zh aw

Angewandte Psychologie



# Start? Klar!
## Toolbox für deine wissenschaftliche Arbeit

Schreiben? Check! am Di, 17. und Do, 19.11.2026

Angewandte Psychologie

# KI beim Schreiben: Was schiefgehen kann – und was hilft

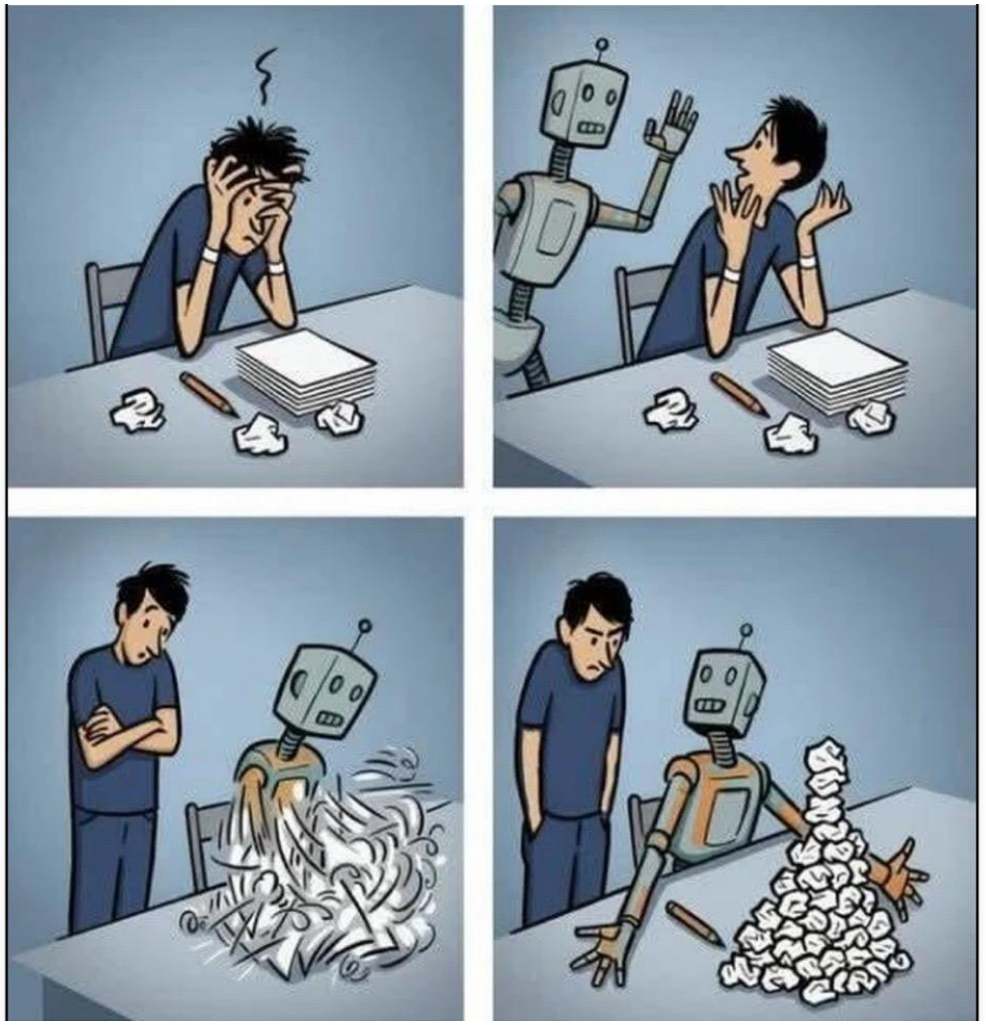
Schreiben? Check! | 19.11.2026




https://stock.adobe.com/images/photo-of-concentrated-student-woman-thinking-while-doing-homework/358052117

# Was schiefgehen kann

# Was schiefgehen kann

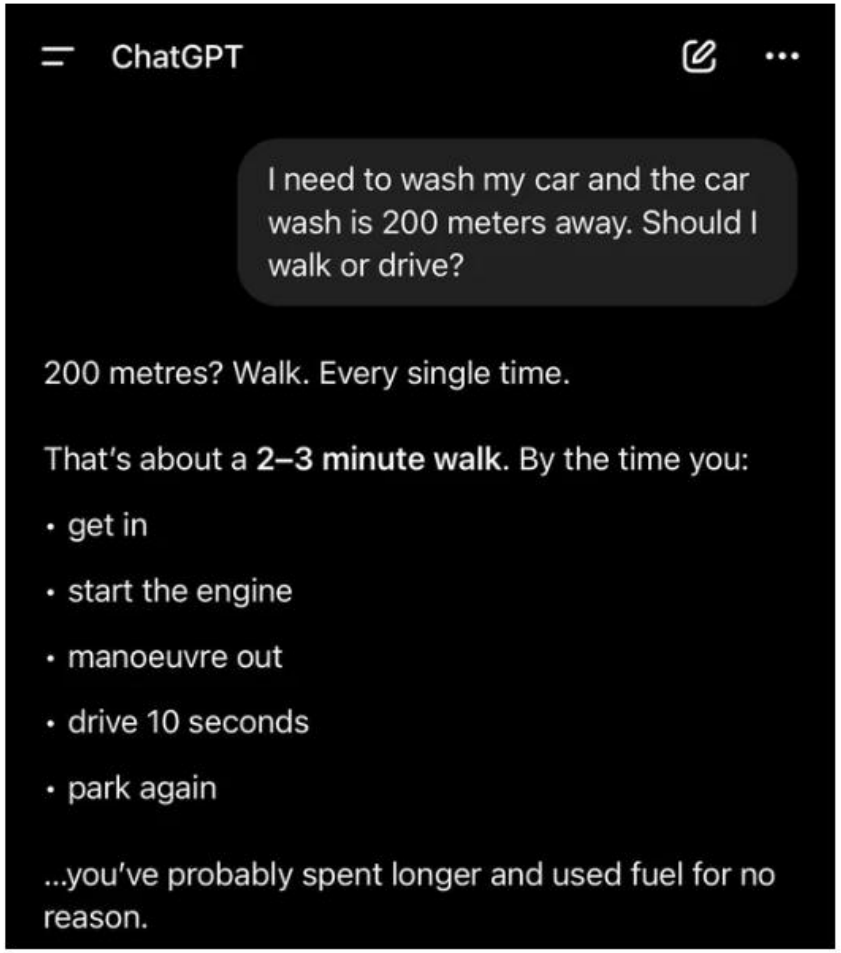


https://www.instagram.com/war.and.peas/

# Was schiefgehen kann: Halluzinationen




@angrytomtweets




https://x.com/tomwarren/status/2024422296983634170

# Was schiefgehen kann: Halluzinationen

«KI-Systeme halluzinieren die ganze Zeit, da sie nichts verstehen und deshalb lediglich halluzinieren können. Häufig sind die ausgegebenen Antworten nur ‚zufällig' richtig, weil die LLMs mit ausreichend viel Material trainiert wurden.»

Buck (2025, p. 37)

"Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot."

Bender et al. (2021, pp. 616–617)



https://www.reddit.com/r/ProgrammerHumor/comments/o10ltp/choose_your_fighter/

# Was schiefgehen kann: Halluzinationen

**Wieso gibt es Halluzinationen?**

- wahrscheinlichkeitsbasierte Aneinanderreihung von Tokens
- mehr Fehler durch programmierte Variation («Temperature»)
- fehlendes Textverständnis
- mehr Fehler bei unzureichender Datengrundlage (aufgrund von «Ausgabezwang»)




https://www.cloud-science.de/ki-halluzinationen/

# Was schiefgehen kann: Sycophancy

**KI bestätigt uns in unseren Ansichten:**

- Verbreitung von Fehlern und Falschinformationen
- Verringerung von kritischem Denken / Hinterfragen
- Polarisierung/Radikalisierung (Echokammern)




https://www.reddit.com/r/ChatGPT/comments/1lg6vyy/sycophancy_has_eaten_technical_accuracy/

# Was schiefgehen kann: KI-Bias

**Verzerrte Daten führen zu verzerrten Ausgaben**

→ Verzerrungen in Verfügbarkeit, Vorverarbeitung und Auswahl von Trainingsdaten

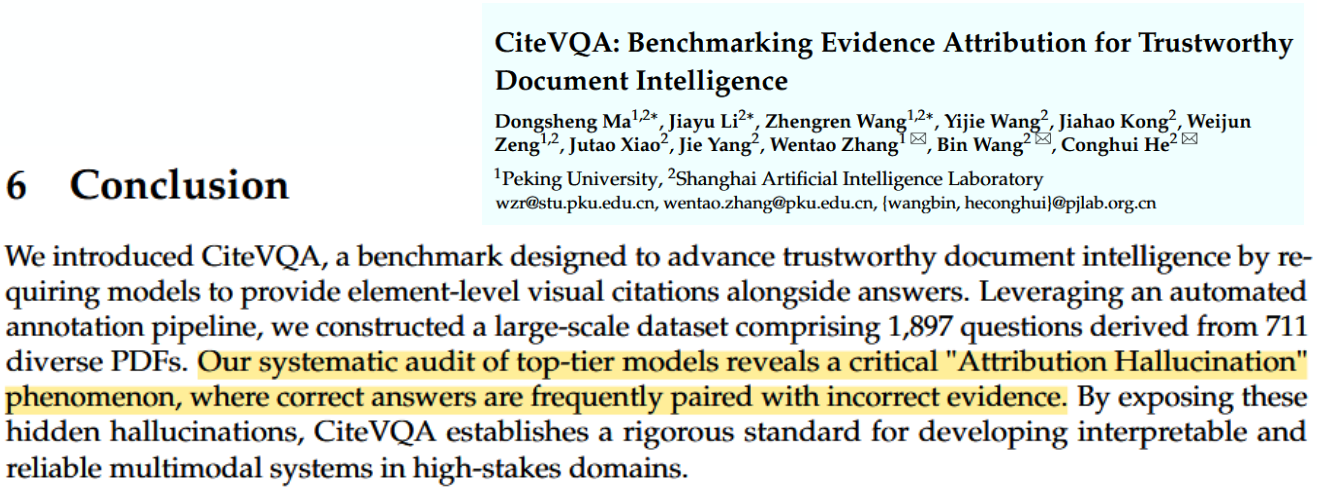



https://marketoonist.com/2025/06/ai-hallucinations-and-reliability.html

# Was schiefgehen kann: Transparenz

**Probleme:**

- Zustandekommen von Outputs: Datengrundlage, deren Verrechnung im Hintergrund sowie allfällige Filter häufig nicht bekannt bzw. nachvollziehbar
- Replizierbarkeit von Outputs: Eingaben in KI-Tools führen häufig nicht zu denselben Outputs
- Scheintransparenz: Häufig Zitierung von passenden Quellen, aber z. B. Verweis auf falsche Textstellen



**CiteVQA: Benchmarking Evidence Attribution for Trustworthy Document Intelligence**

Dongsheng Ma[1,2*], Jiayu Li[2*], Zhengren Wang[1,2*], Yijie Wang[2], Jiahao Kong[2], Weijun Zeng[1,2], Jutao Xiao[2], Jie Yang[2], Wentao Zhang[1] ✉, Bin Wang[2] ✉, Conghui He[2] ✉

[1]Peking University, [2]Shanghai Artificial Intelligence Laboratory

wzr@stu.pku.edu.cn, wentao.zhang@pku.edu.cn, {wangbin, heconghui}@pjlab.org.cn

## 6 Conclusion

We introduced CiteVQA, a benchmark designed to advance trustworthy document intelligence by requiring models to provide element-level visual citations alongside answers. Leveraging an automated annotation pipeline, we constructed a large-scale dataset comprising 1,897 questions derived from 711 diverse PDFs. Our systematic audit of top-tier models reveals a critical "Attribution Hallucination" phenomenon, where correct answers are frequently paired with incorrect evidence. By exposing these hidden hallucinations, CiteVQA establishes a rigorous standard for developing interpretable and reliable multimodal systems in high-stakes domains.



https://doi.org/10.48550/arXiv.2605.12882

# Was schiefgehen kann: Kompetenzentwicklung

**Gefährdung von Kompetenzen wie *critical thinking*, Informationskompetenz und Schreibkompetenz:**

- Illusion des Lernens: Kompetenzen werden durch oberflächliche Beübung gar nie wirklich erworben
  ↑ «kognitive Leichtigkeit»
  ↓ erlebte Anstrengung

- «Deskilling»: Kompetenzen gehen verloren, weil sie weniger geübt werden




https://www.pinterest.com/pin/burn-calories-not-electricity-take-the-stairs--353955470631444497/

# Was schiefgehen kann: Autorenschaft

- **Verwässerung des Schreibstils:** Bestimmte, häufig verwendete Wörter und Wortfolgen werden häufiger generiert, was die sprachliche und emotionale Variation in den ausgegebenen Texten einschränkt.
- **Bleibende Verantwortung für den Text:** Studierende sind «verantwortlich, die Relevanz, den Wahrheitsgehalt und die Genauigkeit sowohl der eingesetzten generativen KI-Systeme als auch des Outputs zu überprüfen.»
  (Merkblatt «Verwendung generativer KI bei Arbeiten», p. 1)
- **Verlust von Urheberschaft:** Rein KI-generierte Inhalte sind grundsätzlich keine menschlichen geistigen Schöpfungen und geniessen daher nicht ohne Weiteres urheberrechtlichen Schutz.

# Was schiefgehen kann: Darf ich dafür überhaupt KI verwenden?

§

- **Datenschutz:** Aus Datenschutzgründen können KI-Tools nicht für alle Aufgaben verwendet werden.
- **Verwendungsrechte:** Gemäss Fachhochschulgesetz (FaHG) gehören die Verwendungsrechte der an der ZHAW verfassten schriftlichen Arbeiten der ZHAW («urheberrechtlich geschützte Werke», § 16 Abs. 1 lit. b FaHG). Studierende sind dafür verantwortlich, dass die Verwendungsrechte auf die ZHAW übergehen können. Dies schliesst gewisse Verwendungsformen von KI-Tools bereits im Vorhinein aus (insb. umfangreiche Generierung von Text).
- **Richtlinien der ZHAW:** KI darf zwar verwendet werden, aber nicht alle Formen der Verwendung sind erlaubt!

# Und: Ersetzbarkeit

**Wer nur KI schreiben lässt, ohne selbst etwas beizutragen, wird ersetzbar (jede:r mit der KI kann euch ersetzen)**

**→ Wenn jede:r KI einsetzen kann, müsst ihr einen eigenen Teil beitragen, um in der Arbeitswelt bestehen zu können!**



https://www.srf.ch/news/kuenstliche-intelligenz-angst-vor-ki-ist-weit-verbreitet-was-ist-realistisch

# Einzelauftrag: Reflexion

2 Minuten überlegen: Welchen Problemen bin ich selbst schon begegnet?

**Halluzinationen**
Fehler / Falschangaben durch KI

**Sycophancy**
übermässige Bestätigung durch KI

**KI-Bias**
verzerrte Ausgaben durch KI

**fehlende Transparenz**
unklares Zustandekommen von Ausgaben

**Illusion des Lernens**
trotz nur oberflächlichem Verständnis

**Einschränkung der Autorenschaft**
z. B. Verwässerung des Schreibstils, fehlendes Verantwortungsgefühl für den Text

**rechtliche Unsicherheiten**
z. B. Darf ich KI so verwenden? Was darf ich hochladen? Wie ist es mit dem Datenschutz?

**Abhängigkeit von KI**
z. B. Impuls, bei einer Aufgabe direkt zur KI zu greifen; Befürchtung, von KI ersetzt werden zu können



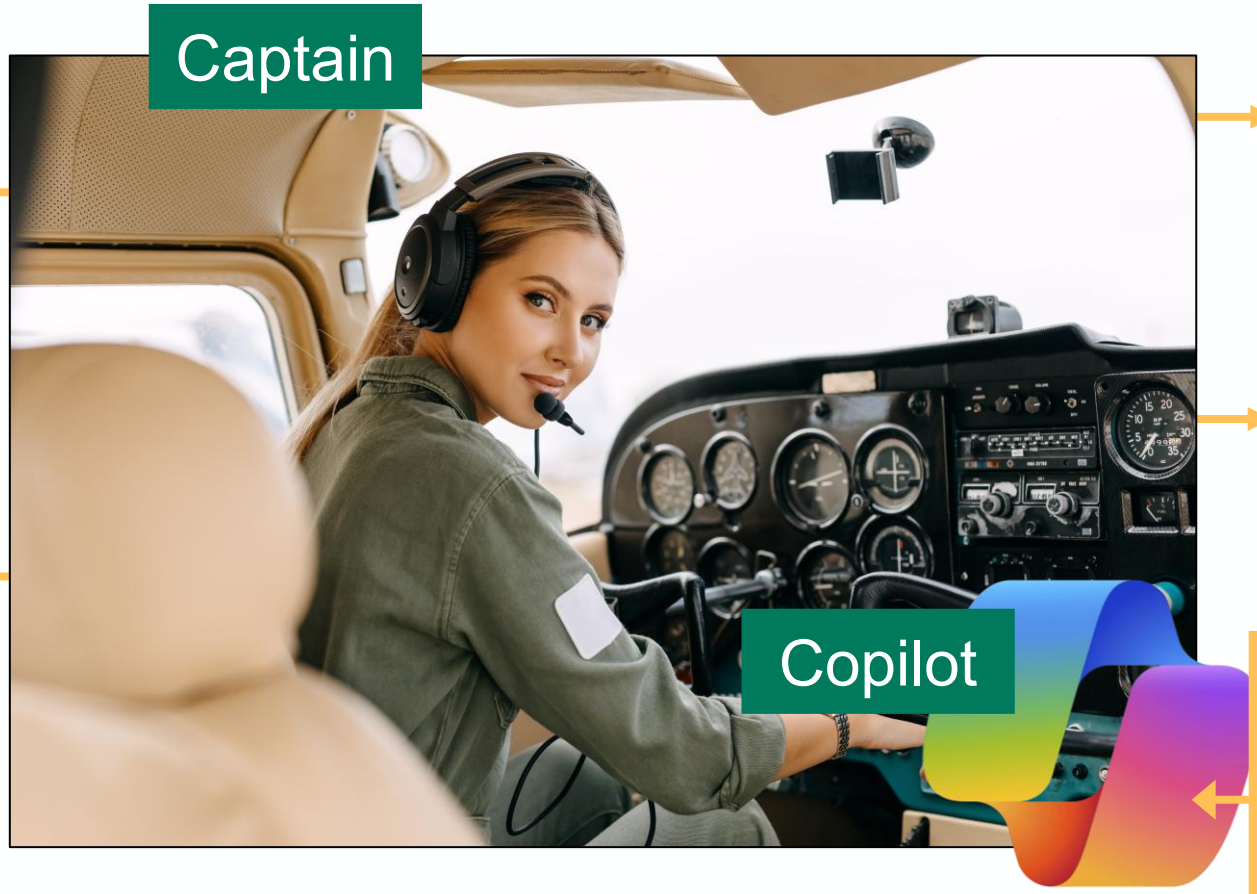

https://stock.adobe.com/images/photo-of-woman-with-headache-rubbing-her-temples-while-doing-homework/358052219

# Was hilft

# Was hilft: Grundhaltung «KI als Copilot:in»

Du bist Captain.

Du musst die Copilot:in überwachen und beurteilen.




Du trägst die Verantwortung.

Du musst auch ohne Copilot:in fliegen *lernen* und können.

Die Copilot:in kann nicht in allen «Gebieten» fliegen.

Schreibende als Captains, welche den Einsatz von KI-Tools als ihre Assistenzen (a) planen, (b) steuern und (c) überprüfen

# Was hilft: Grundhaltung «KI als Copilot:in»

1

## Planung: KI-Tools gezielt einsetzen

→ Aufgaben verteilen: Für welche Aufgaben werden KI-Tools wie eingesetzt?
→ Definieren, anhand welcher Kriterien die Ausgabe der KI-Tools überprüft werden soll

2

## Steuerung: Einsatz von KI-Tools aktiv steuern

→ Klares Formulieren von aus der Planung abgeleiteten Prompts
→ Ausgabe anhand aus der Planung abgeleiteter Kriterien prüfen und ggf. Prompts anpassen

3

## Überprüfung: Ausgaben von KI-Tools kritisch prüfen

→ Inhaltliches Überprüfen anhand externer Quellen, eigener Expertise und kritischem Denken
→ Ausgegebene Quellen kritisch prüfen

# Was hilft: PRÜFE-Framework (Geyer, 2025)



# Was hilft: Zuerst Denken, dann KI

“Writing is thinking on paper.”

Zinsser (2006, p. 147)

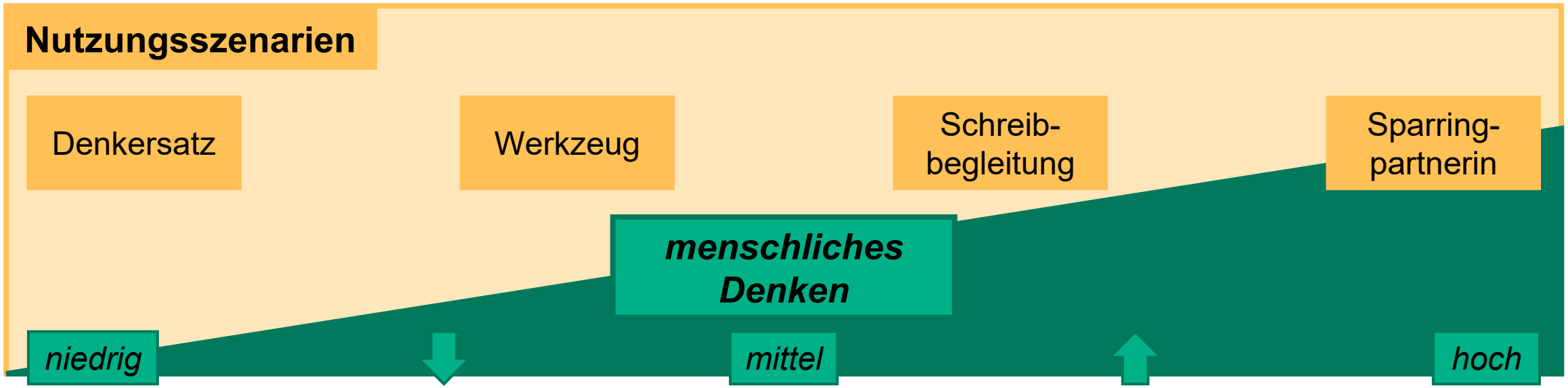
**→ Direkte Textgenerierung nicht zielführend, weil dadurch ein essenzielles Element des Schreibens verloren geht**

“Always generate your own ideas before turning to AI. Write them down, no matter how rough. Just as group brainstorming works best when people think individually first, you need to capture your unique perspective before AI's suggestions can anchor you.”

Mollick (2025, Section The Creative Brain)

**→ zuerst eigene Gedanken machen, dann KI nutzen**

# Was hilft: Nutzungsszenarien von KI-Tools beim Schreiben (Buck & Limburg, 2024)



# Was hilft: Nutzungsszenarien von KI-Tools beim Schreiben (Buck & Limburg, 2024)

## Denkersatz

*Auslagerung des eigenen Denkens*

Beispiel:
Aufgabenstellung eins zu eins als Prompt in ein KI-Tool hochladen und die Ausgabe ohne kritische Prüfung übernehmen

Anforderungen:
keine, aber: ⚠️
nicht empfehlenswert!

## Werkzeug

*Abgabe kognitiv weniger anspruchsvoller Teilaufgaben*

Beispiele:
(a) Formulierung von Fliesstext aus Stichworten
(b) Korrektur/Umformulierung von Entwürfen
(c) Anpassung des Textstils

Anforderungen:
grundlegende Schreibkompetenz

## Schreib-begleitung

*Lösung von konkreten, während dem Schreiben auftretenden Problemen*

Beispiele:
(a) Unterstützung bei Gliederung/Einstieg
(b) Erklärung von Fach-begriffen/Textstellen
(c) Formulierungen/Synonyme

Anforderungen:
grundl. Schreibkompetenz, Wissen über Grenzen von KI

## Sparring-partnerin

*Steigerung der inhaltlichen Qualität*

Beispiele:
(a) Identifikation von Argumentationslücken
(b) Unterstützung bei Vergleich von Versionen
(c) weiterführende Ideen

Anforderungen:
hohe Schreibkompetenz, *AI literacy*

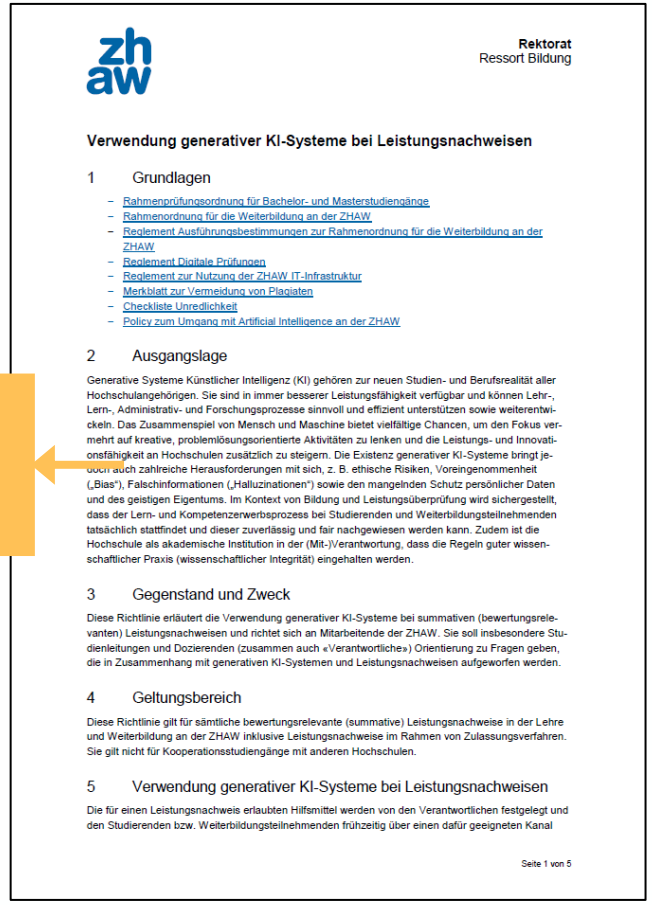
# Was hilft: Eigene Haltung beim Schreiben

- **KI ist keine (sinnvolle) Abkürzung**
  - richtig eingesetzt kann mit KI-Tools ein besseres Ergebnis erreicht werden
  - aber: wirklich schneller ist man dann nicht (z. B. Überprüfung, Steuerung)
  - Haltung: mit demselben zeitlichen Aufwand ein besseres Ergebnis erzielen

- **KI ist kein Ersatz von menschlichen Sparringspartner:innen**
  - menschlicher Austausch (d. h. Feedback von Betreuungspersonen, Mitstudierenden, Verwandten oder Freunden) bleibt wichtig
  - Haltung: KI-Tools als Ergänzung verwenden

# Was hilft: Richtlinien beachten!

## ZHAW-Richtlinien



Rektorat
Ressort Bildung

**Verwendung generativer KI-Systeme bei Leistungsnachweisen**

1 Grundlagen

- Rahmenprüfungsordnung für Bachelor- und Masterstudiengänge
- Rahmenordnung für die Weiterbildung an der ZHAW
- Reglement Ausführungsbestimmungen zur Rahmenordnung für die Weiterbildung an der ZHAW
- Reglement Digitale Prüfungen
- Reglement zur Nutzung der ZHAW IT-Infrastruktur
- Merkblatt zur Vermeidung von Plagiaten
- Checkliste Unredlichkeit
- Policy zum Umgang mit Artificial Intelligence an der ZHAW

2 Ausgangslage

Generative Systeme Künstlicher Intelligenz (KI) gehören zur neuen Studien- und Berufsrealität aller Hochschulangehörigen. Sie sind in immer besserer Leistungsfähigkeit verfügbar und können Lehr-, Lern-, Administrativ- und Forschungsprozesse sinnvoll und effizient unterstützen sowie weiterentwickeln. Das Zusammenspiel von Mensch und Maschine bietet vielfältige Chancen, um den Fokus vermehrt auf kreative, problemlösungsorientierte Aktivitäten zu lenken und die Leistungs- und Innovationsfähigkeit an Hochschulen zusätzlich zu steigern. Die Existenz generativer KI-Systeme bringt jedoch auch zahlreiche Herausforderungen mit sich, z. B. ethische Risiken, Voreingenommenheit („Bias"), Falschinformationen („Halluzinationen") sowie den mangelnden Schutz persönlicher Daten und des geistigen Eigentums. Im Kontext von Bildung und Leistungsüberprüfung wird sichergestellt, dass der Lern- und Kompetenzerwerbsprozess bei Studierenden und Weiterbildungsteilnehmenden tatsächlich stattfindet und dieser zuverlässig und fair nachgewiesen werden kann. Zudem ist die Hochschule als akademische Institution in der (Mit-)Verantwortung, dass die Regeln guter wissenschaftlicher Praxis (wissenschaftlicher Integrität) eingehalten werden.

3 Gegenstand und Zweck

Diese Richtlinie erläutert die Verwendung generativer KI-Systeme bei summativen (bewertungsrelevanten) Leistungsnachweisen und richtet sich an Mitarbeitende der ZHAW. Sie soll insbesondere Studienleitungen und Dozierenden (zusammen auch «Verantwortliche») Orientierung zu Fragen geben, die in Zusammenhang mit generativen KI-Systemen und Leistungsnachweisen aufgeworfen werden.

4 Geltungsbereich

Diese Richtlinie gilt für sämtliche bewertungsrelevante (summative) Leistungsnachweise in der Lehre und Weiterbildung an der ZHAW inklusive Leistungsnachweise im Rahmen von Zulassungsverfahren. Sie gilt nicht für Kooperationsstudiengänge mit anderen Hochschulen.

5 Verwendung generativer KI-Systeme bei Leistungsnachweisen

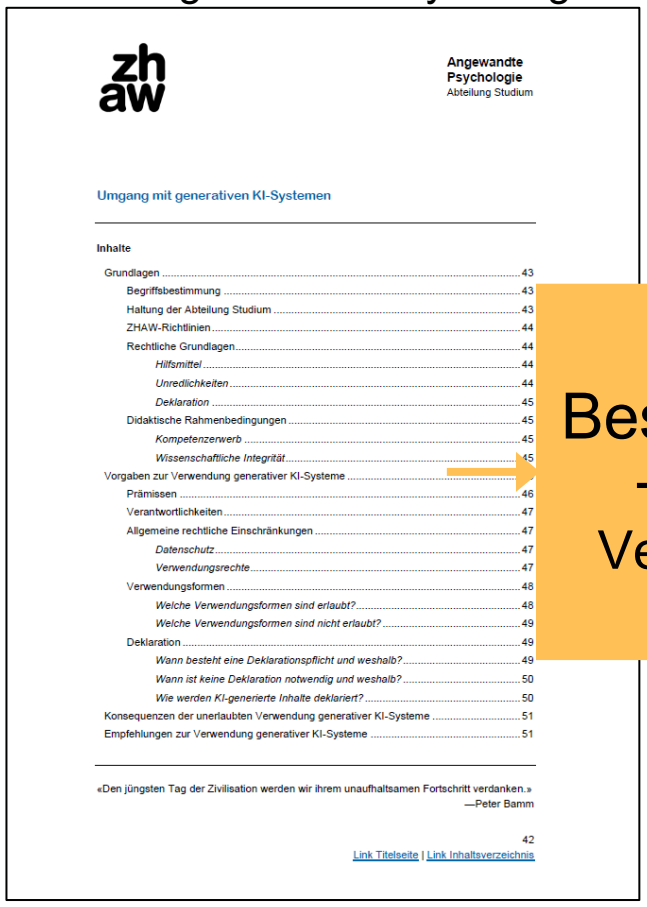
Die für einen Leistungsnachweis erlaubten Hilfsmittel werden von den Verantwortlichen festgelegt und den Studierenden bzw. Weiterbildungsteilnehmenden frühzeitig über einen dafür geeigneten Kanal

Seite 1 von 5

allgemeine Bestimmungen

## Richtlinien Studiengang

z. B. Angewandte Psychologie:



Angewandte Psychologie
Abteilung Studium

**Umgang mit generativen KI-Systemen**

Inhalte


Grundlagen ... 43
Begriffsbestimmung ... 43
Haltung der Abteilung Studium ... 43
ZHAW-Richtlinien ... 44
Rechtliche Grundlagen ... 44
*Hilfsmittel* ... 44
*Unredlichkeiten* ... 44
*Deklaration* ... 45
Didaktische Rahmenbedingungen ... 45
*Kompetenzerwerb* ... 45
*Wissenschaftliche Integrität* ... 45
Vorgaben zur Verwendung generativer KI-Systeme ... 46
Prämissen ... 46
Verantwortlichkeiten ... 47
Allgemeine rechtliche Einschränkungen ... 47
*Datenschutz* ... 47
*Verwendungsrechte* ... 47
Verwendungsformen ... 48
*Welche Verwendungsformen sind erlaubt?* ... 48
*Welche Verwendungsformen sind nicht erlaubt?* ... 49
Deklaration ... 49
*Wann besteht eine Deklarationspflicht und weshalb?* ... 49
*Wann ist keine Deklaration notwendig und weshalb?* ... 50
*Wie werden KI-generierte Inhalte deklariert?* ... 50
Konsequenzen der unerlaubten Verwendung generativer KI-Systeme ... 51
Empfehlungen zur Verwendung generativer KI-Systeme ... 51


«Den jüngsten Tag der Zivilisation werden wir ihrem unaufhaltsamen Fortschritt verdanken.»
—Peter Bamm

42

Link Titelseite | Link Inhaltsverzeichnis

spezifische Bestimmungen: u. a.
→ (nicht) erlaubte Verwendungsformen
→ Deklaration

# Fazit

“Generative AI pretends that there can be writing without writers, which is as nonsensical as suggesting there can be swimming without swimmers, or breathing without breathers.”

Morrison (2023, p. 160)

**→ Selbst schreiben bleibt wichtig!**

“The most important thing we can do is keep asking what to hand over and what to keep for ourselves… and not expect anyone, including the AI, to answer that for us.”

Mollick (2026, Section “Choosing what to keep human”)

**→ Immer überlegen, für welche Aufgaben ich KI wie nutze!**



https://stock.adobe.com/images/photo-of-cheerful-student-woman-using-laptop-while-doing-homework/357811866

# Literaturverzeichnis


Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bettschart, M., & Herrmann, M. (2026). Künstliche Intelligenz – Werkzeug, Schreibbegleitung und Sparringpartnerin. In M. Herrmann & B. Schmugge (Hrsg.), *Schreiben in Wissenschaft und Praxis: Ein Handbuch für Lehre, Studium und Beruf* (S. 729–816). Nomos. https://doi.org/10.5771/9783748960379

Böttcher, R. (2026, April 16). Was wir für Lernen halten, ist oft keines mehr! *Hochschulforum Digitalisierung*. https://hochschulforumdigitalisierung.de/lernen-und-genki/

Buck, I. (2025). *Wissenschaftliches Schreiben mit KI*. utb. https://doi.org/10.36198/9783838563657

Buck, I., & Limburg, A. (2024). KI und Kognition im Schreibprozess: Prototypen und Implikationen. *JoSch - Journal für Schreibwissenschaft, 15*(26), 8–23. https://doi.org/10.3278/JOS2401W002

Geyer, B. (2025, September 29). Nicht alles glauben, was KI sagt. *KI in Lehre und Weiterbildung*. https://barbarageyer.substack.com/p/nicht-alles-glauben-was-ki-sagt

Ma, D., Li, J., Wang, Z., Wang, Y., Kong, J., Zeng, W., Xiao, J., Yang, J., Zhang, W., Wang, B., & He, C. (2026). *CiteVQA: Benchmarking evidence attribution for trustworthy document intelligence* (arXiv:2605.12882). arXiv. https://doi.org/10.48550/arXiv.2605.12882

Mollick, E. (2025, July 7). Against “brain damage.” *One Useful Thing.* https://www.oneusefulthing.org/p/against-brain-damage

Mollick, E. (2026, Mai 26). Choosing to stay human. *One Useful Thing.* https://www.oneusefulthing.org/p/choosing-to-stay-human

Morrison, A. (2023). Meta-writing: AI and writing. *Composition Studies, 51*(1), 155–161. https://compstudiesjournal.com/wp-content/uploads/2023/06/morrison.pdf

Zinsser, W. (2006). *On writing well: The classic guide to writing nonfiction* (7th ed.). HarperCollins.

Zürcher Hochschule für
Angewandte Wissenschaften

**Angewandte Psychologie**
Abteilung Studium

Dr. Martin Bettschart
Dossierverantwortung «Wissenschaftliches Arbeiten»
besn@zhaw.ch
www.zhaw.ch