PÄDAGOGISCHE
HOCHSCHULE
ZÜRICH

PH ZH

Leistungsbeurteilung und KI:

Ein Shift zu integrierten Beurteilungen ist nötig

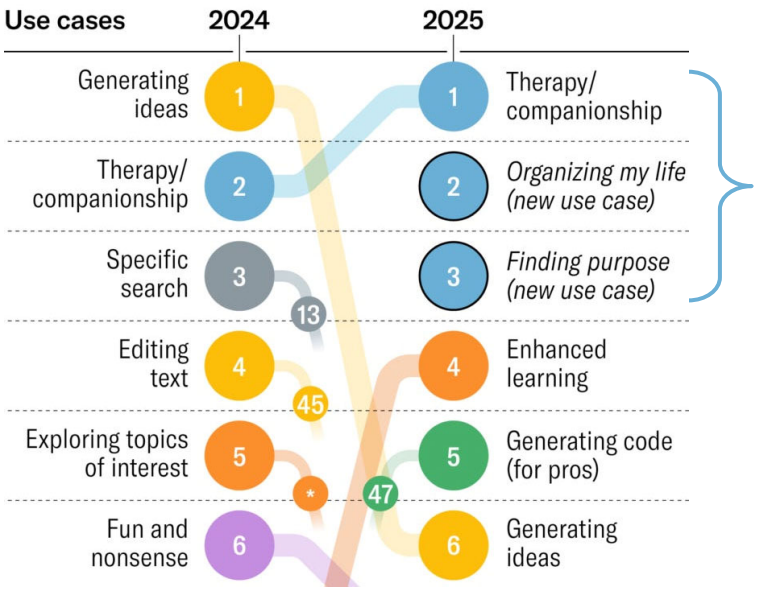
Tobias Zimmermann

Juni 2025

Lagerstrasse 2 8090 Zürich **phzh.ch**

1

KI: Top Use Cases 2024 ⇔ 2025



Wichtigste Veränderung:

++ **Persönlicher und beruflicher Support**

(Quelle: Zao-Sanders 2025)

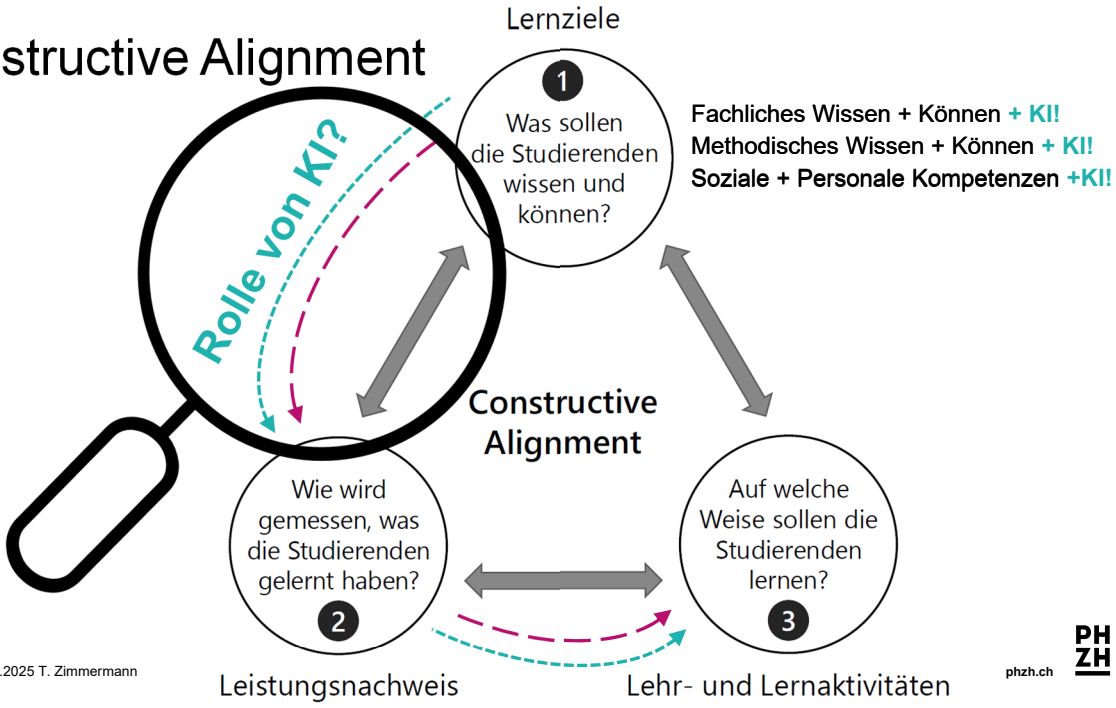
2

Inhaltsverzeichnis

1. Wie wirkt sich die Existenz von KI auf das Constructive Alignment aus?
2. KI, Lernen und Wissen: Risiken und Chancen
3. Was hat Bedeutsamkeit mit KI zu tun?
4. Risiken für die Leistungsbeurteilung im Zeitalter von KI
5. Zukunftsfähige Leistungsbeurteilung: Learning as Assessment

3

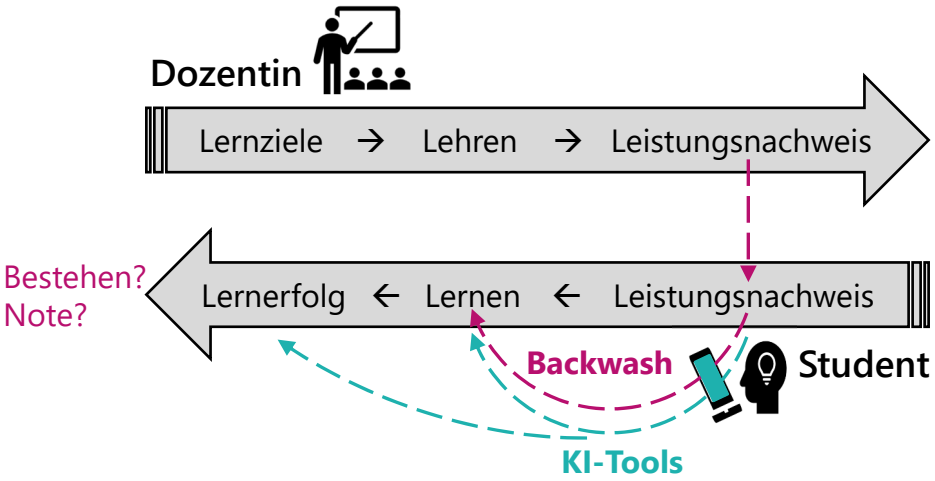
Constructive Alignment



Fachliches Wissen + Können + KI!
Methodisches Wissen + Können + KI!
Soziale + Personale Kompetenzen +KI!

4

Unterschiedliche Perspektiven: Backwash



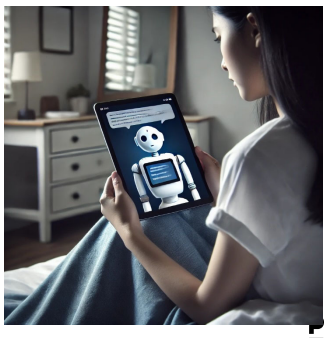
Aktuelle Herausforderungen – Tendenz zunehmend

Studierende und Dozierende!

Overtrust

Deskilling

Social Narrowing



KI und Qualitäten des Lernens

Mensch lässt KI kognitiv komplexe Aufgaben erledigen
Beispiel: KI verfasst Essay zu einem ethischen Dilemma
→ Deskilling und Overtrust

Mensch bearbeitet kognitiv komplexe Aufgaben im intensiven Austausch mit KI
Beispiel: Struktur und Inhalt dieser Darstellung wurden im iterativen Austausch mit zwei KI-Instanzen erarbeitet

Kognitive Tiefe der «Aufgabe»

Interaktionsgrad

Social Narrowing?

Mensch lässt KI kognitiv anspruchslose Aufgaben erledigen
Beispiel: Rechtschreibkorrektur durch ChatGPT

Mensch bearbeitet kognitiv anspruchslose Aufgaben im intensiven Austausch mit KI
Beispiel: Gamifiziertes Vokabellernen mit KI (Duolingo)

7 29. Januar 2025 | Tobias Zimmermann

PH ZH

7

Wieso muss ich etwas können, wenn es die KI für mich erledigen kann?

Wir können KI nur eigenverantwortlich einsetzen, wenn wir das, was die KI kann und weiss, grundsätzlich ebenfalls wissen und können (vgl. Taschenrechner und Kopfrechnen).

→Einzelpersonen

→(Wissens-)Gesellschaft

ABER:

8 10.6.2025 T. Zimmermann

8

Im Fahrersitz bleiben

Kognition

«Falling Asleep at the Wheel» (Dell' Acqua 2022):

- Wir neigen dazu, uns blind auf «gute» KI zu verlassen
  → reduzierte Aufmerksamkeit, schlechtere Entscheidungen
- «Schlechtere» KI hält uns wachsamer
  → kritischer Umgang mit Output, bessere Entscheidungen

Motivation

Selbstbestimmungstheorie:

Motiviert bleiben wir, wenn wir uns

- handlungsfähig
- kompetent und
- sozial eingebunden fühlen.

Das beste Zusammenspiel entsteht, wenn wir aktiv und kritisch mit der KI-Instanz umgehen.

Mensch-KI-Kombinationen können mehr erreichen

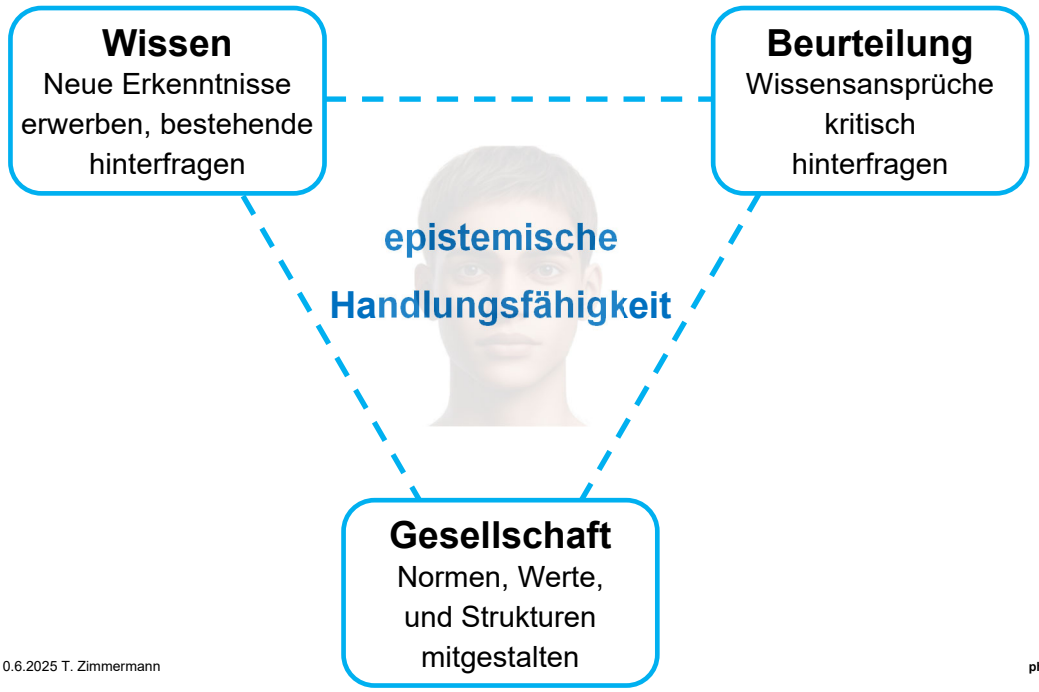
- Radiologie

- Softwareentwicklung

- Personalisiertes Lernen

- Kunst und Kreativität

- Barrierefreiheit



11 10.6.2025 T. Zimmermann phzh.ch

11

Epistemische Handlungsfähigkeit

Wie orientieren wir uns in einer komplexen, wissensbasierten und stark medial vermittelten Gesellschaft?

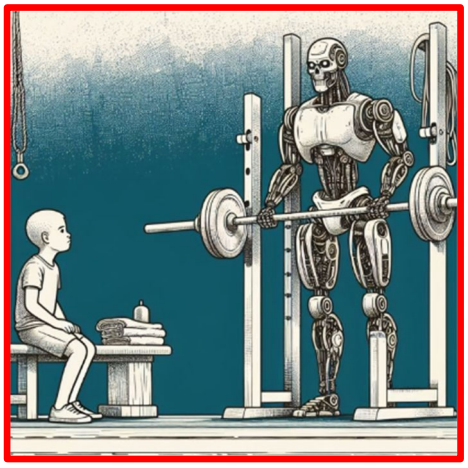
→ Wir benötigen epistemische Handlungsfähigkeit

Definition
Epistemische Handlungsfähigkeit beschreibt die Fähigkeit, Wissen in einem kritischen und sozialen Kontext zu bewerten, zu produzieren und zu transformieren. Sie umfasst die Entwicklung epistemologischer Reife hin zu einer evaluativen Perspektive, in der Unsicherheit akzeptiert und Meinungen anhand von Argumenten und Evidenz verglichen werden. (Nieminen und Ketonen 2023)

12 10.6.2025 T. Zimmermann phzh.ch

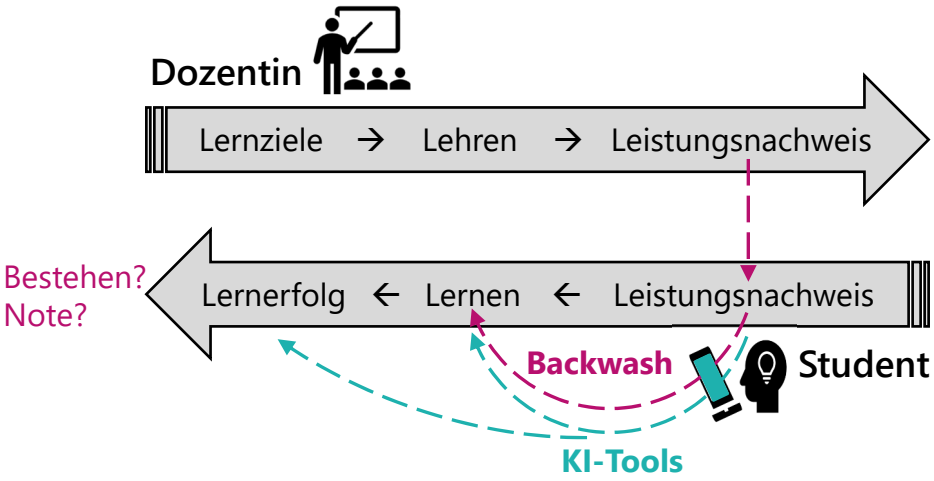
12

KI lernt nicht für mich…



…kann mich aber dabei unterstützen!

Unterschiedliche Perspektiven: Backwash



Zur Rolle von Bedeutsamkeit beim Lernen

«Die Durchführung einer Handlung genügt noch nicht, um etwas daraus zu lernen – auch nicht, wenn sie wiederholt durchgeführt wird. Entscheidend ist, dass die Tätigkeit und/oder ihr Resultat als bedeutsam empfunden werden. Dann erfolgen neurologische Vorgänge, welche die betroffenen Stellen im neuronalen Netzwerk des Hirns plastisch und somit offen für Veränderung machen (vgl. Eagleman 2020, 149–53).»

(Zimmermann 2024, S. 99f.)

Zahlenbasierte Anreizsysteme, die ausserhalb der eigentlich durch sie zu steuernden Tätigkeit liegen, schaffen extrinsische Motivation. Dies gilt für Fallzahlen in Krankenhäusern, Abschlussquoten in Schulen oder den h-Index von Forschenden. Zum Ziel wird schliesslich das gute Abschneiden im Anreizsystem, das meistens auch relativ unabhängig von der Qualität der eigentlich zu steuernden Tätigkeit erreicht werden kann.

(Zimmermann 2024, S. 101)

Noten: → Studierende/SuS wollen gute Noten erreichen = extrinsische Motivation
→ intrinsische Bedeutsamkeit der Lerninhalte wird ausgeblendet
→ Es wird weniger und weniger nachhaltig/tiefgreifend gelernt

phzh.ch

15 10.6.2025 T. Zimmermann

15

Gaming the System: Einfacher mit KI

GtS = «Spielregeln» formal einhalten, Zweck unterminieren

Klassisches GtS:

- Spitäler fokussieren risikoarme Operationen; Insider nutzen Informationen für Aktienhandel
- Schwächere SuS sind am Tag der PISA-Testung systematisch «abwesend»
- SuS/Studierende, die Aufgaben abschreiben, indem sie diese paraphrasieren

GtS mit KI:

- Berechnungen, wie sich formale Anreizsysteme (aus)nutzen lassen, werden zugänglicher
- Paraphrasieren und Übersetzen übernimmt KI in Sekundenschnelle
- KI löst auch viele komplexe Aufgabenstellungen (sehr schnell)

16 10.6.2025 T. Zimmermann phzh.ch

16

Risiko #1: KI aus Leistungsbeurteilung ausschliessen

Technische Probleme

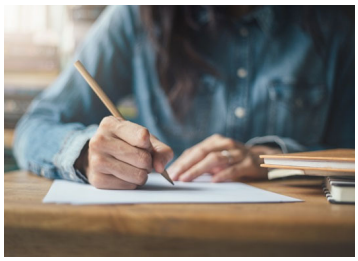
- Technisch kann kaum verhindert werden, dass KI zum Einsatz kommt (ist fast überall «drin»)
- KI-Gebrauch nicht detektierbar (gilt für Texte - wohl bald auch für Bilder/Videos)

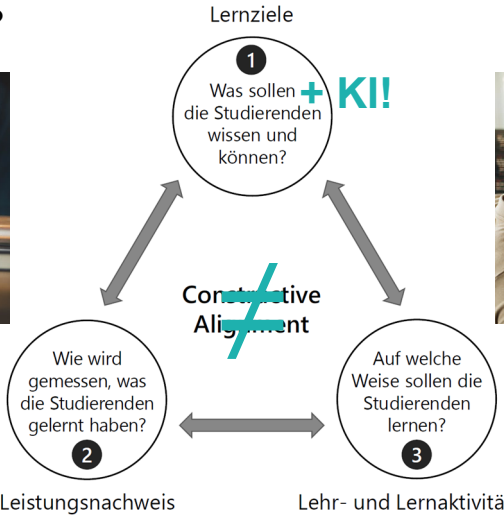
Pädagogische Probleme

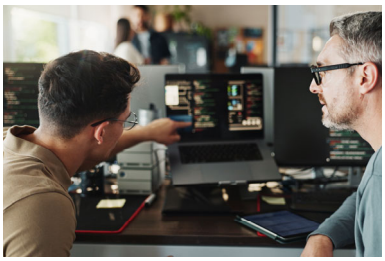
- Studierende sollen für den beruflichen Alltag fit gemacht werden – dazu gehören auch KI-Kompetenzen!
- Reduktion auf Papier, Bleistift und mündliche Prüfungen führt zu unerwünschtem Lernverhalten: Begünstigt Bulimie-Lernen – in kurzer Zeit Gelerntes geht grösstenteils wieder vergessen oder bleibt träges Wissen.
- Lebensweltbezug? Studierende können epistemische Handlungsfähigkeit kaum ausbauen

Risiko #1: KI aus Leistungsbeurteilung ausschliessen

Fit für beruflichen Alltag?

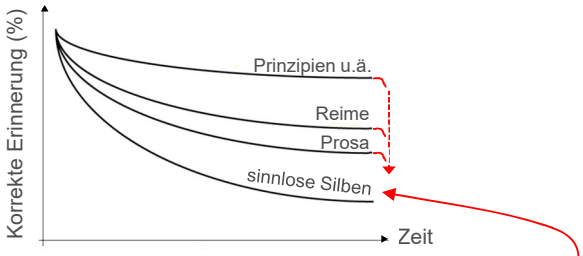






Risiko #1: KI aus Leistungsbeurteilung ausschliessen

Auswendiglernen, Closed Books, Prüfungsgespräche:



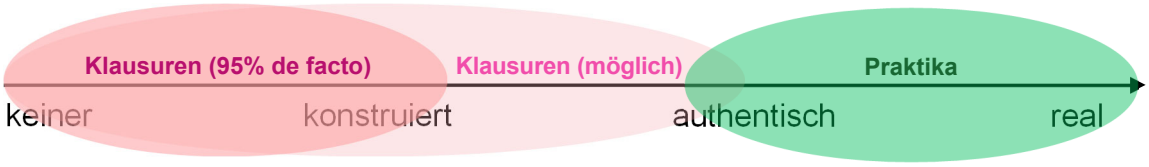
→ Kaum nachhaltiges Lernen (vieles nach wenigen Wochen wieder vergessen)

→ Anwendungsbezug schwer herzustellen → für Lernende wenig sinnhaft

→ Lückenhaftes und träges Wissen

Risiko #1: KI aus Leistungsbeurteilung ausschliessen

Lebensweltbezug von Aufgaben



Risiko #2: So tun, als gäbe es KI nicht

Risiko #2a: Empfehlen, KI nicht zu verwenden

Vollständige Liste aller Vorteile dieses Ansatzes

Risiko #3: KI übernimmt summative Bewertung (Noten)

Ethische und pädagogische Probleme (exemplarisch)

- Ethik: Beurteilung von Menschen sollte Aufgabe von Menschen bleiben (oder möchte irgendjemand Rechtsfälle vor einem KI-Gericht austragen?)
- Motivation: Wenn Studierende von Maschinen beurteilt werden, führt das im Sinne der Selbstbestimmungstheorie zu einer Demotivationsspirale – für Studierende wie Dozierende.

Technische Probleme

- Blackbox: Die Funktionsweise von KI-Systemen ist bis auf Weiteres so komplex, dass sie nicht als transparent eingeschätzt werden kann. KI-Systeme sind auch nicht selbstreflexiv, können als nicht über ihr eigenes Denken Auskunft geben.
- Ein überzeugendes Beispiel, wie KI zuverlässig komplexe Urteile vornehmen kann, steht noch aus (jenseits der ethischen und pädagogischen Probleme).

Risiko #3: KI übernimmt summative Bewertung (Noten)

Ethische und pädagogische Probleme (exemplarisch)

- Ethik: Beurteilung von Menschen [illegible] bleiben (oder möchte irgendjemand Re[illegible]ragen?)
- Motivation: Wenn Studierende [illegible] wurden, führt [illegible] im Sinne der Selbst[illegible] zu einer Demotivation [illegible] Studierende wie Dozierende.

Technische Probleme

- Black[illegible] Antwortweise von KI-Systemen [illegible] dass sie nicht als transparent [illegible] können als nicht über ihr eigenes Denken Auskunft geben.
- Ein überzeugendes Beispiel, wie KI zuverlässig komplexe Urteile vornehmen kann, steht noch aus (jenseits der ethischen und pädagogischen Probleme).

Chance: Formativer Gebrauch

- KI-Bot, der Feedback zu Laborprotokollen gibt, bevor sie offiziell eingereicht werden
- KI-Bot unterstützt Studierende bei der Vorbereitung auf eine mündliche Prüfung
- etc.

Risiken

Risiko Nr. 1: Ausschluss von KI aus der Leistungsbeurteilung

Risiko Nr. 2: KI weitgehend ignorieren

Risiko Nr. 3: KI übernimmt die summative Beurteilung (Noten/Prädikate)

Chance: Learning as Assessment

Assessment of Learning: Summative Beurteilung
Assessment for Learning: Formative Beurteilung
Assessment as Learning: Selbstevaluative Beurteilung

} Beurteilung als Add-On, als Wurmfortsatz des Lernprozesses

→ «Learning as Assessment»:
Lernformen, in denen die erfolgreiche Bewältigung der Aufgabenstellung einen erheblichen Teil der Lernziel-Erreichung gewährleistet.

→ exemplarisches Lernen
→ kritisches Denken
→ kreatives Problemlösen
→ Wissenstransfer, vernetztes Denken/Handeln

} Epistemische Handlungsfähigkeit

Learning as Assessment: Beispiele

Alle Beispiele:

– Prozess- und Produktbeurteilung möglich

– Gelegenheiten für formative Rückmeldungen (Dozierende und/oder Peers)

– Mehrphasige Leistungsbeurteilungen möglich: formativ → formativ → summativ

Learning as Assessment: Beispiel 1

Challenge-based Learning

- Komplexe, realitätsnahe Problemstellungen eigenständig bearbeiten
- Kreative Lösungsansätze entwickeln und umsetzen
- Iterativer Prozess: kontinuierliches Weiterentwickeln (Lösungsweg oft nicht linear)
- «Misserfolgskompetenz»: Überwinden von Durststrecken, Entwickeln von kognitiver und affektiver Resilienz
- Neben Lösung/Produkt kann auch Prozess (Zwischenprodukte, Reflexionen, Gespräche, Journal/Portfolio) beurteilt werden

Bildquelle: Adobe Stock

Learning as Assessment: Beispiel 2

Service-oriented Learning

- Bearbeiten sozialer oder gesellschaftlicher Herausfor-derungen ist sinnstiftend und motivierend
- Theorie, Konzepte und Modelle in realen Kontexten anwenden und vertiefen
- Multiperspektivität: Kontakt mit verschiedenen Menschen in diversen Kontexten erweitert die soziale Wahrnehmung
- Erfahrungen reflektieren und mit persönlichen Werten verbinden (Gespräche, Journal, Portfolio)
  → Beurteilungsmöglichkeit

Bildquelle: Adobe Stock

Learning as Assessment: Beispiel 3

(Teilweise) begleitete Praktika

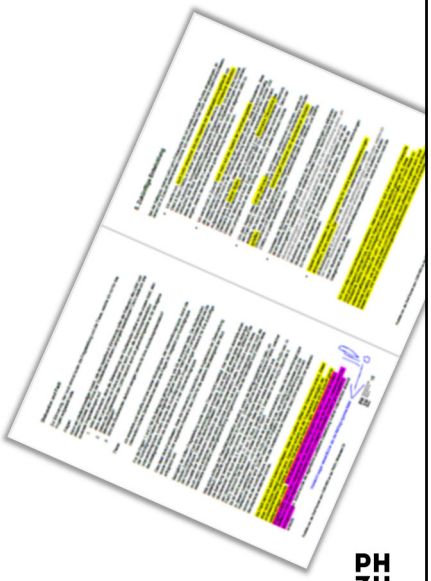
- Theorie-Praxis-Bezug: Studierende wenden Wissen in realen beruflichen Kontexten an
- Gezielte Begleitung und Förderung durch Mentor/-innen
- Dokumentation und Reflexion des Lernprozesses (Implizites explizit machen)
- Mitverantwortung: Studierende gestalten Praktikum aktiv mit

Bildquelle: OST Ostschweizer Fachhochschule
www.ost.ch/de/studium/technik/bachelor-erneuerbare-energien-und-umwelttechnik/studieninhalt-und-aufbau/vertiefung-solartechnik

Learning as Assessment: Beispiel 4

Portfolioarbeit mit persönlichen Entwicklungszielen und regelmässigen formativen Gesprächen

- Studierende setzen (auch) eigene Entwicklungsziele
- Regelmässige Portfoliogespräche mit Dozierenden (und evtl. Peers) fördern Reflexion
- Summative Beurteilung spielt nur eine untergeordnete Rolle (idealerweise als abschliessendes Pass/Fail)



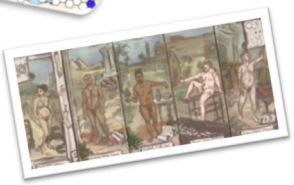
Learning as Assessment: Beispiel 5

Posterwerkstatt und Posterpräsentation (nicht: Poster als Produkt)

- Feedback von Peers und Dozierenden verbessert Poster und Präsentationsskills
- Austausch in Werkstatt und Präsentation ist sinnstiftend und motivierend
- Förderung visueller Darstellungskompetenzen
- Wissenschaftskommunikation: Wirksames Präsentieren von Forschungsergebnissen

Bildquelle: Adobe Stock

Learning as Assessment: Beispiel 6



Unessay

- Vielfalt: Erkenntnisse und Resultate in vielfältigen Formaten (z.B. Videos, Podcasts, Spiele) präsentieren
- Fördert Kreativität und Individualität
- Offene Form regt kreative, kritische Auseinandersetzung an
- Austausch über Formate unterstützt die Reflexion von Lernprozessen und Lernwegen

Bildquelle: https://hopkinshistoryofmedicine.org/undergraduate-courses/unessay-project-as-140-105/

Learning as Assessment: Beispiel 7

Einen KI-Bot trainieren – Beispiel

- Interviews mit echten Menschen (Bankkunden, Patienten, SuS, ...)
- Transkription der Interviews, qualitative Analyse (KI-Unterstützung möglich)
- KI-Bot mit Transkripten und Analysebefunden trainieren, bis er interviewte Menschen möglichst gut imitieren kann
- Führen eines „Laborjournals“ von der Vorbereitung der Interviews bis zum Testen des KI-Outputs

33 10.6.2025 T. Zimmermann

Bildquelle: Adobe Stock

phzh.ch

33

Learning as Assessment: Beispiel 8

Einen KI-Bot unterrichten

- KI-Bot wird von Lehrenden so gepromptet, dass er bestimmten Könnensstand aufweist und sich wie ein lernender Mensch verhält
- Studierende unterrichten Bot, bis er das Lernziel erreicht (≈ Lernziel der Studierenden)
- Studierende führen ein «Unterrichtsjournal»
- Evtl. Bot-Unterricht in mehreren Etappen, dazwischen Feedback von Dozierenden/Peers

34 10.6.2025 T. Zimmermann

Bildquelle: Adobe Stock

phzh.ch

34

Hinweis auf Kurztagung:

Assessment und KI: Gemeinsam Praktische Ansätze erkunden

20. November 2025, Tagungszentrum Schloss Au, Wädenswil

→ http://tiny.phzh.ch/auki

phzh.ch

35

Literatur

Dell' Acqua, Fabrizio. 2022. «Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters.» Working Paper. https://www.almendron.com/tribuna/wp-content/uploads/2023/09/falling-asleep-at-the-whee.pdf Zugriff 26.1.2025.

Eagleman, David. 2020. *Livewired: The Inside Story of the Ever-Changing Brain*. Edinburgh: Canongate Books.

Hardman, Philippa. 2025. Your Learners Are Using AI to Redesign Your Courses. https://drphilippahardman.substack.com/p/your-learners-are-using-ai-to-redesign. Zugriff 6.6.2025.

Maier, Uwe, Marc Kleinknecht, Kerstin Metz und Thorsten Bohl. 2010. «Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben.» *Beiträge zur Lehrerbildung* 28 (1): 84–96. doi:10.25656/01:13734.

Moldoveanu, Mihnea und George Siemens. 2024. «Interactionalism:Re-Designing Higher Learning for the Large Language Agent Era.» https://doi.org/10.48550/arXiv.2501.00867. Zugriff 26.1.2025

Mollick, Ethan. 2024a. *Co-Intelligence. Living and Working with AI*. New York: Penguin Random House.

Nieminen, Juso Henrik, und Laura Ketonen. 2023. «Epistemic Agency: A Link Between Assessment, Knowledge and Society.» *Higher Education* (88), 777–794. doi: 10.1007/s10734-023-01142-5. (Open Access)

Reinmann, Gabi. 2023. «Deskilling durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik.» Diskussionspapier Nr. 25. Berlin: Hochschulforum Digitalisierung. https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_DP_25_Deskilling.pdf. Zugriff 26.1.2025

Zao-Sanders, Marc. 2025. How People Are Really Using Gen AI in 2025. Harvard Business Review. https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025. Zugriff 6.6.2025

phzh.ch

36 10.6.2025 T. Zimmermann

36

Literatur

Zimmermann, Tobias. 2024. *Leistungsbeurteilungen an Hochschulen lernförderlich gestalten. Prüfen, Beurteilen und Rückmelden von Lernleistungen*. Opladen, Berlin, Toronto: Verlag Barbara Budrich.

z.B. Kapitel 7: Formate von Leistungsnachweisen

Kapitel 8: Formative Leistungsrückmeldung («Feedback»)

Kapitel 9: Summative Beurteilen: Bewerten und Benoten

Kapitel 15: Künstliche Intelligenz und menschliche Leistung

37

Bonus: Die beliebtesten Prompts zum Lernen

Raffinierte Lernstrategien

– «Ich möchte durch Lehren lernen. Stelle mir Fragen zu [Thema], damit ich dir die Kernkonzepte erklären kann.»
«Erstelle mir anhand des Materials ein Übungsquiz. Stellen mir jede Frage einzeln.»

Emotionale und motivationale Bedürfnisse

– «Ich fühle mich heute lustlos. Kannst du mir helfen, diese Vorlesung trotzdem zu verstehen?»

Zugänglichkeit und Personalisierung

– «Kannst du mir anhand der folgenden Folien helfen, den Inhalt interessanter zu machen, so dass ich ihn besser lernen kann?»
– «Entwickle ein Spiel, das mir hilft, [Lernziele] zu erreichen.»

Quelle: https://edunewsletter.openai.com/p/top-20-chats-for-finals
vgl. auch Hardman 2025

38

Bonus: Die beliebtesten Prompts zum Lernen

Prozessunterstützung und Scaffolding

- «Gib mir eine Schritt-für-Schritt-Anleitung, damit ich [Projekt] fertigstellen kann. Die Schritte sollten so klein und erreichbar wie möglich sein.»
- «Sei mein Coach für öffentliche Reden und gib mir Feedback, damit ich mich verbessern kann.»

Kritisches Denken und akademische Genauigkeit

- «Ich möchte meine These einem Stresstest unterziehen, bevor ich weiter schreibe. Schlage mir bestehende Gegenargumente und eventuelle Schwachstellen in meiner Argumentation vor.»
- «Ich möchte mehrere Perspektiven berücksichtigen. Finde drei Expert*innen mit unterschiedlichen Standpunkten und vergleiche deren Meinungen.»

Quelle: https://edunewsletter.openai.com/p/top-20-chats-for-finals
vgl. auch Hardman 2025