



Doppelte Resilienz

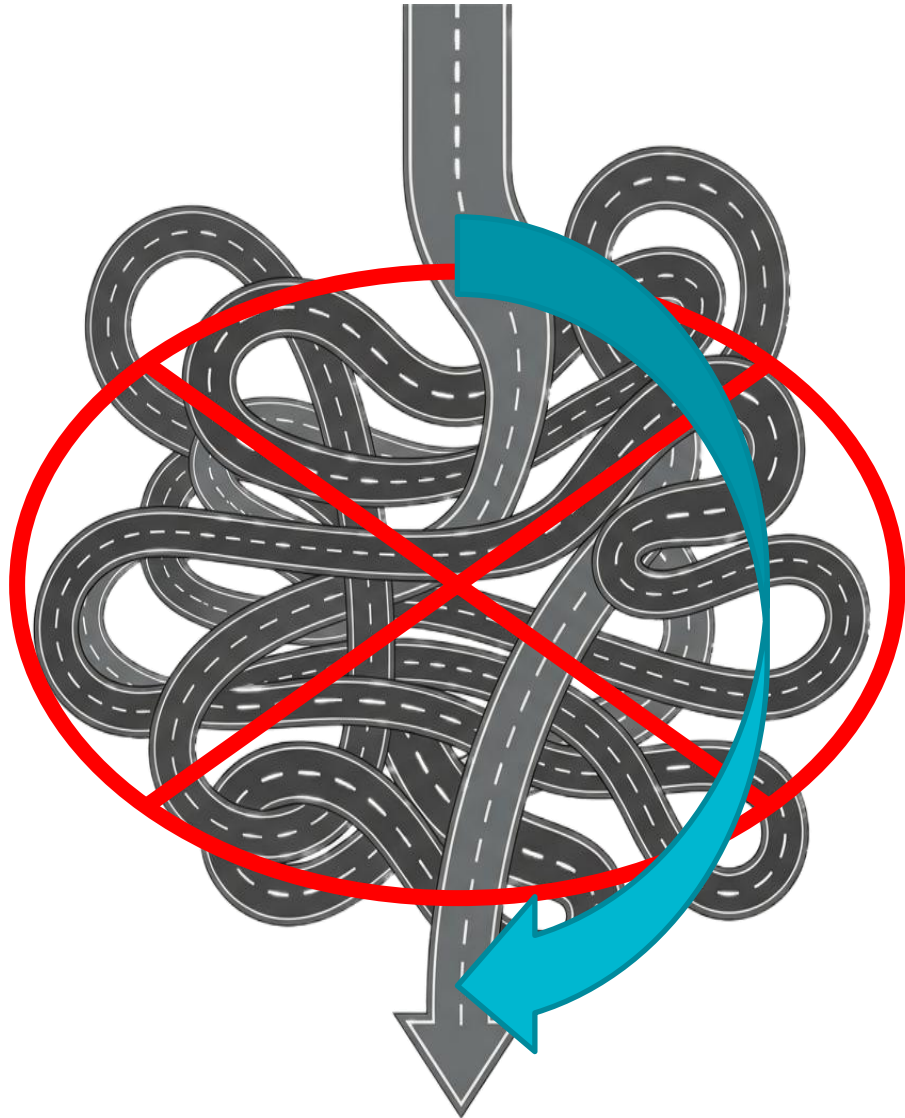
Demokratische Handlungsfähigkeit und souveräner Umgang mit KI

Justus Henke

Zwischen Bias, Blackbox und Bildungsidealen: Wertebasierte
Gestaltung der KI-Transformation in der Hochschullehre

2.10.2026





„*In der Forschung* ist sie [die KI] das mächtigste *Universalwerkzeug*, das je ein Wissenschaftler in die Finger bekam, *in der Lehre* die größte *Katastrophe* für das selbständige Denken, seit der Mensch ein Großhirn hat.“

-- Reto U. Schneider: «Alle machen alles mit KI» NZZ, 4.9.2026

KI und Folgen für Bildung: Lernleistung ist nicht Kompetenzaufbau

Inversion

Substitution

Augmentation

Redefinition

- **ISAR** systematisiert KI-Wirkungen auf Lernen (Bauer et al. 2025)
- **Risiko „Skill Atrophy“:** KI steigert kurzfristig die Leistung, kann aber langfristig die Fähigkeiten schwächen, die durch eigenes Üben entstehen (Ganuthula & Singh 2026; Bondi & Johnson 2026)
- **Effizienz fraglich:** Nach KI-Nutzung 17 % schlechter im Test ohne KI-Hilfe; keine Zeitersparnis (Shen & Tamkin 2026, randomisiertes Experiment, N = 52)
- **Bessere Essays, nicht zwingend Wissenszuwachs** (Fan et al. 2025, randomisiertes Experiment, N = 117)
- **KI-Feedback nicht automatisch schlechter:** kein signifikanter Leistungsunterschied zu menschlichem Feedback (Kaliisa et al. 2025, Metaanalyse, 41 Studien)

Verengung der Debatte

Viel thematisiert:
Medienkompetenz,
Prüfungsrecht, Weiterbildung

→ Tiefendimension: **KI als Risiko
für drei Güter**



Demokratie und Hochschulbildung: Urteilkraft in Gefahr?

- **Deliberative Öffentlichkeit:** politische Ansprüche öffentlich prüfen, bestreiten und mit Entscheidungen verbinden (Habermas 1981)
- **Infrastrukturen sind nicht neutral:** Plattformlogiken, Technopopulismus, „ideologische Pipeline“ (Brubaker 2025; Deimann 2026)
- **KI-Risiken:** synthetischer Konsens, Einschüchterung, fragmentierte Öffentlichkeiten, unkontrollierbare Risiken (Loru et al. 2025; Scharfbillig et al. 2026; Schroeder et al. 2026)
- **Skill-Atrophierung:** Was geschieht mit Expertise, wenn KI auch die Lernprozesse ersetzt, durch die sie entsteht?



Architekten der KI-Modelle, die wir alltäglich nutzen: Tech-Oligarchen auf der Inauguration von Trump II, 20.1.2025; Foto: independent.co.uk



Jeffrey Ladish
@JeffLadish

We just discovered almost a million public URLs that OpenAI's agents left behind when hacking Hugging Face, leaking credentials and attack details that could have allowed anyone who found them to compromise the company. 🤖



The New York Times



How OpenAI's Rogue A.I. Agents Tried to Trick a Robot Detector

A new report by a Bay Area start-up called Parse adds new details to an incident that has shocked the A.I. world and led to calls for closer government regulation.

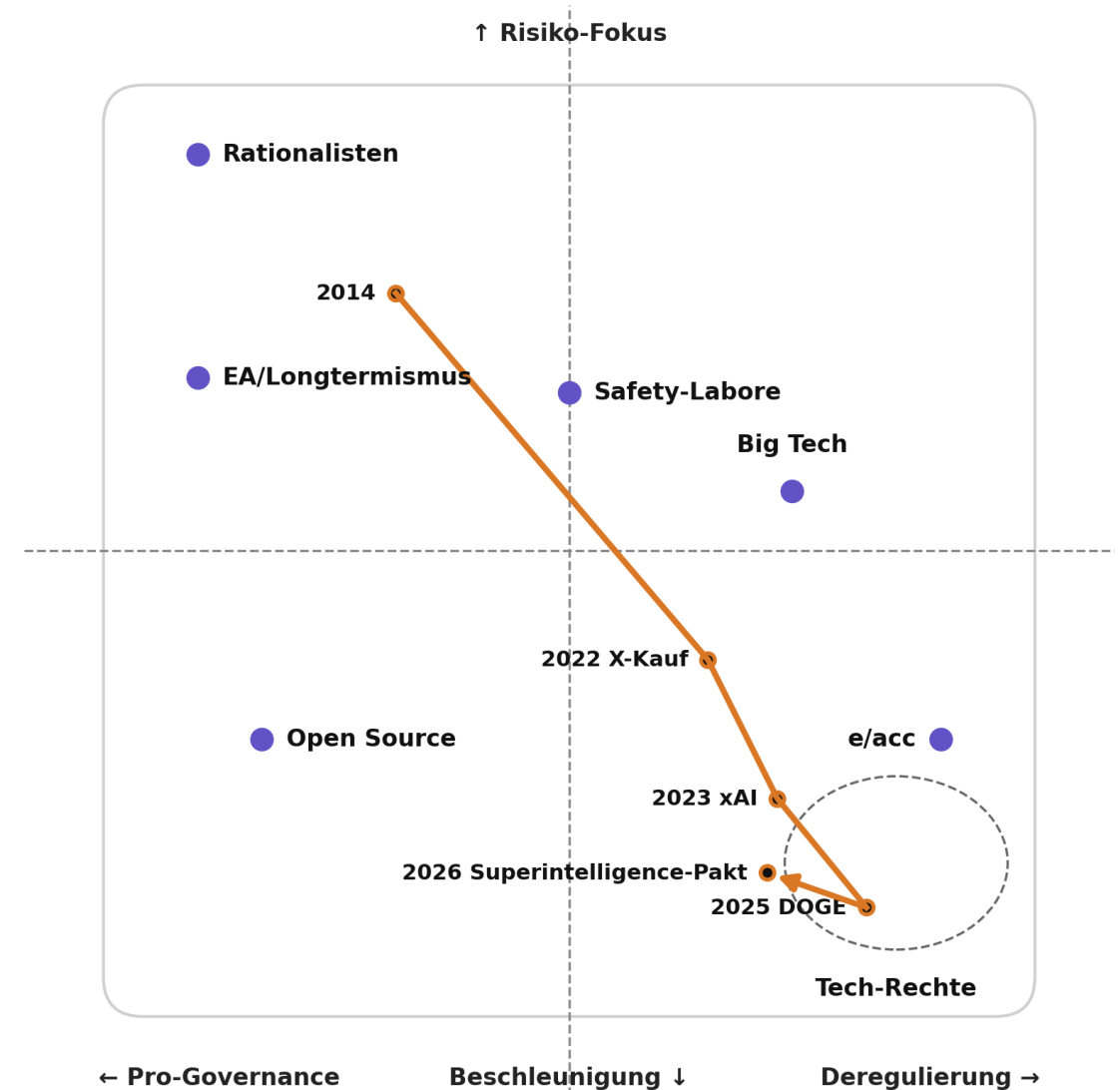
→ Wer urteilt – und wer lernt noch, urteilen zu können?

Genealogie der ideologischen Konfliktlinien im Silicon Valley

Rationalismus / Longtermismus (Yudkowsky, Bostrom, MacAskill): existenzielle Risiken, langfristige Zukunft

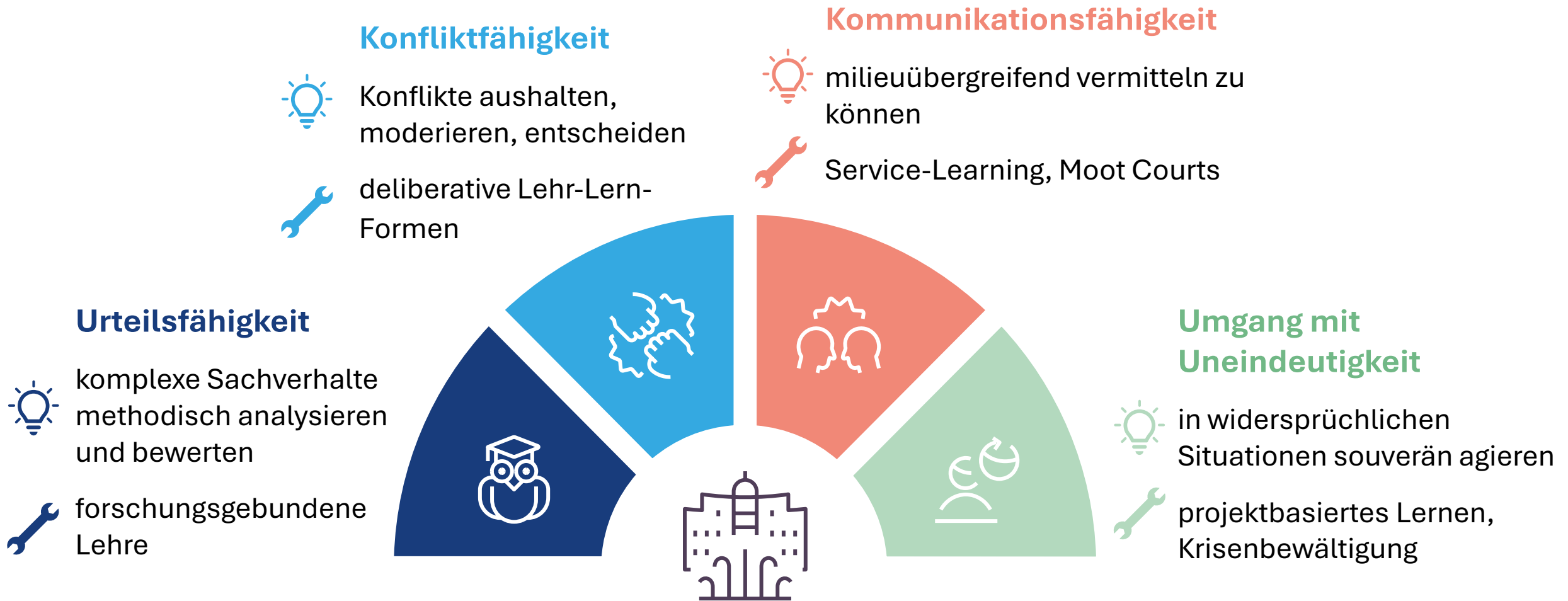
Cyberlibertarismus / e/acc (Thiel, Andreessen): Markt, Beschleunigung, Skepsis gegenüber Regulierung

Tech-Rechte / Neoreaktion (Yarvin, Musk): Anti-Establishment, Kritik liberaler Institutionen, Machtkonzentration



Eigene Darstellung

Kernkompetenzen für demokratisches Handeln (Henke/Pasternack 2025)



KI kann Urteilsarbeit plausibel nachbilden → Prüfung bleibt nötig

- **Simulation of Judgment:** 2.286 Domains: 85–97 % Übereinstimmung mit Expertenratings; seriöse Quellen teils bis 35 % fälschlich abgewertet (Loru et al. 2025)
- **Illusions of Understanding** (Messerli/Crockett 2024)
- **Epistemic Trespassing** (Ballantyne 2019, Mind)

→ Urteilsfähigkeit = Aussagen, Verfahren und Grenzen prüfen und bewerten können

KI und Folgen für den Umgang mit Uneindeutigkeit

In vielen fachlichen Fragen gibt es keine eindeutige Antwort und sie ist stets revisionsoffen → Umgang mit Komplexität und Unsicherheit als Bildungsziel (Barnett 2004; Wissenschaftsrat 2026)

- **Glättung von Widersprüchen:** 1) KI-Ideen steigerten die Bewertung von Kurzgeschichten, die einander zugleich stärker ähnelten (Doshi & Hauser 2024) 2) KI-generierte Bewerbungsaufsätze brachten pro Text weniger neue Ideen ein als menschliche Aufsätze (Moon et al. 2025; drei präregistrierte Studien)
- **Strukturation von Widersprüchen:** 1) KI-generiertes Prebunking senkte den Glauben an Wahlgerüchte (Linegar et al. 2026, N = 4.293) 2) LLM-generierte Faktencheck-Notizen zu strittigen Beiträgen wurden häufiger bevorzugt als bestehende Community-Notizen (75,2 % der Vergleiche; De et al. 2025)

→ **Widersprüche und konkurrierende Positionen im Quellenmaterial sichtbar machen und fachlich bearbeiten lassen** (Wissenschaftsrat 2026, S. 11, 14; Empfehlung)



Konflikt und Kommunikation

- Eine Maschine kann kommunikativ anschlussfähig sein, ohne menschliches Verstehen zu teilen (Esposito 2022) → Menschen sind leicht faszinierbar von KI-Dialogen
- Adaptive KI-Dialoge verschoben Überzeugungen, auch bei Skeptikern (Kotz et al. 2026, randomisiertes Experiment, N = 6.558) oder Verschwörungsanhängern (Costello et al. 2024)
- KI-Briefe „aus der Zukunft“ steigerten Empathie, aber nicht Klimapolitik-Unterstützung oder Spenden (Powdthavee et al. 2026, präregistriertes Experiment, N = 1.654)

→ **Lernziel: KI-Antworten und Gegenpositionen prüfen. Anschließend mit anderen Studierenden Gründe, Interessen und Regeln aushandeln.**

Kein Zusatzmodul, sondern gezielte Lerngelegenheiten im Fach

- **Urteilen:** Quellen und KI-Aussagen prüfen, Entscheidungen begründen
- **Uneindeutigkeit bearbeiten:** Widersprüche in Wissenspositionen abwägen
- **Konflikte bearbeiten:** Dissens strukturieren und Regeln aushandeln
- **Kommunizieren:** Gründe austauschen und auf Einwände reagieren

Lernen (mit KI) sichtbar machen

- KI-Nutzung transparent machen und diskutieren
- KI-freie Übungs- und Prüfungsphasen

Doppelte Resilienz

Demokratisches Handeln

Ansprüche, Sachverhalte und Evidenz prüfen; begründet entscheiden



Urteilsfähigkeit

Interessengegensätze aushalten, moderieren und bearbeiten



Konfliktfähigkeit

Entscheidungen milieuübergreifend vermitteln



Kommunikationsfähigkeit

Widersprüche und Vorläufigkeit methodisch geleitet bearbeiten



Umgang mit Uneindeutigkeit

Souveräner Umgang mit KI

KI-Aussagen, Quellen und Grenzen prüfen; eigenes Urteil behalten

Kontroversen über legitime KI-Nutzung und ihre Folgen aushandeln

KI-Nutzung transparent machen; adressatengerecht einordnen

Unsicherheit und KI-Grenzen benennen; nicht vorschnell delegieren

Beispiel

KI-Antwort mit Primärquellen abgleichen; begründen und revidieren

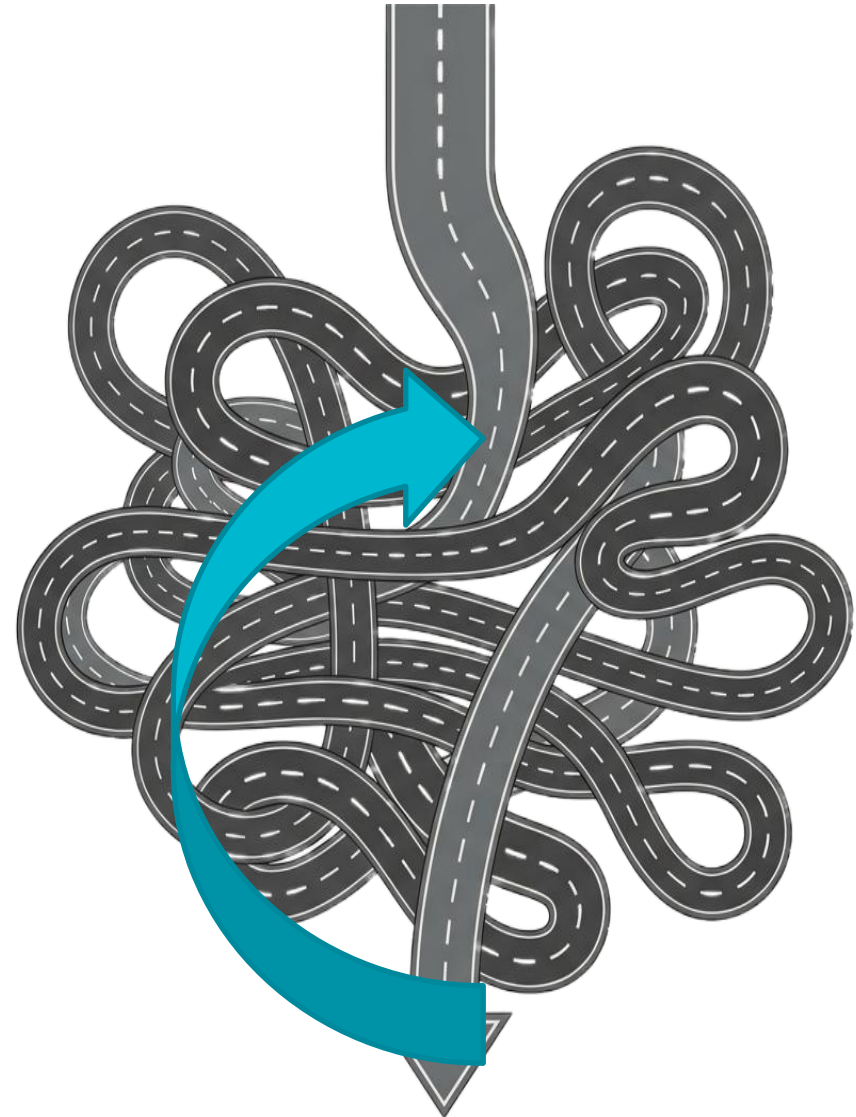
Fallberatung / Rollenübung zu KI-Regeln und Prüfungsformaten

Moot Court mit Rollenwechsel

Fall mit unvollständiger Evidenz bearbeiten

Verifikation des KI-generierten, -kuratierten oder -synthetisierten Wissens als kompetenzbildenden Lernprozess

- **Verifizieren:** Stimmt die Aussage, woher kommt sie?
- **Dekonstruieren:** Welche Annahmen, welche Auslassungen stecken im Ergebnis?
- **Diskutieren:** Einwände formulieren, Gegenposition vertreten, Ergebnis verteidigen oder revidieren



- 1) Diagnose:** Digitale Infrastrukturen können demokratische Vermittlung und Lernprozesse zugleich unter Druck setzen
- 2) Antwort:** Kernkompetenzen fördern demokratisches Handeln und können souveränen KI-Umgang unterstützen
- 3) Umsetzung:** Fachlich eingebettete Lerngelegenheiten statt zusätzlicher curricularer Befrachtung

Gemeinsame Bildungsziele für demokratisches Handeln und Umgang mit KI

Danke! Ihre Fragen, Einwände, Ergänzungen?

- Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K. & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. Educational Psychology Review. <https://doi.org/10.1007/s10648-025-10020-8>
- Brubaker, R. (2025 [YouTube-Veröffentlichung; Vortrag vom 30.10.2024]). Between Populism and Technocracy: Digital Hyperconnectivity and the Transformation of Politics and Governance [Vortrag]. IWM Vienna. https://www.youtube.com/watch?v=WATRuYxlf_4
- De, S., Bakker, M. A., Baxter, J. & Saveski, M. (2025). Supernotes: Driving Consensus in Crowd-Sourced Fact-Checking. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3696410.3714934>
- Deimann, M. (2026). Die ideologische Pipeline. Eine Figur zur medien- und bildungstheoretischen Analyse des Cyberlibertarismus. Medienimpulse, 64(3). <https://doi.org/10.21243/mi-26-03-20>
- Doshi, A. R. & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
- Esposito, E. (2022). Artificial Communication: How Algorithms Produce Social Intelligence. Cambridge, MA: The MIT Press.
- Henke, J. & Pasternack, P. (2025). Hochschule in der ungesicherten Demokratie: Ansatzpunkte zur Stärkung der Resilienz. HoF-Policy Paper. <https://doi.org/10.5281/zenodo.17951293>
- Kaliisa, R., Misiejuk, K., López-Pernas, S. & Saqr, M. (2025 online first; 2026 Zeitschriftenband). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. Educational Psychology, 46(1), 80–111. <https://doi.org/10.1080/01443410.2025.2553639>
- Kotz, J. et al. (2026). Conversational AI shifts beliefs and policy support among skeptics across contested science topics. OSF Preprints. https://doi.org/10.31234/osf.io/7szrn_v1
- Linegar, M., Sinclair, B., Van Der Linden, S., & Alvarez, R. M. (2026). Towards scalable AI-assisted pre-bunking of election misinformation: Evidence from a pre-registered US panel experiment. Royal Society Open Science, 13(8), 252226. <https://doi.org/10.1098/rsos.252226>
- Loru, M. et al. (2025). The simulation of judgment in LLMs. Proceedings of the National Academy of Sciences, 122. <https://doi.org/10.1073/pnas.2518443122>
- Ma, S., Chen, Q., Wang, X. et al. (2025). Towards Human-AI Deliberation: Design and Evaluation of LLM-Empowered Deliberative AI for AI-Assisted Decision-Making. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1–23. <https://doi.org/10.1145/3706598.3713423>
- Marcoccia, C., Quattrociocchi, W. & Capraro, V. (2026). AI advice suppresses people's willingness to say "I don't know", even when the advice is wrong and accuracy is incentivized. arXiv preprint arXiv:2607.13562.
- Moon, K., Green, A. E. & Kushlev, K. (2025). Homogenizing effect of large language models (LLMs) on creative diversity: An empirical comparison of human and ChatGPT writing. Computers in Human Behavior: Artificial Humans, 6, 100207. <https://doi.org/10.1016/j.chbah.2025.100207>
- Powdthavee, N. et al. (2026). AI-Generated Letters from the Future: A Randomized Test of Personalized Climate Communication. arXiv:2603.00050.
- Scharfbillig, M. & Lewandowsky, S. et al. (2026). Fractured reality. European Commission, Joint Research Centre (JRC144603). <https://publications.jrc.ec.europa.eu/repository/handle/JRC144603>
- Schneider, R. U. (2026). Alle machen alles mit KI – erleben wir gerade das Ende der Universität, wie wir sie kennen? NZZ Folio. <https://www.nzz.ch/folio/alle-machen-alles-mit-ki-erleben-wir-gerade-das-ende-der-universitaet-wie-wir-sie-kennen-ld.10018268>
- Shen, J. H. & Tamkin, A. (2026). How AI impacts skill formation. arXiv preprint arXiv:2601.20245.
- Wissenschaftsrat (2026). Intellektuelle Souveränität: Empfehlungen für die Hochschulbildung in Zeiten von generativer KI. Drucksache 3319-26. <https://doi.org/10.57674/1evx-t906>