

Richtlinie für Vertragspartner:innen zum Umgang mit generativen Modellen Künstlicher Intelligenz

Version: 30. 6. 2025

1. Einleitung

Das Österreichische Institut für Wirtschaftsforschung (nachstehend "WIFO" genannt) verpflichtet sich zu einem verantwortungsvollen, transparenten und ethischen Einsatz künstlicher Intelligenz (KI) in allen seinen Tätigkeitsbereichen. Die Richtlinie fasst den aktuellen Stand der Regeln zum Einsatz von KI für Organisationen und Personen, die für das WIFO als Vertragspartner:innen wie (Sub-)Auftragnehmer:innen oder Konsortialpartner:innen (nachstehend "Vertragspartner:in" genannt) tätig werden, zusammen.

Der fachgerechte Einsatz von künstlicher Intelligenz bietet zahlreiche Chancen für mehr Effizienz, Effektivität und Kreativität und wird daher vom WIFO ausdrücklich begrüßt. Der rasche Wandel erfordert ein hohes Maß an Aufmerksamkeit und Flexibilität sowie die Bereitschaft kontinuierlich zu lernen. Im Mittelpunkt stehen **generative Modelle** der künstlichen Intelligenz, die sich dadurch auszeichnen, dass sie aus vorhandenen Daten vorhandene Muster, Strukturen oder Eigenschaften lernen und daraus **neue, originäre Inhalte erzeugen**¹⁾.

Soweit der Einsatz von KI-Tools bei der Erbringung der vertraglich vereinbarten Leistungen zulässig ist, **darf der:die Vertragspartner:in KI-Tools ausschließlich in Übereinstimmung mit dieser KI-Richtlinie sowie allen geltenden Anforderungen verwenden**. Die Nichteinhaltung dieser Bestimmungen stellt einen wesentlichen Verstoß dar, kann somit zu rechtlichen Konsequenzen führen und ist umgehend und nachweislich dem WIFO mit entsprechender Begründung zu melden.

Der:Die Vertragspartner:in stellt sicher, dass all jenen Personen, die im Rahmen der Beauftragung für das WIFO tätig werden, die vorliegende Richtlinie nachweislich zur Kenntnis gebracht wurden.

2. Wissenschaftliche Integrität

Neben den Anforderungen an die technische Ausstattung und die persönlichen Fähigkeiten im Umgang mit KI ist die wissenschaftliche Integrität eine wesentliche Voraussetzung für Qualität, Vertrauen und positive gesellschaftliche Auswirkungen von Forschung. Grundsätzlich sind

¹⁾ Die hier verwendete Definition des Begriffs "generative KI-Modelle" basiert auf einer Abfrage von ChatGPT sowie weiteren Recherchen (Google, Wikipedia) am 14. 3. 2025.

beim Einsatz von KI wie bei allen anderen Forschungsaktivitäten die geltenden nationalen, europäischen und internationalen Rechtsvorschriften zu beachten. Ebenso gelten dieselben Anforderungen und Regeln, wie sie z. B. im Europäische Verhaltenskodex für wissenschaftliche Integrität (All European Academies, 2023) oder in den WIFO-Richtlinien zur wissenschaftlichen Integrität festgelegt sind²⁾. Jede Form von **Fehlverhalten**, wie z. B. das Erfinden, Verfälschen oder Plagieren von Daten und Ergebnissen ist ebenso auszuschließen wie das Verletzen von **Persönlichkeitsrechten, Urheberrechten und Datenschutz**³⁾. Wissenschaftliche Integrität stellt zusätzlich folgende besondere Anforderungen an die Forschung:

- **Transparenz** in Bezug auf Auftraggeber:innen, finanzielle Beiträge, Interessenkonflikte;
- korrektes **Zitieren** aller relevanten **Quellen** für Ideen, Texte, Daten usw.;
- sorgfältiges **Prüfen** der Ergebnisse unter Berücksichtigung kritischer Annahmen, Robustheit und Grenzen der Aussagekraft der verwendeten Methoden;
- **Nachvollziehbarkeit** der Resultate durch das Aufbewahren erforderlicher Unterlagen (z. B. Daten, Methoden, Fragebögen, Stichprobenplan) für mindestens 10 Jahre.

3. Besondere neue Herausforderungen

Die vielfältigen neuen Möglichkeiten in Verbindung mit der bislang beispiellosen und rasch ansteigenden Leistungsfähigkeit **generativer KI-Modelle** bringen zusätzliche Herausforderungen mit sich (Europäische Kommission, 2023, Deutsche Forschungsgemeinschaft, 2023). Die folgenden Grundsätze ergänzen und präzisieren daher die WIFO-Richtlinien zur wissenschaftlichen Integrität im Hinblick auf den Einsatz von generativer KI.

Eigenverantwortung

Der rechtssichere und regelkonforme Umgang mit generativer KI liegt in der Eigenverantwortung der natürlichen Person, die sie aktiv einsetzt oder ihre Ergebnisse weiterverwendet. Dazu gehört auch die Bereitschaft, sich im Umgang mit generativen KI-Modellen kontinuierlich weiterzubilden, z. B. durch Schulungen oder die Teilnahme an Anwendungsforen. Dieser **Eigenverantwortung** entsprechend steht jeder Person, die generative KI-Modelle verwendet, für die Integrität der verwendeten Inhalte ein und

- prüft KI-Ergebnisse sorgfältig;
- trägt dafür Sorge, dass sie keine KI-generierten Erfindungen, Fälschungen oder Plagiate verwendet oder verbreitet;
- pflegt einen kritischen Umgang mit den Ergebnissen;
- ist sich der Grenzen der Werkzeuge (z. B. in Bezug auf mögliche Halluzinationen, Verzerrungen und Ungenauigkeiten sowie fehlende Reproduzierbarkeit) bewusst.

²⁾ Siehe als Anhang zum (Sub-)Werkvertrag sowie auch unter <https://www.wifo.ac.at/institut/evidenzbasierung-transparenz-und-unabhaengigkeit/>.

³⁾ Vgl. z. B. Birnbauer (2025).

In diesem Zusammenhang hält etwa die Leibniz Gesellschaft (2024, 3) fest, dass KI-Systeme "lediglich unterstützende Instrumente" sind. Deren "Resultate haben als solche keine wissenschaftliche Validität und sind kein Ersatz für authentische Erkenntnis- und Entscheidungsprozesse." Neue KI-generierte Inhalte müssen daher immer sorgfältig von der verantwortlichen natürlichen Person auf ihre Richtigkeit überprüft werden. Bleiben dennoch berechtigte Zweifel an deren Richtigkeit, gilt es den Einsatz der KI-Hilfsmittel entweder auf eine erste Ideenfindung (Brainstorming) mit nachfolgender umfassender und eigenverantwortlicher Recherche zu beschränken oder auf deren Verwendung zu verzichten.

Aus dieser Eigenverantwortlichkeit ergibt sich für alle Anwendungen generativer KI-Modelle das Prinzip des **Human in the Loop**, d. h. der Vorrang menschlicher Kontrolle und Verantwortung, wonach **am Anfang und am Ende jeder Anwendung** eine eigenverantwortlich handelnde natürliche Person stehen muss⁴⁾. Daraus folgt z. B., dass es keine automatisierte Publikation von KI-generierten Inhalten ohne vorherige Kontrolle durch eine für diese Inhalte verantwortlichen Person geben darf.

Transparenz

Die **Autorenschaft** von Inhalten, die mithilfe generativer KI-Modelle erzeugt werden, liegt bei den menschlichen Nutzer:innen der KI⁵⁾. Der Einsatz von generativer KI im Forschungsprozess erfordert aber bei der Veröffentlichung von Ergebnissen einen konkreten Hinweis darauf, welche Werkzeuge, wann und für welche Aufgabe eingesetzt wurden. Die Regeln für das **Zitieren** von KI-generierten Inhalten befinden sich international noch in Entwicklung und werden regelmäßig aktualisiert. Sie werden von verschiedenen Forschungsgesellschaften oder wissenschaftlichen Verlagen teilweise sehr unterschiedlich festgelegt. Bei Veröffentlichungen sind jeweils die Regeln von Auftraggebern, Verlagen bzw. Herausgeber:innen von Forschungsergebnissen bzw. der genutzten KI-Werkzeuge zu beachten.

Folgende Regeln und Leitlinien bieten eine allgemeine **Orientierung**:

- Sofern die Daten nicht öffentlich zugänglich sind, werden diese als **persönliche Kommunikation** behandelt und entweder in Fußnoten⁶⁾ oder direkt im Text (ChatGPT, 6. April 2025) zitiert. Eine Aufnahme in das Literaturverzeichnis ist nicht notwendig.
- Die Nutzung von KI mit Einfluss auf das Gesamtergebnis aber ohne direkt zuordenbare konkrete Inhalte, z. B. zur allgemeinen Ideenfindung ("Brainstorming"), wird in den **allgemeinen Hinweisen** offengelegt⁷⁾.

⁴⁾ Vgl. z. B. Austria Presse Agentur (2023) oder Österreichischer Rundfunk (2025).

⁵⁾ So setzt etwa auch ein urheberrechtlicher Schutz voraus, dass "diese (i) eigentümlich und (ii) geistige Schöpfungen eines Menschen sind" (Birnbauer, 2025, Reg. 6, Kap 1.2, 1).

⁶⁾ Z. B.: "Diese Erklärung beruht auf einer Frage an/Interaktion mit ChatGPT vom 6. April 2025, <https://chat.openai.com>."

⁷⁾ Z. B. in den "Acknowledgements" oder in einem Abschnitt "Weitere Quellen" im Anschluss an das Literaturverzeichnis.

- Der Einsatz von **rein technischen Hilfsmitteln**, die selbst keine originär neuen Inhalte generieren, kann davon ausgenommen sein⁸⁾. Die Einschätzung, welche KI-unterstützten Inhalte in nennenswertem Umfang originärer Natur sind, obliegt den Autor:innen. In Zweifelsfällen, insbesondere bei umfangreicher Nutzung, sollten diese aber in den allgemeinen Hinweisen angeführt werden.
- Soweit es für die spätere **Nachvollziehbarkeit** der Forschungsergebnisse erforderlich ist, gilt es auch die Eingaben (Prompts) sowie Ausgaben der KI für mindestens 10 Jahre zu speichern. Davon ausgenommen sind Daten und Dokumente, die aufgrund von besonderen Bestimmungen (z. B. Datenschutz, Urheberrechte, Nutzungsbedingungen) nach Verarbeitung zwingend eine Löschung erfordern. D. h., der Schutz von Daten, Urheberrechten, usw. hat im Zweifel Vorrang. Die durch Löschung entstandenen Lücken müssen nachvollziehbar dokumentiert werden.

Persönlichkeitsrechte und Datenschutz

Sensible oder geschützte Daten dürfen keinesfalls online in ein KI-System geladen werden, es sei denn, es wurde durch eine fachlich kompetente Stelle auf Seiten des:der (Sub-)Auftragnehmers:in geprüft und bestätigt, dass das verwendete Tool diese Daten nicht speichert oder weitergibt und z. B. für das Training zukünftiger Sprachmodelle wiederverwendet. Zusätzlich sind etwaige Nutzungsvereinbarungen (einschließlich Urheberrechte) für die verwendeten Datenbestände zu beachten. Beispiele für sensible oder geschützte Daten sind

- Betriebsgeheimnisse (z. B. unveröffentlichte Strategiedokumente, Entwürfe für Prognosen, Stellungnahmen oder Pressemitteilungen),
- Personaldaten (z. B. Bewerbungsunterlagen, Protokolle von Personalgesprächen),
- personenbezogene Inhalte Dritter in E-Mails (z. B. Signaturen, Meinungen, persönliche Aussagen) sowie
- unveröffentlichte und insbesondere fremde Forschungsarbeiten.

⁸⁾ In den Veröffentlichungen des WIFO können solche Ausnahmen z. B. für KI-unterstützte Sprachkorrekturen und Übersetzungsdienste, dem Coding von Programmscripts oder der Nutzung von Software zur Transkription von Interviews zur Anwendung kommen, sofern diese nicht den Nutzungsbedingungen der KI-Werkzeuge widerspricht.

Literaturhinweise

All European Academies – ALLEA (2023). *Europäischer Verhaltenskodex für wissenschaftliche Integrität – überarbeitete Fassung 2023*. <https://allea.org/wp-content/uploads/2024/09/CoC2023-German.pdf>.

Austria Presse Agentur – APA (2023). *Leitlinie zum Umgang mit künstlicher Intelligenz*.

Birnbauer D. (Hrsg.; 2025). *Künstliche Intelligenz in der Praxis*, WEKA.

Deutsche Forschungsgemeinschaft – 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. <https://www.dfg.de/resource/blob/289674/ff57cf46c5ca109cb18533b21fba49bd/230921-stellungnahme-praesidium-ki-ai-data.pdf>.

Europäische Kommission (2023). *Living guidelines on the responsible use of generative AI in research. ERA Forum Stakeholders' document*. European Commission, Directorate-General for Research and Innovation. https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en.

Europäische Union (2024). *Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 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)*. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401689.

Leibniz Gesellschaft (2024). *Empfehlung zur Sicherung der guten wissenschaftlichen Praxis beim Umgang mit Künstlicher Intelligenz*. Beschlossen durch das Präsidium der Leibniz-Gemeinschaft am 29. November 2024. https://www.leibniz-gemeinschaft.de/fileadmin/documents/3085-2906/Leibniz_Empfehlung_zur_KI_Beschluss_Praesidium_2024_11_29_.pdf.

Österreichischer Rundfunk – ORF (2025). *KI-Guidelines: glaubwürdig, verlässlich und innovativ*.

Weitere Quellen

Bei der Erstellung dieses Leitfadens wurde DeepLWrite für die sprachliche Korrektur und ChatGPT für einzelne spezifische Recherchen (z. B. Begriffsdefinitionen, Zitierregeln) verwendet. Die Autor:innen haben die verwendeten Inhalte überprüft und bearbeitet und übernehmen für diese die volle Verantwortung.