- und Audioproduktion) umfasst alle Anwendungen, bei denen mittels KI-Tools Inhalte auf der Basis von Prompts erstellt werden, insbesondere:
 - Formulierung von neuen Texten (auch auf der Basis eigener Notizen),
 - Korrektur bzw. Umformulierung von bestehenden Texten,
 - Erstellung von Zusammenfassungen längerer Texte,
 - Übersetzungen und Transkriptionen,
 - Generierung und Bearbeitung von Bild- und/oder Audiomaterial.

Vereinbarungen zur regelmäßigen Überarbeitung der Leitlinien

Da die Entwicklung der Tools und Anwendungsfälle sehr dynamisch ist, werden diese Leitlinien regelmäßig überprüft und bei Bedarf überarbeitet.

2 Leitlinien

Verantwortung für die Verwendung von Ergebnissen

- Die volle Verantwortung für die Verwendung von KI-generierten Ergebnissen liegt bei den Personen, die diese Inhalte verwenden. Nur Menschen können Autor/innen sein. Die Ergebnisse bzw. Outputs der Tools sind stets zu überprüfen. Insbesondere bei Inhalten, die zur Veröffentlichung bestimmt sind, und bei der Nutzung für andere wichtige Entscheidungen darf KI nicht die einzige Quelle von Informationen sein.
- Dies betrifft insbesondere die Verantwortung für die Richtigkeit und Ausgewogenheit der Inhalte, die Verantwortung für Plagiate und die Einhaltung von Rechten bezüglich geistigen Eigentums.

Offenlegung der Verwendung

- In internen wie zu publizierenden TAB-Ergebnissen soll die Nutzung von KI-Tools in einer dem Format angemessenen Weise offengelegt werden. Dies kann in der Einleitung, im Methodenkapitel, in einer Fußnote oder in einem Verzeichnis am Ende der Publikation erfolgen, die genaue Form und die Position der Offenlegung liegen in der Verantwortung der Verfasser/innen, sofern keine formellen Vorgaben für ein Format bestehen. Formulierungsbeispiel (hier sind in kursiver Schrift Beispielanwendungen genannt; weitere, hier nicht genannte sollten spezifisch für die jeweilige Arbeit hinzugefügt werden):

Zur Erstellung dieses Berichts wurden generative KI-Tools verwendet, und zwar [Name des Tools] für [das Formulieren, die Korrektur bzw. Umformulierung von Texten], [Name des Tools] für [die Erstellung von Zusammenfassungen], [Name des Tools] für [die Übersetzung von Texten] und [Name des Tools] für [die Generierung von Bild- und Audiomaterial]. Alle KI-generierten Inhalte wurden sorgfältig überprüft und bearbeitet, um Richtigkeit und Ausgewogenheit sicherzustellen. Die Nutzung erfolgte im Rahmen der TAB-Leitlinien für generative KI. Die endgültige Verantwortung für die inhaltliche Richtigkeit, die kritische Reflexion und die Interpretation der Ergebnisse liegt bei den Autor/innen.

Diese Offenlegung betrifft nicht die Nutzung von Tools zur Überprüfung von Grammatik, Rechtschreibung oder Ähnlichem.

- Bei Bildern, Videos und Audiomaterial ist in der Quellenangabe anzugeben, dass es sich um von KI generierte bzw. manipulierte Inhalte handelt.
- Bei Übersetzungen und Transkriptionen ist ein Hinweis hinzuzufügen, dass diese maschinell oder mithilfe von KI-basierter Software erstellt wurde.

Qualitätssicherung

- Die Ergebnisse bzw. Outputs der Tools sind systematisch insbesondere mit Blick auf Richtigkeit, Aktualität, Kohärenz, Ausgewogenheit, Präzision und etwaige Verzerrungen beispielsweise durch Prompting zu überprüfen. Outputs sind stets durch Triangulation zu prüfen.
- Bei Übersetzungen ist der übersetzte Text durch Personen zu überprüfen, die mit dem Thema und der Sprache vertraut sind.

Daten- und Urheberrechtsschutz

- Nutzer/innen tragen Verantwortung für die Inhalte, die dem KI-Tool zur Verfügung gestellt werden, und für potenzielle Datenschutzverletzungen.
- Personenbezogene Inhalte dürfen nur in DSGVO-konforme Tools eingegeben oder hochgeladen werden.
- Vertrauliche Arbeitsinhalte (z.B. Zwischenergebnisse für Publikationen, unveröffentlichte Gutachten, selbst erstellte Datensets oder nicht öffentlich zugängliche Daten) speisen wir nur in Lösungen ein, die DSGVO-konform sind, Daten nicht langfristig speichern und Daten nicht für Trainingszwecke nutzen.
- Urheberrechtlich geschütztes Material (Text, Bild etc.) wird nur eingegeben, wenn zuvor die Erlaubnis der/s Urheberin/s eingeholt wurde.
- Tools für explorative Recherchen können unabhängig davon genutzt werden, wo und wie lange die Daten gespeichert werden, da lediglich einzelne Wissensfragen an das Tool gestellt werden.
- Für die Veröffentlichung generierter Bildinhalte sollte nur auf Anwendungen zugegriffen werden, die mit lizenzierten Inhalten, frei lizenzierten Inhalten und Public-Domain-Inhalten trainiert sind, deren Urheberrechte abgelaufen sind.

Nachhaltigkeit

- Bei der Toolauswahl sollen Informationen zu ökologischen und sozialen Auswirkungen geprüft werden und Tools mit möglichst geringen Auswirkungen bevorzugt werden.
- Vor dem Einsatz generativer KI soll sorgfältig abgewogen werden, inwiefern der Einsatz generativer KI für die Aufgabe einen eindeutigen Mehrwert mitbringt und welches Tool sich zur Durchführung der Aufgabe am besten eignet. Prompts sollen sorgfältig verfasst und möglichst präzise formuliert werden, um unnötige zusätzliche Anfragen zu vermeiden.
- Die kontinuierliche Weiterbildung zu generativer KI soll auch deren soziale und ökologische Auswirkungen sowie Strategien für einen möglichst ressourcenschonenden Einsatz umfassen.

Nutzungsausschluss von Anwendungsfällen bzw. Tools

- Die ungeprüfte Übernahme von KI-generierten Texten ist ausgeschlossen. Unveränderte Abschnitte sind im Text zu kennzeichnen und die Verwendung von KI-Tools ist wie unter „Offenlegung der Verwendung“ beschrieben anzugeben.
- KI-Tools, denen Zugriff auf persönliche, durch Passwort geschützte Dienste zur stellvertretenden Durchführung von Tätigkeiten gestattet wird, sind selbst bei Nutzung mit nicht personalisierten Daten vorerst ausgeschlossen. Eine erneute Prüfung erfolgt, sobald die Ausgestaltung der menschlichen Aufsicht in solchen Systemen (Art. 14 KI-Verordnung) weiter fortgeschritten ist.
- Generative KI darf nicht zur automatisierten Entscheidungsfindung etwa in der Bewertung von Gutachten, Angeboten oder Berichten herangezogen werden. Entsprechende Bewertungsprozesse erfordern menschliches Urteilsvermögen und die Übernahme von Verantwortung. Daher sollte generative KI nur für nicht substanzielle Aspekte solcher Bewertungs- und Beurteilungsprozesse verwendet werden, z.B. im Rahmen von ergänzenden Recherchen.

3 Literatur

- DFG (2023): Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG. Deutsche Forschungsgemeinschaft, <https://www.dfg.de/resource/blob/289674/ff57cf46c5ca109cb18533b21fba49bd/230921-stellungnahme-praesidium-ki-ai-data.pdf> (6.2.2025)
- DFG (2019): Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Stand 2024, Deutsche Forschungsgemeinschaft, DOI: 10.5281/zenodo.14281892
- EK (2019): Ethikleitlinien für vertrauenswürdige KI. Europäische Kommission, <https://digital-strategy.ec.europa.eu/de/library/ethics-guidelines-trustworthy-ai> (4.4.2025)
- EK (2024): Living guidelines on the Responsible Use of Generative AI in Research. Europäische Kommission, https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf (6.2.2025)
- Ganjavi, C. et al. (2024): Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. In: BMJ 384, Art: e077192, <https://doi.org/10.1136/bmj-2023-077192>
- Helmholtz-Gemeinschaft (2024): Handlungsempfehlungen zum Einsatz von Künstlicher Intelligenz. Helmholtz-Gemeinschaft Deutscher Forschungszentren, https://www.helmholtz.de/assets/helmholtz_gemeinschaft/Downloads/Helmholtz_Handlungsempfehlungen_zum_Einsatz_von_KI_Version_1.0.pdf (6.2.2025)
- KIT (2025): Leitlinien zum Einsatz generativer KI am KIT. Karlsruher Institut für Technologie, <https://www.kit.edu/downloads/KI-Leitlinien-de.pdf> (14.8.2025)
- Leopoldina (2023): Leitlinien zur Nutzung von großen Sprachmodellen bei Veröffentlichungen der Leopoldina. Deutsche Akademie der Naturforscher Leopoldina – Nationale Akademie der Wissenschaften, https://www.leopoldina.org/fileadmin/redaktion/Ueber_uns/2023_Leitlinien_zur_Nutzung_von_gro%C3%9Fen_Sprachmodellen.pdf (6.2.2025)

Herausgeber

Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag
Neue Schönhauser Straße 10
10178 Berlin

Telefon: +49 30 28491-0

E-Mail: buero@tab-beim-bundestag.de

www.tab-beim-bundestag.de

2025

DOI: 10.5445/IR/1000188419

Das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB) berät das Parlament und seine Ausschüsse in Fragen des wissenschaftlich-technischen Wandels. Das TAB wird seit 1990 vom Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie (KIT) betrieben. Grundlage ist ein Vertrag mit dem Deutschen Bundestag. Hierbei kooperiert es seit September 2013 mit dem IZT – Institut für Zukunftsstudien und Technologiebewertung gGmbH sowie dem Institut für Innovation und Technik (iit) in der VDI/VDE Innovation + Technik GmbH.