

## FAQ KÜNSTLICHE INTELLIGENZ UND GUTE WISSENSCHAFTLICHE PRAXIS

Dieses FAQ versammelt Fragen, die uns häufig im Zusammenhang mit künstlicher Intelligenz (KI) und guter wissenschaftlicher Praxis (GWP) erreichen. Die Antworten sollen bei der Orientierung in einem schnelllebigen Thema helfen, ohne dabei präskriptiv zu sein. Sie stellen keine offizielle Positionierung des *Ombudsman für die Wissenschaft* (OfdW) dar, sondern beschreiben den Status Quo und ordnen bereits bestehende Empfehlungen aus Sicht der GWP ein, identifizieren Lücken und verweisen auf weiterführende Literatur. Dieses FAQ richtet sich primär an Forschende. Für die Nutzung von KI in der Lehre und in studentischen (Qualifikations-)Arbeiten sind i.d.R. universitäre KI-Richtlinien, angepasste Prüfungsordnungen und Selbstständigkeitserklärungen sowie Entscheidungen individueller Lehrpersonen maßgeblich. Daher werden eventuelle Besonderheiten von KI in der Lehre und in Prüfungsangelegenheiten in diesem FAQ nicht besprochen.

Zusammengestellt und verfasst von Katrin Frisch

Stand: 6. November 2024

## Inhalt

Zu welchen Aspekten hinsichtlich KI und GWP besteht Konsens? .....	2
Warum sollte die Nutzung von KI angegeben werden? .....	2
Was genau bedeutet „die Nutzung von KI transparent und angemessen angeben“? .....	3
Wo im Manuskript muss ich die Nutzung von KI angeben? .....	3
Gibt es Zitiervorschläge für die Angabe von KI? .....	3
Müssen Prompts dokumentiert werden? .....	3
Was ist mit der Angabe der Nutzung von KI für Software? .....	4
Muss ich die Nutzung von KI angeben, wenn ich diese nur nutze um meinen Text stilistisch/sprachlich zu verbessern? .....	4
Muss ich die Nutzung von KI für Übersetzungen angeben? .....	4
Muss ich angeben, dass ich KI nur zur Inspiration genutzt habe? .....	4
Welche Anwendungen fallen unter den Begriff KI? .....	5
Darf ich KI für das Schreiben von Anträgen nutzen? .....	5
Darf KI zum Generieren von Bildern/Graphiken genutzt werden? .....	5
Darf KI in der Peer-Review eingesetzt werden? .....	5
Was ist bei der Nutzung von KI in Autorschaftsteams zu beachten? .....	6
Was ist der Zusammenhang von KI-generierten Texten und Plagiaten? (Kann ich durch die Nutzung von KI aus Versehen andere Texte plagiiieren?) .....	6
Auf welche Schwächen und Risiken von KI sollten Forschende achten? .....	6
Kann die Nutzung von KI in einem Text durch Software-Tools nachgewiesen werden? Wie erkenne ich KI generierte Texte? .....	7
Ich bin Reviewer:in oder Editor:in und vermute, ein Text bzw. Teile davon wurden mithilfe einer KI geschrieben, diese Nutzung ist jedoch nicht transparent angegeben. Was sollte ich tun? .....	7
Was passiert, wenn ich beschuldigt werde KI genutzt zu haben, ohne dies in der Publikation anzugeben? .....	7
Gibt es aus GWP-Sicht besonders empfehlenswerte KI-Anwendungen? .....	7
Wie ist der Zusammenhang zwischen Urheberrecht und KI-generierten Inhalten? .....	8
Wie kann ich Konflikte, die aus KI-Nutzung resultieren könnten, vorbeugen? .....	8
Welche KI-Richtlinien sind für mich maßgeblich? .....	8
Warum gibt es keine klaren Empfehlungen vom <i>Ombudsman für die Wissenschaft</i> zum Thema KI und GWP? .....	8
Referenzen .....	9
Literatur .....	9
Policies, Handreichungen, Empfehlungen .....	10

## 1. Zu welchen Aspekten hinsichtlich KI und GWP besteht Konsens?

In Richtlinien und Empfehlungen zu KI hat sich seit der Veröffentlichung von ChatGPT und anderer generierender KI für ein größeres Publikum zu zwei Aspekten ein Konsens herausgebildet.

- 1) KI kann nicht als Autorin fungieren. Dies wird damit begründet, dass KI keine Verantwortung für die Inhalte eines Manuskriptes übernehmen und auch nicht der finalen Fassung zustimmen kann. Beides wird von menschlichen Autor:innen gem. GWP grundsätzlich gefordert.
- 2) Die Nutzung von KI muss transparent und angemessen im Manuskript angegeben werden.

Die genaue Umsetzung der transparenten Angabe von KI wird jedoch teils unterschiedlich ausgelegt bzw. oft nicht genauer definiert.

In der folgenden Kurzübersicht findet sich eine Zusammenfassung der Editorial Policies in Bezug auf KI der großen Verlage und verwandter Organisationen (Stand 6. November 2024)

	KI und Autorschaft	Angabe der Nutzung von KI	KI-generierte Abbildungen	KI und Peer-Review	Link zur KI-Policy
<b>Elsevier</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	sehr eingeschränkt erlaubt	nicht erlaubt	<a href="https://www.elsevier.com/de-de/about/policies-and-standards/generative-ai-policies-for-journals">https://www.elsevier.com/de-de/about/policies-and-standards/generative-ai-policies-for-journals</a>
<b>ICMJE</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	keine Angabe	sehr eingeschränkt erlaubt	<a href="https://www.icmje.org/icmje-recommendations.pdf">https://www.icmje.org/icmje-recommendations.pdf</a>
<b>Science</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	sehr eingeschränkt erlaubt	nicht erlaubt	<a href="https://www.science.org/content/page/science-journals-editorial-policies#image-and-text-integrity">https://www.science.org/content/page/science-journals-editorial-policies#image-and-text-integrity</a>
<b>Springer Nature</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	sehr eingeschränkt erlaubt	sehr eingeschränkt erlaubt	<a href="https://www.springernature.com/gp/policies/editorial-policies">https://www.springernature.com/gp/policies/editorial-policies</a>
<b>Taylor &amp; Francis</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	nicht erlaubt	sehr eingeschränkt erlaubt	<a href="https://taylorandfrancis.com/our-policies/ai-policy/?_ga=2.202881233.1550617214.1729762987-990921615.1729762987">https://taylorandfrancis.com/our-policies/ai-policy/?_ga=2.202881233.1550617214.1729762987-990921615.1729762987</a>
<b>Wiley</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	keine Angabe	sehr eingeschränkt erlaubt	<a href="https://authorservices.wiley.com/ethics-guidelines/index.html#22">https://authorservices.wiley.com/ethics-guidelines/index.html#22</a>
<b>WAME</b>	KI ≠ Autorin	Nutzung muss im Manuskript angegeben werden	keine Angabe	sehr eingeschränkt erlaubt	<a href="https://wame.org/page2.php?id=106">https://wame.org/page2.php?id=106</a>

## 2. Warum sollte die Nutzung von KI angegeben werden?

Die Angabe der Nutzung von KI entspricht den geltenden Transparenzstandards der GWP (siehe [Leitlinien 12](#) und [13](#) im DFG-Kodex „Leitlinien zur Sicherung guter wissenschaftlicher Praxis“). Angaben zur Nutzung von KI erlauben Lesenden und Reviewer:innen die Ergebnisse, Methoden und Arbeitsschritte besser nachzuvollziehen. Aufgrund der Vielzahl von unterschiedlichen KI-Anwendungen und Funktionen gibt es bzgl. der genauen Gestaltung der Angaben jedoch unterschiedliche Empfehlungen. Auch werden einige Aspekte der Angabe zur Nutzung von KI in der Forschungscommunity noch vermehrt diskutiert, z.B. die Dokumentation von Prompts ([siehe Frage 6](#)).

### 3. Was genau bedeutet „die Nutzung von KI transparent und angemessen angeben“?

Darauf gibt es keine einheitliche Antwort. Die meisten Editorial Policies spezifizieren die Angabe der Nutzung nicht oder nur minimal. Es gibt jedoch auch Vorstöße von kleineren Verlagen, z.B. [Berlin Universities Publishing \(BUP\), die genaue Vorschläge erarbeitet haben](#). Vorschläge zur Angabe von KI in der Literatur unterscheiden sich in den Details bzw. der Komplexität der Angabe, allen gemein ist jedoch ein gewisser Minimalkonsens. Dazu gehört die Angabe

- der genutzten KI-Anwendung, inklusive Version, Datum der Nutzung, URL
- wofür die KI-Anwendung wie genutzt wurde

Weitere Vorschläge beinhalten die Angabe, welche Person aus einem Team für die Nutzung der KI verantwortlich war (siehe [Hosseini et al. 2023](#)), aber auch ausführliche Überlegungen zur technischen Arbeitsweise der genutzten KI und ihren Limitierungen ([Resnik und Hosseini 2024](#)). Insbesondere letzterer Vorschlag geht weit über grundsätzlich geforderte Angaben hinaus, kann aber durchaus als Denkanstoß im Umgang mit KI dienlich sein, um zu prüfen, ob die favorisierte KI-Anwendung zur Nutzung passt, und auf welche Schwächen geachtet werden muss. Generell sollte die Angabe der Nutzung den Transparenzbedürfnissen der Leser:innen und Reviewer:innen gerecht werden, aber auch der tatsächlichen Arbeitsweise mit KI entsprechen. Beides kann sehr fachspezifisch ausfallen und sollte deswegen auch mit der eigenen Community besprochen werden. Dies betrifft auch Fragen, welche Nutzungsarten und KI-Anwendungen der Dokumentationspflicht unterliegen ([siehe dazu Fragen 7-11](#)).

### 4. Wo im Manuskript muss ich die Nutzung von KI angeben?

Dazu gibt es in den Richtlinien und Empfehlung nicht immer genaue Angaben. Oft wird jedoch der Methodenteil oder die Danksagung als Ort für die Angabe der KI-Nutzung vorgeschlagen. Welcher Teil des Manuskriptes sich für die Angabe anbietet, kann auch von fachspezifischen Konventionen sowie der Form der Angabe abhängen. So kann zum Beispiel eine sehr detaillierte Angabe der KI-Nutzung inklusive Prompts und Chatverläufen auch als Supplement hinterlegt werden (wie z.B. [ASC Nano](#) vorschlägt).

### 5. Gibt es Zitiervorschläge für die Angabe von KI?

Sofern institutsintern oder vom Journal keine konkreten Zitationsvorschläge zur Angabe von KI vorgegeben sind, können bestehende Vorschläge etablierter Zitationsstile zur Orientierung genutzt werden. Dazu gehören die [American Psychological Association](#) (APA), das [Chicago Style Manual](#) sowie die [Modern Language Association](#) (MLA). Dabei sollte darauf geachtet werden, dass die Angabe alle wichtigen Elemente enthält, die vom Zielmedium gefordert werden.

### 6. Müssen Prompts dokumentiert werden?

Dazu gibt es momentan noch keine einheitlichen Angaben. Von den großen Verlagen macht bis jetzt nur Science genaue Vorgaben (Prompts sollen demnach im Methodenteil angegeben werden). Auch die APA, das Chicago Style Manual sowie die MLA empfehlen die Angabe von Prompts mit unterschiedlichen Vorschlägen, wie und wo diese erfolgen kann. Unter Forschenden wird die Nützlichkeit der Angabe von Prompts z.T. hinterfragt, mit der Begründung, dass sie nicht der tatsächlichen (oft iterativen) Arbeitsweise mit KI entsprechen würde und die KI-generierten Antworten nicht reproduzierbar seien und somit selbst bei gleichlautenden Prompts keine deckungsgleichen Antworten erfolgen würden. Die Frage, ob Prompts angegeben werden müssen, sollte daher weiterhin insbesondere von den Fachcommunitys diskutiert werden. Zu überlegen wäre, welche Funktion die Angabe der Prompts für die Fachcommunity bzw. für die jeweilige Publikation erfüllt. Auch wenn Ergebnisse der KI nicht reproduzierbar sind, können die genauen Prompts, die eingesetzt wurden, ggf. Aufschlüsse über die Arbeitsweise der Autor:innen geben.

## 7. Was ist mit der Angabe der Nutzung von KI für Software?

In den Editorial Policies der Verlage wird meist nicht dezidiert auf Software eingegangen. Die [World Association of Medical Editors empfiehlt](#), dass KI-Tools, die bei der Erstellung von Code beteiligt waren, im Manuskript angegeben werden sollten. Aus Transparenzgründen sollte generell für Software gelten, dass die Nutzung von KI bei der Erstellung und Überarbeitung des Codes transparent angegeben werden sollte. Dies ist insbesondere dann wichtig, wenn die Software anderen Forschenden zur Verfügung gestellt bzw. publiziert wird.

## 8. Muss ich die Nutzung von KI angeben, wenn ich diese nur nutze um meinen Text stilistisch/sprachlich zu verbessern?

Einige Editorial Policies unterscheiden zwischen der Verwendungsart von KI bzw. zwischen verschiedenen KI (meist in generative KI auf der einen und Anwendungen zum Überprüfen von Grammatik und Stil, wie z.B. Grammarly, auf der anderen Seite) und spezifizieren, dass die Richtlinien zur Angabe von KI sich nicht auf Letztere beziehen (siehe beispielsweise die [AI Policy von Elsevier](#) oder [Wiley](#)). Auch im Rahmen der [Stellungnahme der DFG](#) werden KI-Anwendungen, „die sich nicht auf den wissenschaftlichen Inhalt des Antrags [auswirken] (bspw. Grammatik-, Stil-, Rechtschreibprüfung, Übersetzungsprogramme)“ von der Kennzeichnungspflicht ausgenommen.

Innerhalb der Fächer kann diese Frage auch unterschiedlich bewertet werden, je nachdem welche Stellung Text bzw. sprachliche Formulierung innehaben. Insbesondere in den Geisteswissenschaften ist der Sprachstil eng an Personen bzw. Schulen geknüpft. In diesem Fall wäre die Angabe der Nutzung von Anwendungen, die den Sprachstil verbessern, denkbar.

## 9. Muss ich die Nutzung von KI für Übersetzungen angeben?

Übersetzung mithilfe von KI wird in den Editorial Policies und Empfehlungen meist nicht dezidiert adressiert. Eine Ausnahme bildet die Empfehlung der DFG, die Übersetzungsprogramme als Beispiel für nicht deklarationspflichtige KI-Anwendungen anführt ([siehe Frage 8](#)). Auch die [BUP-Leitlinie](#) klassifiziert Übersetzungsprogramme als Hilfsmittel, wonach „nicht zwingend eine explizite Nennung“ erforderlich ist. Autor:innen sollte bewusst sein, dass beim maschinellen Übersetzen von Texten wichtige Informationen verloren gehen oder verzerrt werden können. Auch hier gilt, dass die übersetzten Texte sorgfältig geprüft werden müssen und Autor:innen die Verantwortung für etwaige Fehler übernehmen.

Übersetzungen können außerdem als wichtige Eigenleistung und Skill bzw. Selbstverständnis in einem Fach (z.B. den fremdsprachlichen Philologien) angesehen werden. Insbesondere der Übersetzung von literarischen Texten kommt ein hoher Stellenwert zugute: „Eine Übersetzung ist das Ergebnis eines individuellen Umgangs mit einem Ausgangswerk. Dieser Umgang muss gewissenhaft verantwortet werden, nicht nur im eigenen Namen, sondern auch in dem des Autors bzw. der Autorin des Originals“ ([VdÜ / A\\*ds / IGÜ – Offener Brief zur KI-Verordnung](#)). In Fächern wie den Philologien und bei der Übersetzung von literarischen Texten (für die wissenschaftliche Arbeit) wäre die Angabe der Nutzung von Übersetzungstools empfehlenswert.

## 10. Muss ich angeben, dass ich KI nur zur Inspiration genutzt habe?

Dazu gibt es teils konkrete Empfehlungen, z.B. bei [BUP](#). Diese empfiehlt einen allgemeinen Hinweis auf die Nutzung als Anmerkung oder ggf. im Methodenteil. Generell gilt es auch hier zu prüfen, ob im anvisierten Journal/Verlag dazu Vorgaben gemacht werden. Fehlen konkrete Vorgaben sollte eine Entscheidung mit Blick auf fachspezifische Leseerwartungen sowie Bedeutung/Umfang der KI-generierten Ideen getroffen werden.

## 11. Welche Anwendungen fallen unter den Begriff KI?

Im Allgemeinen Sprachgebrauch dient in der Debatte um künstliche Intelligenz meist der Chatbot „ChatGPT“ als Synonym für künstliche Intelligenz, insbesondere für Large Language Models (LLM). KI-Anwendungen, die für Forschende relevant sind, gibt es derzeit viele (siehe z.B. [Liste von KI-Ressourcen zusammengestellt durch VK:KIWA](#) oder den [Ithaka Product Tracker](#)). Im Zuge der steten technischen Weiterentwicklung werden auch bereits bestehende Anwendungen teilweise durch KI-Komponenten ergänzt, z.B. bei gängigen Suchmaschinen. Eine klare Grenzziehung ist deswegen teils schwierig. Es ist möglich den Begriff KI sehr weit auszulegen, wie das beispielsweise der [EU AI Act](#) tut. In Richtlinien für Forschende wird sich momentan zumeist auf generative KI, insbesondere LLM, beschränkt. KI-Anwendungen, die nur zur Sprachbearbeitung/-verbesserung genutzt werden, fallen in eine Grauzone ([siehe Frage 8](#)). KI-Anwendungen, die bei den ersten Schritten eines Forschungsprojektes unterstützen (Literatursuche, Hypothesenfindung, Literatur-Review), werden nicht immer in Empfehlungen dezidiert mitgenannt. Eine Ausnahme bilden die [Living Guidelines on the Responsible Use of Generative AI in Research](#) der Europäischen Kommission, in denen diese KI als „substantial use“ klassifiziert werden und demnach deklarationspflichtig wären. Einige Verlags-Policies grenzen deutlich ab, auf welche KI sich welche Regelungen und Transparenzangaben beziehen. Die [Policy von Elsevier](#) beispielsweise unterscheidet zwischen KI-Anwendungen für den Schreibprozess, für den Forschungsprozess sowie weiteren assistierenden Anwendungen: Für alle drei gibt es unterschiedliche Regelungen. Autor:innen sollten sich im Vorfeld mit den für sie maßgeblichen Regelungen vertraut machen und im Zweifelsfall eher den größtmöglichen Transparenzbedingungen gerecht werden.

## 12. Darf ich KI für das Schreiben von Anträgen nutzen?

Hier sind die Angaben der jeweiligen Forschungsförderer zu beachten. Die Deutsche Forschungsgemeinschaft (DFG) hat in ihrer [Stellungnahme zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG](#) (2023) die Nutzung von KI in Anträgen stattgegeben, wenn diese klar markiert ist ([siehe auch Frage 3](#)). Ferner führt die Stellungnahme aus, dass die Nutzung von KI in Anträgen neutral bewertet wird, also weder negativ noch positiv seitens der Gutachter:innen geltend gemacht werden soll. Zu beachten ist hingegen, dass die Stellungnahme die Nutzung von KI im Rahmen der Begutachtung verbietet ([siehe auch Frage 14](#)).

## 13. Darf KI zum Generieren von Bildern/Graphiken genutzt werden?

Wie in der Kurzübersicht einsehbar, haben Journals häufig sehr restriktive Vorgaben. Meist ist die Nutzung nur in einem sehr begrenzten Rahmen erlaubt, z.B. in thematischen Beiträgen, die sich dezidiert mit dem Thema KI auseinandersetzen. Die Nutzung muss dann ebenfalls ausreichend gekennzeichnet sein. Ausnahmen sind z.B. der Verlag Frontiers, der die Nutzung von KI zum Erstellen von Bildern ausdrücklich erlaubt, mit dem Verweis, dass die Nutzung klar gekennzeichnet werden muss ([siehe Frontiers Artificial intelligence: fair use and disclosure policy](#)).

Für die Nutzung von KI bei anderen wissenschaftlichen Formaten, z.B. Vorträge oder Poster-Präsentationen, existieren meines Wissens momentan keine übergreifenden Vorgaben. Hier sollte zuerst mit der eigenen Projektgruppe/Fachcommunity gesprochen werden. Bei Bildern, die rein zu illustrativen Zwecken auf Vortragsfolien genutzt werden, spricht aus Sicht der GWP nichts gegen eine Nutzung von KI. Bei Graphiken, die Forschungsergebnisse illustrieren (z.B. Diagramme oder Schemata), ist die Nutzung von KI denkbar, wenn diese vermerkt wird. Wie bei KI-generiertem Text obliegt es den Autor:innen das Ergebnis auf Korrektheit zu überprüfen.

## 14. Darf KI in der Peer-Review eingesetzt werden?

Für die Begutachtung ist die Nutzung von KI in vielen Richtlinien und Empfehlungen entweder starken Einschränkungen unterworfen oder gar nicht erlaubt ([siehe Verlagsübersicht unter Frage 1](#)). Generell

gilt, dass die Einspeisung des zu begutachtenden Manuskriptes (oder Projektantrags) in eine KI nicht erlaubt ist. Angeführt werden dafür die Aspekte Vertraulichkeit und Datenschutz, die dadurch ggf. kompromittiert werden. Auch sollte bedacht werden, dass die Funktion der Begutachtung auch der Wissensproduktion und Weiterentwicklung des eigenen Faches dient. Diese lenkenden Aufgaben sollten nicht an eine KI ausgelagert werden (siehe auch [Hosseini und Horbach 2023](#)).

KI darf, wenn überhaupt, dann nur zur sprachlichen Nachbearbeitung des Reviews genutzt werden ([siehe Verlagsübersicht unter Frage 1](#)). Reviewer:innen sollten jeweils prüfen, welche Vorgaben für sie gelten.

### 15. Was ist bei der Nutzung von KI in Autorschaftsteams zu beachten?

Aufgrund der relativen Neuheit vieler KI-Anwendungen sollte in Autorschaftsteams zu Projektbeginn geklärt werden, ob alle Autor:innen mit der Nutzung von KI-Tools und dem Umfang der Nutzung einverstanden sind. Insbesondere bei trans- und interdisziplinären Autorschaftsteams ist dies wichtig, da hier z.B. unterschiedliche Bewertungen von Textarbeit aufeinandertreffen. In Teams bietet sich außerdem eine gute, interne Dokumentation der Nutzung von KI an. In Fällen von nachträglich auftretenden Konflikten oder GWP-Verstößen kann der Prozess dann besser nachvollzogen werden. Autor:innen, die in Teams publizieren, können zudem überlegen, ob sie der Zitationsempfehlung von [Hosseini et al 2023](#) folgen, die bei der Angabe der Nutzung von KI zusätzlich die anwendende Person notiert ([siehe auch Frage 3](#)).

### 16. Was ist der Zusammenhang von KI-generierten Texten und Plagiaten? (Kann ich durch die Nutzung von KI aus Versehen andere Texte plagiiere?)

Da generative KI-Modelle auf Grundlage einer riesigen Anzahl an Trainingsdaten stochastisch arbeiten, gehört unbeabsichtigtes Plagiiere nicht zu den häufigen Risiken von KIs. Es gibt Beispiele, die aufzeigen, dass bekannte KIs trotz ihrer stochastischen Natur bestimmte Texte fast wortwörtlich reproduzieren können ([Henderson et al. 2023](#) oder der Fall New York Times versus OpenAI, siehe z.B. [Pope 2024](#)). In den Tests wurde per Prompt jedoch genau auf dieses Resultat hingearbeitet. In bestimmten Fächern der Geisteswissenschaften, in denen Text und Sprache eine übergeordnete Rolle einnehmen, können jedoch schon einzelne Begrifflichkeiten bestimmten bekannten Theoretiker:innen zugeordnet werden. Fehlt im generierten Text die Ausweisung, auf wen der Begriff zurückgeht, kann dies daher entweder als unzureichende Referenzierung oder gar als Plagiat angesehen werden (siehe Seadle „For the humanities, words matter. [...] A stolen word is a stolen thought“ (42)). Forschende, die sich länger in einer Disziplin bewegen, sind sich der typischen Fachdiskurse zumeist bewusst; deswegen sind besonders interdisziplinär Forschende angehalten, KI-generierte Inhalte gründlich zu prüfen. Aber auch generell sollte immer eine umfassende Überarbeitung von KI-generiertem Text durch die Forschenden erfolgen.

### 17. Auf welche Schwächen und Risiken von KI sollten Forschende achten?

Für generative KI ist besonders das sogenannte Halluzinieren eine bekannte Schwäche. Es gibt jedoch eine ganze Reihe von Risiken und Schwächen, die bekannte Anwendungen mit sich bringen. Dazu gehören fehlende oder falsche Referenzen, Fehler in wörtlichen Zitaten, komplett fabrizierte Zitate oder Referenzen, Weitergabe veralteter Informationen sowie die Reproduktion von Bias bzw. vorurteilsbehafteten Aussagen. Der große Anteil englischsprachiger (oft auch US-amerikanischer) Quellen im Trainingsmaterial lädt zu Vorsicht bei der Nutzung von KI in anderen Sprachen ein. Eine Typologie der Schwächen und Risiken von generativen KI-Anwendungen findet sich beispielsweise bei [Oertner 2024](#). Autor:innen übernehmen die Verantwortung für die von KI produzierten Falschangaben und Regelverstöße. Durch gutes Prompting und gutes Prüfen lassen sich die Risiken der Nutzung minimieren.

## 18. Kann die Nutzung von KI in einem Text durch Software-Tools nachgewiesen werden? Wie erkenne ich KI generierte Texte?

Es gibt mehrere Studien, die sich mit Detektions-Tools befasst haben, die teilweise zu unterschiedlichen Ergebnissen hinsichtlich des genauen Potentials von Software-Tools kommen ([Weber-Wulff et al. 2023](#), [Gao et al. 2023](#)). Generell lässt sich mit Blick auf die Studien jedoch festhalten, dass es keine hinreichend zuverlässigen Tools zur Erkennung von KI-generierten Inhalten auf dem Markt gibt. Auch menschliche Reviewer:innen konnten in Studien nur unzureichend KI-generierte Texte von menschlich-geschriebenen unterscheiden. Von einer eindeutigen Feststellung der Nutzung von KI in Texten ist im Moment also nicht auszugehen.

## 19. Ich bin Reviewer:in oder Editor:in und vermute, ein Text bzw. Teile davon wurden mithilfe einer KI geschrieben, diese Nutzung ist jedoch nicht transparent angegeben. Was sollte ich tun?

Das kommt darauf an, was genau diese Vermutung auslöst. Es gibt einige deutliche Anzeichen, dass eine KI zum Einsatz kam, z.B. gängige Phrasen in Texten („Certainly, here is an introduction for you“, „As an AI language model, I cannot...“, „as of my last knowledge update“) oder Nonsense-Wörter bzw. verzerrte Schriftzeichen bei der Erstellung von Bildern. Hierbei würde es sich um einen belegten und somit eindeutigen Verstoß gegen die gängigen Vorgaben bzgl. der KI-Nutzung handeln. Ob dies direkt zu einer Ablehnung eines Manuskript führen sollte oder den Autor:innen die Möglichkeit eingeräumt werden sollte, den Text nachzubessern und die nötigen Angaben nachzutragen, sollte journalintern festgelegt werden. Einige Verlagspolicies spezifizieren rudimentär das Vorgehen, falls nicht-deklariertes KI-Inhalt in Manuskripten vermutet wird. Ist die Vermutung auf weniger konkrete Anzeichen zurückzuführen und macht sich eher am Sprachstil generell fest bzw. an der Nutzung von bestimmten Wörtern, die mittlerweile oft als Signalwörter für KI-Nutzung angesehen werden (z.B. „delve“, „meticulous“, „commendable“), sollte ein klärendes Gespräch mit den Autor:innen gesucht werden, denn weder die genannten Signalwörter noch eine Prüfung mit einer Software kann hier zweifelsfrei die Nutzung von KI belegen. Eine vorschnelle Anschuldigung eines Verstoßes gegen die GWP sollte mit Blick auf [Leitlinie 18](#) des DFG-Kodex vermieden werden, in der es heißt: „[b]ewusst unrichtig oder mutwillig erhobene Vorwürfe können selbst ein wissenschaftliches Fehlverhalten begründen“.

## 20. Was passiert, wenn ich beschuldigt werde KI genutzt zu haben, ohne dies in der Publikation angegeben zu haben?

Dazu gibt es erste Erfahrungsberichte ([Wolkovich 2024](#)). Wie in [Frage 18](#) aufgeführt gibt es keine Software-Tools, die die Nutzung von KI sicher nachweisen können. Autor:innen sollten sich die genaue Begründung geben lassen, wieso ihr Text bzw. Teile davon im Verdacht stehen KI-generiert zu sein. Hilfreich ist es, den Arbeits- und Schreibprozess für sich selbst intern zu dokumentieren, damit anhand der Dokumentation ggf. nachvollzogen bzw. nachgewiesen werden kann, dass nicht mit KI gearbeitet wurde.

## 21. Gibt es aus GWP-Sicht besonders empfehlenswerte KI-Anwendungen?

Welche KI-Anwendung die geeignetste für eine Aufgabe ist, hängt von verschiedenen Faktoren und deren Gewichtung ab. So sollten Forschende neben der Bewertung nach Funktionen und Leistung von KI-Anwendungen auch prüfen, ob die gewählte KI-Anwendung gesetzliche Vorschriften und Bestimmungen einhält. Datenschutz und Vertraulichkeit spielen dabei eine übergeordnete Rolle. Viele KI-Anwendungen sind außerdem dafür bekannt, dass sie weitere Schwächen und Risiken bzgl. der generierten Inhalte mit sich bringen ([siehe Frage 17](#)). Autor:innen sollten sich der Schwächen und Risiken bewusst sein und die generierten Inhalte daraufhin sorgfältig prüfen. Es stellen sich zudem



weitere ethische Fragen im Zusammenhang mit bestimmten KI-Anwendungen sowie KI-Anwendungen im Allgemeinen. Die Entscheidung mit KI oder einer bestimmten KI zu arbeiten, sollte deswegen im Vorfeld mit Bedacht getroffen werden. Eine Art Leitfaden zur Auseinandersetzung mit KI aus Sicht der GWP können dabei die Überlegungen von [Resnik und Hosseini \(2024\)](#) bieten, die eine sehr detaillierte Auflistung zur transparenten Angabe von KI-Anwendungen zusammengestellt haben.

## 22. Wie ist der Zusammenhang zwischen Urheberrecht und KI-generierten Inhalten?

Fragen im Zusammenhang mit Urheberrecht und KI-Anwendungen können sowohl die generierten Inhalte betreffen (z.B. Habe ich das Urheberrecht an den von mir mithilfe einer KI-generierten Inhalten?) als auch die in eine KI eingespeisten Inhalte (Kann ich das Urheberrecht Anderer durch meine Arbeit mit KI verletzen?). Mit diesen Fragen haben sich bereits [Roman Konertz 2023](#) und [Ulrike Verch 2024](#) eingehend beschäftigt auf deren Publikationen wir an dieser Stelle verweisen.

## 23. Wie kann ich Konflikte, die aus KI-Nutzung resultieren könnten, vorbeugen?

Besonders im jetzigen Stadium, in dem die Arbeit mit KI bei vielen Forschenden noch nicht vollständig etabliert ist und Empfehlungen/Richtlinien noch im Entstehen sind, sollten Forschende sich zu Beginn mit den für sie maßgeblichen KI-Richtlinien auseinandersetzen sowie ihre eigene Expertise in Bezug auf KI richtig einschätzen. Arbeiten Forschende im Team oder einem langfristigen Forschungsprojekt mit fluktuierendem Personal (z.B. Editionsprojekte), sollte es stets eine offene Kommunikation über die Nutzung von KI sowie eine transparente interne Dokumentation darüber geben. Für andere Forschende im Team/Projekt sollte ebenso ersichtlich sein, ob und wie KI zum Einsatz kam. Dies gilt auch für andere Produkte wissenschaftlicher Arbeit abseits von „klassischen“ Publikationen, z.B. Software, Vortragsmanuskripte und Folien (die ggf. geteilt werden), Vorlagen für Posterpräsentationen und Weiteres.

## 24. Welche KI-Richtlinien sind für mich maßgeblich?

Dies hängt u.a. davon ab, wozu KI-Anwendungen in welchem Kontext genutzt werden. Für die Erstellung von Publikationen sollten neben institutionellen Richtlinien – sofern vorhanden – die KI-Policies der Verlage beachtet werden. Existieren Richtlinien innerhalb der eigenen Fachcommunity sollten diese Anwendung finden, insbesondere wenn diese strengere Kriterien beinhalten. Bei Promotionen gelten Promotions- und Prüfungsordnungen. Außerdem kann und sollte das Thema mit den Betreuungspersonen besprochen werden. Bei Anträgen gelten die Angaben aus [Frage 12](#). Sollten sich Widersprüche auftun – z.B. zwischen institutionellen Richtlinien und Verlagsrichtlinien – sollte diese frühzeitig kommuniziert werden.

## 25. Warum gibt es keine klaren Empfehlungen vom *Ombudsman für die Wissenschaft* zum Thema KI und GWP?

Der OfdW hat sich Anfang 2023 mit der Frage beschäftigt, ob zum Thema KI und GWP Empfehlungen herausgegeben werden sollten. In diesem Zusammenhang konzipierten und organisierten die Mitarbeiter:innen des Projektes, Katrin Frisch, Felix Hagenström und Nele Reeg, einen Expert:innenworkshop, in dem die wichtigsten Fragen zu KI entlang der bestehenden GWP-Leitlinien des DFG Kodex diskutiert wurden. Die Ergebnisse dieses Workshops mündeten in einem [Kurzbericht](#) sowie [einer Publikation zum Thema](#), die im Dezember 2023 in der Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) erschien. Da sich während des Workshops herauskristallisierte, dass viele Fragen zur GWP-konformen Nutzung von KI entweder schon durch den Kodex implizit adressiert werden oder aber einer Klärung in den einzelnen Fachcommunitys bedürfen, um verschiedener Fachspezifika gerecht zu werden, wurden keine allgemeinen Empfehlungen erarbeitet. Der OfdW kann aber bei der Aufsetzung von Empfehlungen beraten.

## Referenzen

Alle Referenzen finden sich als Direktlink ebenso an den jeweiligen Stellen im Text, an denen auf sie verwiesen wird. Eine [Literatursammlung zu KI und GWP](#) findet sich auch auf der Webseite des OfdW.

## Literatur

Frisch, Katrin, Felix Hagenström und Nele Reeg. Textgenerierende KI und gute wissenschaftliche Praxis. *Zeitschrift für Bibliothekswesen und Bibliographie* 70, Nr. 6 (15. Dezember 2023): 326–36. <https://doi.org/10.3196/186429502370667>.

Gao, Catherine A., Frederick M. Howard, Nikolay S. Markov et al. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers, *npj Digital Medicine*, Volume 6, Issue 75. <https://doi.org/10.1038/s41746-023-00819-6>

Henderson, Peter, Xuechen Li, Dan Jurafsky, et al. (2023). Foundation Models and Fair Use [Preprint]. *Arxiv*. <https://doi.org/10.48550/arXiv.2303.15715>

Hosseini, Mohammad, David B. Resnik und Katie Holmes (2023). The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*. Volume 19, Issue 4. <https://doi:10.1177/17470161231180449>

Hosseini, Mohammad, Serge P.J.M. Horbach (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Res Integr Peer Rev* 8, 7. <https://doi.org/10.1186/s41073-023-00133-5>

Konertz, Roman (2023). Urheberrechtliche Fragen der Textgenerierung durch Künstliche Intelligenz: Insbesondere Schöpfungen und Rechtsverletzungen durch GPT und ChatGPT. *Wettbewerb in Recht und Praxis*, Jahrgang 69. <https://doi.org/10.18445/20230613-111651-0>

Oertner, Monika (2024). ChatGPT als Recherchetool?: Fehlertypologie, technische Ursachenanalyse und hochschuldidaktische Implikationen. *Bibliotheksdienst* 58, Nr. 5 : 259-297. <https://doi.org/10.1515/bd-2024-0042>

Pope, Audrey (2024). NYT v. OpenAI: The Times's About-Face. *Harvard Law Review*. <https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/>

Resnik, David B., Hosseini, Mohammad (2024). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI Ethics*. <https://doi.org/10.1007/s43681-024-00493-8>

Seadle, Michael (2017): *Quantifying Research Integrity*. San Rafael, California: Morgan & Claypool.

Verch, Ulrike (2024). Per Prompt zum Plagiat? Rechtssicheres Publizieren von KI-generierten Inhalten. *API*, Band 5, Nr. 1. <https://doi.org/10.15460/apimagazin.2024.5.1>.

Weber-Wulff, Debora, Alla Anohina-Naumeca, Sonja Bjelobaba, et al. (2023): Testing of detection tools for AI-generated text. *International Journal for Education Integrity*, Volume 19, Issue 26. <https://doi.org/10.1007/s40979-023-00146-z>

Wolkovich, E.M. (2024). 'Obviously ChatGPT' — how reviewers accused me of scientific fraud. *Nature; Career Column*. <https://doi.org/10.1038/d41586-024-00349-5>

Policies, Handreichungen, Empfehlungen

ACS Nano: Best Practices for Using AI When Writing Scientific Manuscripts, 2023, 17, 5, 4091–4093.  
<https://doi.org/10.1021/acsnano.3c01544>

American Psychological Association: How to Cite ChatGPT: <https://apastyle.apa.org/blog/how-to-cite-chatgpt>

Berlin Universities Publishing: „Leitlinie zum Umgang mit künstlicher Intelligenz“ und „Handreichung zur Zitation von KI-Tools“, 2024 <https://www.berlin-universities-publishing.de/ueber-uns/policies/ki-leitlinie/index.html>

Chicago Manual of Style: Citing content developed or generated by artificial intelligence:  
<https://www.chicagomanualofstyle.org/qanda/data/faq/topics/Documentation/faq0422.html>

Deutsche Forschungsgemeinschaft: Leitlinien zur Sicherung guter wissenschaftlicher Praxis, 2019:  
<https://wissenschaftliche-integritaet.de/>

Deutsche Forschungsgemeinschaft: Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG, 2023: <https://wissenschaftliche-integritaet.de/verwendung-generativer-modelle/>

Elsevier: Generative AI policies for journals: <https://www.elsevier.com/de-de/about/policies-and-standards/generative-ai-policies-for-journals>

EU AI Act (Gesetz über künstliche Intelligenz (Verordnung (EU) 2024/1689):  
<https://artificialintelligenceact.eu/de/das-gesetz/>

Europäische Kommission: Living guidelines on the responsible use of generative AI in research, 2024:  
[https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc\\_en?filename=ec\\_rtd\\_ai-guidelines.pdf](https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf)

Frontiers: Artificial intelligence: fair use and disclosure policy:  
<https://www.frontiersin.org/guidelines/policies-and-publication-ethics>

ICMJE: Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals, 2024: <https://www.icmje.org/icmje-recommendations.pdf>

Modern Language Association: How do I cite generative AI in MLA style?: <https://style.mla.org/citing-generative-ai/>

Science Journals: Editorial Policies: <https://www.science.org/content/page/science-journals-editorial-policies#image-and-text-integrity>

Springer Nature: Editorial Policies: Artificial Intelligence:  
<https://www.springernature.com/gp/policies/editorial-policies>

Taylor & Francis: AI Policy: [https://taylorandfrancis.com/our-policies/ai-policy/?\\_ga=2.202881233.1550617214.1729762987-990921615.1729762987](https://taylorandfrancis.com/our-policies/ai-policy/?_ga=2.202881233.1550617214.1729762987-990921615.1729762987)

WAME: Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications, 2023: <https://wame.org/page3.php?id=106>

Wiley: Best Practice Guidelines on Research Integrity and Publishing Ethics: Artificial Intelligence:  
<https://authorservices.wiley.com/ethics-guidelines/index.html#22>

