

## KI-Leitfaden für philosophisches Arbeiten<sup>1</sup>

Stand: 04.03.2025

### 1. Grundsätzliches

Künstliche Intelligenz (KI) und Large Language Models (LLMs) wie etwa ChatGPT erzielen beeindruckende Erfolge und dringen immer mehr in (Hoch)Schulen als ein unverzichtbares Instrument ein – zumeist jedoch nur unsystematisch und wenig reflektiert. Es ist daher keine Frage mehr, *ob* KI an (Hoch)Schulen verwendet werden sollte, sondern *wie*.<sup>2</sup> Dieser Leitfaden konzentriert sich vor allem auf den Einsatz von KI im Hochschulkontext, kann aber auch auf schulische Kontexte angewendet werden. Konkret ergeben sich folgende Fragen: Wie *darf* KI bei der Erstellung von schriftlichen Leistungen eingesetzt werden, damit diese am Ende noch als *eigene* Leistungen gelten können? Oder anders gefragt: Wie *sollte* KI bei der Erstellung von schriftlichen Leistungen eingesetzt werden, um als ein sinnvolles Hilfsmittel noch bessere Leistungen, Ergebnisse und Lerneffekte zu ermöglichen? Dieser Leitfaden möchte Orientierung und Empfehlungen bei diesen Fragen geben und zugleich geeignete Hilfsmittel benennen. Im Gegensatz zu KI-Leitfäden anderer Universitäten, die oftmals rechtliche Aspekte in den Vordergrund stellen,<sup>3</sup> fokussiert dieser Leitfaden speziell auf den Einsatz von Künstlicher Intelligenz im Fach Philosophie und auf ihren philosophischen Gebrauch. Dieser Gebrauch muss freilich unter Berücksichtigung von rechtlichen Rahmenbedingungen stattfinden, was auch den Schutz von privaten und geschützten Daten betrifft. Auch gilt es zu berücksichtigen, dass die Entwicklung von KI ständig Wandlungen unterliegt, die in diesem Leitfaden (Stand 03/2025) noch nicht antizipiert werden können. Deswegen wird dieser Leitfaden in regelmäßigen Abständen aktualisiert und den neuen technischen Herausforderungen angepasst.<sup>4</sup>

### 2. Zur Funktionsweise von KI

Künstliche Intelligenz, wie sie gegenwärtig vor allem im Sinne von künstlichen neuronalen Netzen und sogenannter „generativer KI“ verwendet wird, basiert auf den Prinzipien von Mustererkennung und Wahrscheinlichkeitskalkulation. Auf Basis von gewaltigen Datenmengen werden KI-Systeme darauf trainiert, Muster zu abstrahieren, zu erkennen und auf andere Bereiche (neu) anzuwenden. Solche Muster sind u.a. sprachlicher Art. Auf diese Weise ist es möglich, mit Chatbots wie ChatGPT Gespräche zu führen, die sich teilweise von zwischenmenschlicher

---

<sup>1</sup> Ich danke Georg Gasser, Nurten Güler Güçlü, Johanna Häusler, Thomas Heichele und Uwe Voigt für hilfreiche Kommentare zum Leitfaden.

<sup>2</sup> Vgl. auch <https://intranet.uni-augsburg.de/de/services/portal-lehre/ki-in-der-hochschullehre>.

<sup>3</sup> Vgl. etwa die allgemeinen Leitfäden der Fernuni Hagen ([https://www.fernuni-hagen.de/zli/docs/6716\\_ki-leitfaden\\_-\\_din\\_a4\\_-\\_web.pdf](https://www.fernuni-hagen.de/zli/docs/6716_ki-leitfaden_-_din_a4_-_web.pdf)) und der Universität Basel ([https://www.unibas.ch/dam/jcr:e46db904-bf0f-475a-98bc-94ef4d16ad2e/Leitfaden-KI-zitieren\\_v2.2.pdf](https://www.unibas.ch/dam/jcr:e46db904-bf0f-475a-98bc-94ef4d16ad2e/Leitfaden-KI-zitieren_v2.2.pdf)).

<sup>4</sup> Die jeweils aktuelle Version findet sich unter folgender Adresse <https://joerg-noller.com/ki-leitfaden-philosophie>.

Kommunikation nicht mehr unterscheiden lassen. Ebenso ist es möglich, Chatbots durch sogenannte „Prompts“ konkrete Instruktionen zu geben, die diese dann immer besser ausführen (z.B.: „Fasse mir den Text in 100 Wörtern zusammen.“). Der Grund dafür liegt darin, dass KI diejenigen sprachlichen Muster identifiziert, die am wahrscheinlichsten auf eine von Menschen gestellte Frage oder Anweisung passen. Die Ergebnisse von KI werden also nicht infolge von guten Gründen oder wissenschaftlicher Einsicht gewonnen, die im Detail transparent nachvollzogen und erklärt werden können, sondern durch komplexe Wahrscheinlichkeitskalkulation, die den Anschein von Wissenschaftlichkeit erweckt.

### **3. Zur Problematik von KI im Hochschulkontext**

Aufgrund der beeindruckenden sprachlichen Leistungen von KI ist es verlockend, Chatbots auch für die Erstellung von wissenschaftlichen Texten an Universitäten zu benutzen oder gar ganze Hausarbeiten durch sie selbst verfassen zu lassen, etwa durch die Anweisung: „Schreibe eine 4000 Wörter umfassende wissenschaftliche Hausarbeit über die Problematik von Kants kategorischem Imperativ“. Auch wenn dies generell von Seiten der Universitäten verboten ist, so besteht doch von Seiten der Studierenden und Lehrenden nach wie vor ein nicht unerheblicher Klärungsbedarf von Fragen, die den konkreten und erlaubten Einsatz von KI betreffen. Deswegen sollen im Folgenden sinnvolle und erlaubte Wege für den Einsatz von KI im Hochschulkontext aufgezeigt werden, jenseits von KI-Plagiat und digitaler Unmündigkeit. Grundsätzlich gilt, dass der Einsatz von KI im Hochschulkontext nur insofern erfolgen sollte, als er unsere eigenen wissenschaftlichen Fähigkeiten erweitert, d.h. wir dadurch einen Lern- und Erkenntniseffekt erzielen. Nicht verwendet werden sollte KI, sofern unsere wissenschaftlichen Fähigkeiten dadurch ersetzt oder nur simuliert werden. In neueren Studien werden die negativen Auswirkungen von unkritischer KI-Nutzung auf unser kritisches Urteilsvermögen durch kognitive Entlastung („cognitive offloading“) diskutiert. Ebenso gilt, dass kein Produkt von KI unreflektiert und unverarbeitet unmittelbar in eine wissenschaftliche Hausarbeit eingehen sollte. Denn aufgrund ihres Wahrscheinlichkeitsprinzips ist es immer möglich, dass die KI auf den ersten Blick stimmige Ergebnisse nur fingiert oder simuliert. Dies wird in neueren Studien im Sinne von Harry Frankfurt's Begriffs des „Bullshits“ diskutiert, also von Urteilen, die nicht an der Wahrheit interessiert sind.

### **4. Sinnvoller und erlaubter Einsatz**

Sofern KI im Rahmen einer wissenschaftlichen Arbeit verwendet wurde, sollte am Ende immer ihr konkreter methodischer Einsatz unter der Rubrik „Hilfsmittel“ angegeben werden.

#### 4.1 Textrecherche und Vorverarbeitung

KI eignet sich hervorragend dazu, längere Texte nach bestimmten Suchbegriffen und Themen zu durchsuchen und zu ordnen. Ein Beispiel dafür ist Googles KI „NotebookLM“ (<https://notebooklm.google.com>). Damit ist es möglich, große Textkorpora (etwa Kants „Kritik der reinen Vernunft“) einzulesen und mithilfe von konkreten Anweisungen verarbeiten zu lassen, etwa in der Form: „Nenne mir die fünf häufigsten Kontexte, in denen der Begriff der Freiheit in diesem Text vorkommt.“ Insofern kann die KI ein umfangreiches Begriffsregister ersetzen, dessen Recherche sehr zeitaufwändig ist. Der Vorteil von NotebookLM besteht darin, dass die KI diejenigen Passagen eines Textes ausweist bzw. zitiert, auf deren Basis sie zu ihren Ergebnissen gelangt ist. Hier ist also wissenschaftliche Transparenz und Nachvollziehbarkeit gegeben.

#### 4.2 Zusammenfassungen und Übersicht

Oftmals müssen für das Erstellen von wissenschaftlichen Arbeiten große Textmengen, etwa an Forschungsliteratur, hinzugezogen und durchforstet werden, von denen nicht auf den ersten Blick ersichtlich ist, ob sie für die gewählte Fragestellung relevant sind. Hier empfiehlt es sich, mit Hilfe von geschickt gewählten Prompts Zusammenfassungen erstellen zu lassen: „Fasse den Inhalt dieses Textes in 200 Wörtern zusammen.“ Diese Zusammenfassungen dienen dann als Basis für die weitere Beschäftigung mit dem Thema und dem Text, etwa indem man sich einen bestimmten Forschungsaufsatz genauer ansieht oder die KI mit weiterführenden, spezifischeren Fragen und Anweisungen auf den Text ansetzt („Wo befasst sich die Autorin in ihrem Text mit Kants kategorischem Imperativ?“).

#### 4.3 Anregung, Assoziation, Diskussion und Problematisierung

Insbesondere Chatbots wie ChatGPT eignen sich hervorragend dafür, um mit ihnen in ein philosophisches Gespräch einzutreten, das wesentlich problemorientiert ist. Dies ermöglicht es, durch KI eine fremde oder gar neue Perspektive auf selbst gewählte Themen zu erhalten und sich philosophisch anregen zu lassen. Eine mögliche Instruktion wäre etwa: „Welche Gegenargumente wurden in der Philosophie gegen Kants These des kategorischen Imperativs vorgebracht?“. Auch hier gilt, dass die Anregungen immer nur als Ausgangspunkt für die eigene Reflexion dienen sollten und dass sie auf ihren Wahrheitsgehalt und ihre Plausibilität hin überprüft werden müssen. Durch Mustererkennung ist KI besonders gut darin, ähnliche Phänomene und Themen zu erkennen und zu assoziieren. Dies erlaubt es z.B. bei der Themenfindung solche Gegenstände und Werke zu identifizieren, die strukturell miteinander zusammenhängen und deren Zusammenhang ansonsten nur schwer zu erkennen ist. So verwendet, wird KI nicht als *Ersatz* des eigenen wissenschaftlichen Denkens, sondern als dessen *Erweiterung* gebraucht. Wann die Verwendung von KI bei wissenschaftlichen Arbeiten angegeben werden muss, hängt davon ab, wie spezifisch

ihre Antworten sind. Je fachspezifischer und konkreter die Antworten einer KI sind, umso eher sollte ihre Verwendung genannt werden, so wie es auch üblich ist, einer anderen Person in einer Fußnote für eine spezifische Anregung zu danken.

## 5. Übersicht über aktuelle Chatbots und ihren philosophischen Einsatz

Name	Eigenschaften	Philosophischer Einsatz
ChatGPT ( <a href="https://chatgpt.com">https://chatgpt.com</a> )	Vorteile: Sehr leistungsstark, kann Bilder generieren, kann über individuelle ChatGPT-Versionen erweitert werden  Nachteile: Bislang keine Quellennachweise der Ergebnisse, keine ethischen Richtlinien implementiert; es ist aber möglich, nach den Quellen zu fragen	Kann als philosophischer Diskussionspartner verwendet werden  Kann philosophische Texte zusammenfassen und erlaubt dazu weitergehende Rückfragen (z.B.: „Was könnte man gegen diese These einwenden?“)
Claude ( <a href="https://claude.ai">https://claude.ai</a> )	Vorteile: Ethische Richtlinien und Werte implementiert („Constitutional AI“)  Nachteile: Bislang keine Quellennachweise der Ergebnisse	Kann als philosophischer Diskussionspartner verwendet werden.  Kann philosophische Texte zusammenfassen und erlaubt dazu weitergehende Rückfragen
NootebookLM ( <a href="https://notebooklm.google.com">https://notebooklm.google.com</a> )	Vorteile: Gibt verwendete Quellen präzise an, lässt spezifische Quellen selektieren, verwenden und verarbeiten, erlaubt es, verschiedene Quellenprojekte („Notebooks“) anzulegen  Nachteile: Kann nicht ohne Quellen verwendet werden	Kann philosophische Texte zusammenfassen und erlaubt dazu weitergehende Rückfragen, auch als Diskussionspartner  Erlaubt es, Texte unter spezifischen Fragestellungen zu analysieren und Rückfragen dazu zu stellen, gibt Quellen präzise an
Perplexity ( <a href="https://www.perplexity.ai">https://www.perplexity.ai</a> )	Vorteile: Sehr leistungsstark, gibt die verwendeten Quellen an  Nachteile: Kann keine Bilder generieren	Kann philosophische Texte zusammenfassen und erlaubt dazu weitergehende Rückfragen, auch als Diskussionspartner

## 6. Fazit

Insbesondere NotebookLM erweist sich momentan als geeignetes philosophisches Hilfsmittel, da es eine präzise und transparente argumentative Quellenarbeit mit philosophischen Texten erlaubt.