10 MINUTEN DigiTaL

Die wöchentliche 10-Minuten-Reihe | DigiTaL-Projekt
Digital Transformation Lab for Teaching and Learning

→ **DigiTaL | Lehrservice**

Stiftung Innovation in der Hochschullehre

LEUPHANA
UNIVERSITÄT LÜNEBURG

Feedback durch künstliche Intelligenz – eine Revolution in der Hochschulbildung?

10 Minuten DigiTal

Lucas Jacobsen & Kira Elena Weber

Leuphana Universität Lüneburg

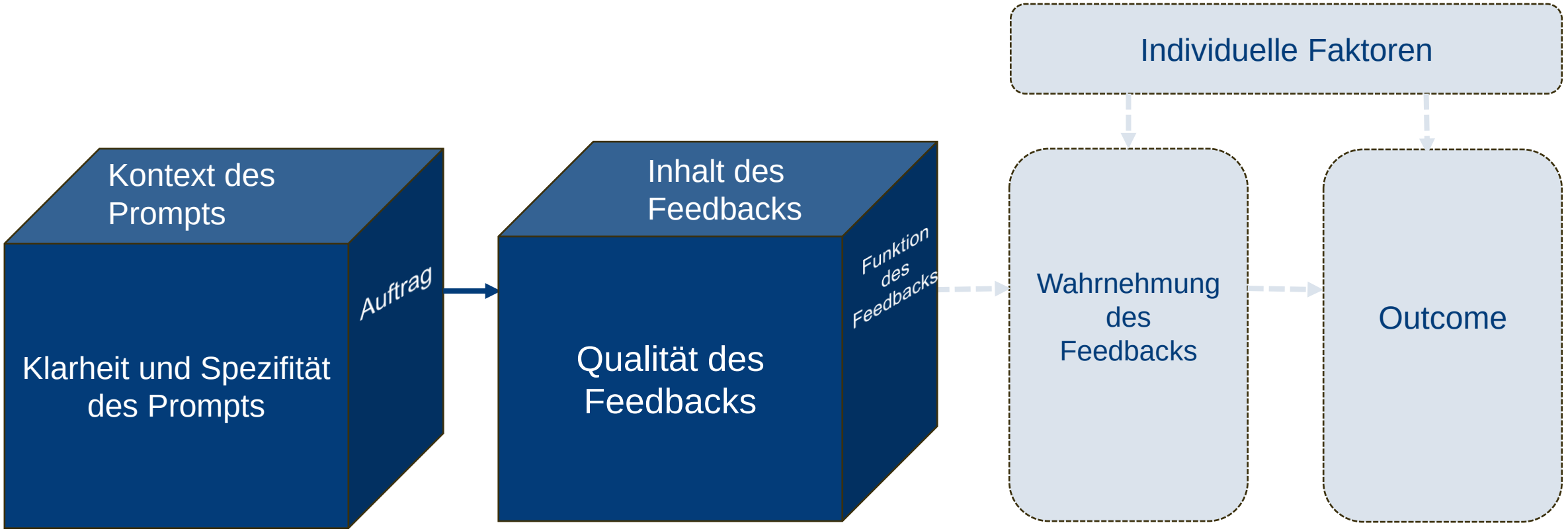
1. THEORETISCHER HINTERGRUND

Relevanz von Feedback

- Feedback gilt als integraler Bestandteil von Bildungsprozessen in der Hochschule (Henderson et al., 2019)
- Feedback in der Hochschule kann von verschiedenen Akteuren gegeben werden (Peers [Novizen]; Expert:innen und KI)
- Expert:innenfeedback ist qualitativ hochwertiger als Noviz:innen Feedback (Prilop et al., 2019) und fördert stärker professionelle Kompetenzen bei Lehramtsstudierenden (Weber et al., 2018; 2020; Prilop et al., 2021)
- ABER: In der Hochschule mangelt es oft an personellen und finanziellen Ressourcen, um qualitativ hochwertiges Expert:innen Feedback zu geben (Demszky et al., 2023)
- KI-Feedback könnte hier eine ökonomische Alternative sein!

1. THEORETISCHER HINTERGRUND

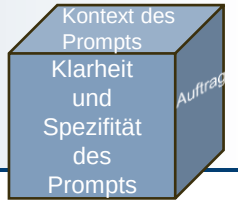
Ein heuristisches Modell zu Ursache und Wirkung von KI-Feedback



Angelehnt an das *Model of factors and effects of external feedback* (Narciss, 2008)

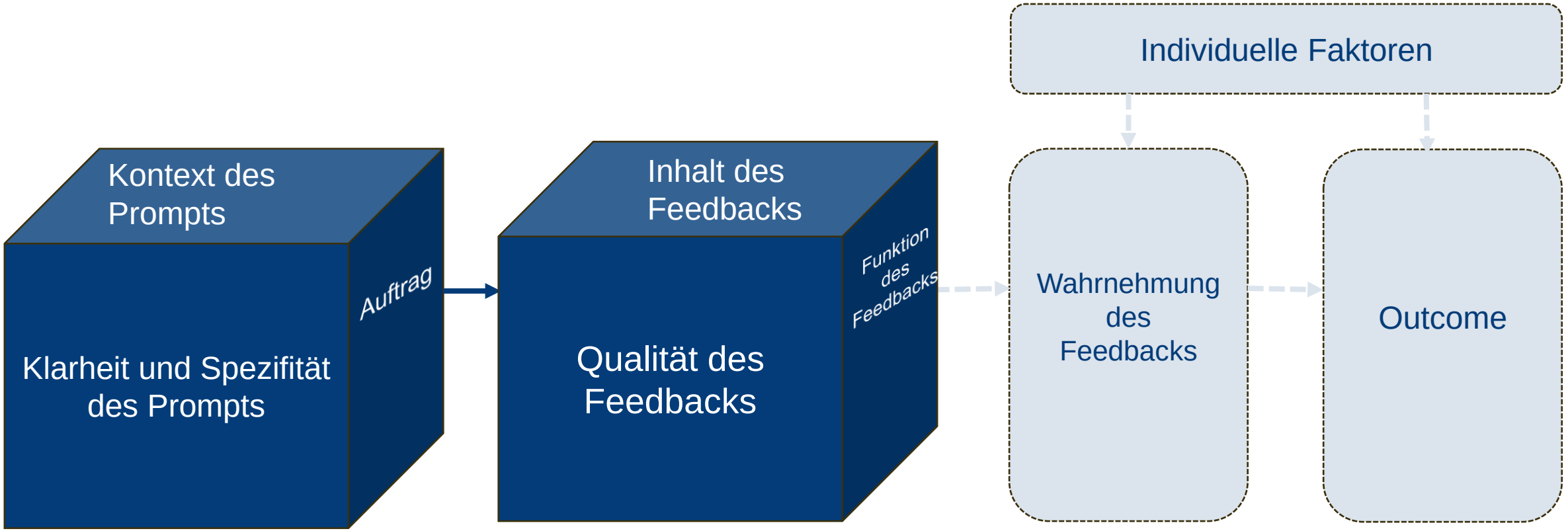
1. THEORETISCHER HINTERGRUND

Prompt Engineering: Kodiermanual für qualitativ hochwertige Prompts



Kategorie	Guter Prompt	Code	Durchschnittlicher Prompt	Code	Suboptimaler Prompt	Code
Rolle	Die Rolle von ChatGPT und von der Frage stellenden Person wird erläutert	2	Nur die Rolle von ChatGPT wird erläutert.	1	Aspekte eines guten Lernziels werden nicht angesprochen	0
Zielgruppe	Es gibt eine klar definierte und beschriebene Zielgruppe.	2	Die Zielgruppe wird grob beschrieben.	1	Die Zielgruppe wird nicht benannt.	0
Kanal	Der Kanal wird klar beschrieben.	2	Der Kanal wird grob beschrieben.	1	Der Kanal wird nicht genannt.	0
Mission/Frage	Auftrag an die KI wird klar und deutlich beschrieben.	2	Auftrag an die KI wird grob beschrieben	1	Auftrag an die KI wird nicht deutlich.	0
Format und Beschränkungen	Es werden sowohl stilistische Eigenschaften als auch Längenangaben beschrieben.	2	Es werden entweder stilistische Eigenschaften beschrieben oder eine Längenangabe gemacht.	1	Es werden weder stilistische Eigenschaften noch Längenangaben gemacht.	0
Prägnanz	Der Prompt beinhaltet nur Informationen, die direkt auf die Ausgabe bezogen und für diese Relevant sind. Er ist klar und prägnant.	2	Der Prompt besteht aus einer prägnanten Eingabeaufforderung mit nur sehr wenig überflüssigen Informationen.	1	Der Prompt beinhaltet viele Informationen, die irrelevant für die Mission/Ausgabe sind.	0
Domänenspezifität	Fachbegriffe werden korrekt genutzt und geben dem LLM die Möglichkeit, in der Antwort auf diese Rekurs zu nehmen.	2	Fachbegriffe werden vereinzelt oder ohne Erklärung genutzt.	1	Es wird kein spezifisches Vokabular, dass für den Fachbereich der Frage relevant ist, genutzt.	0
Logik	Der Prompt besitzt einen sehr guten Lesefluss, innere logische Kohärenz, eine sehr stimmige Abfolge von Informationen und klar verständliche Verbindung von Inhalten und Aufforderungen.	2	Der Prompt erfüllt nur Teile der Bedingungen der Kodierung „2“.	1	Der Prompt ist unlogisch aufgebaut.	0

ChatGPT, & Ekin, S. (2023). Prompt Engineering for ChatGPT: A quick quide to techniques, tips and best practice. doi.org/10.36227/techrxiv.22683919
Kipp, Michael (2023). Wie sag ich’s meiner KI? Hintergründe und Prinzipien zum #Prompting bei #ChatGPT, https://www.youtube.com/watch?v=cfI7q1llkso&t=2382s
Leo S. Lo, The CLEAR path: A framework for enhancing information literacy through prompt engineering, The Journal of Academic Librarianship, 49(4), 2023, 102720, https://doi.org/10.1016/j.acalib.2023.102720.



Angelehnt an das *Model of factors and effects of external feedback* (Narciss, 2008)

Feedback sollte immer *konkret*, *empathisch, aktivierend* und *korrekt* sein

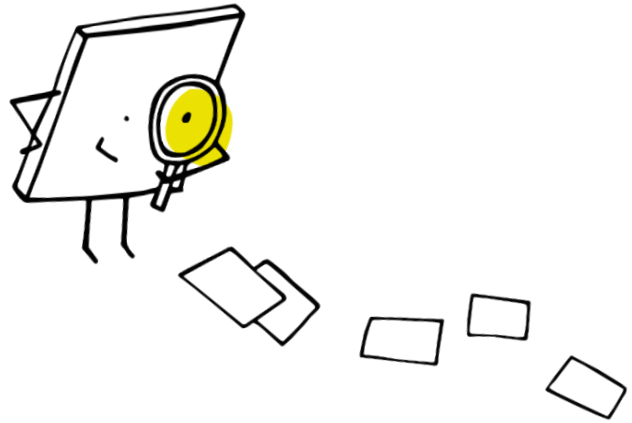
Kriterien guten Feedbacks

- **Konkret** → Benennen und bewerten Sie spezifische Aspekte anhand von Fachbegriffen/ theoretischen Modellen (**Angemessenheit**) und erläutern Sie deren Relevanz (**Erläuterung**).
- **Empathisch** → Senden Sie Ich-Botschaften (**Stil**). Betten Sie kritische Anmerkungen positiv ein (**Valenz**).
- **Aktivierend** → Stellen Sie **Fragen**! Geben Sie Hinweise zur Verbesserung (**Alternativen**) und erläutern Sie die Alternativen (**Erläuterung von Alternativen**).
- **Korrekt** → Benennen Sie verbesserungswürdige Aspekte (**Spezifität**) und achten Sie auf die inhaltliche/ fachliche Korrektheit Ihres Feedbacks (**Fehler**).

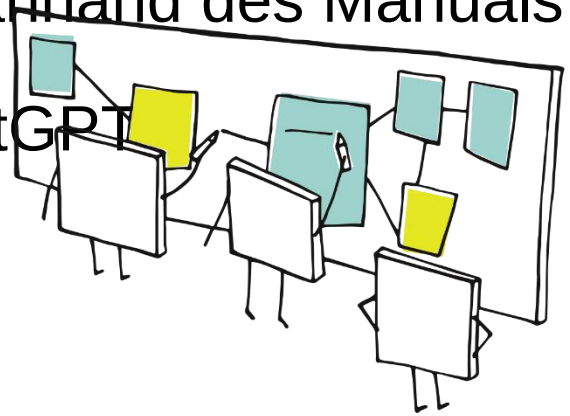
Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2020). Effects of digital video-based feedback environments on pre-service teachers' feedback competence. *Computers in Human Behavior, 102*, 120–131.
Wu, Y., & Schunn, C. D. (2021). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science, 49*(3), 365-394.

2. FRAGESTELLUNG

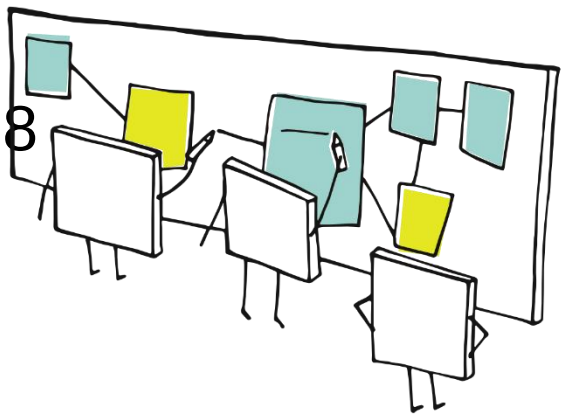
Welche Unterschiede zeigen sich zwischen Noviz:innen-, Expert:innen- und KI-Feedback hinsichtlich der Feedbackqualität?



1. Lernzielformulierung (Wittwer et al. 2020) mit drei Fehlern
   - ➢ Die Schülerinnen und Schüler erkennen ein rechtwinkliges Dreieck und verstehen den Satz des Pythagoras.
2. Theoriegeleitete (ChatGPT & Enkin, 2023; Leo S. Lo, 2023, Michael Kipp, 2023) Entwicklung eines Manuals zur Feststellung von Promptqualität
3. Entwicklung von drei Prompts unterschiedlicher Qualität anhand des Manuals
4. Generierung von Feedback mithilfe der Prompts und ChatGPT



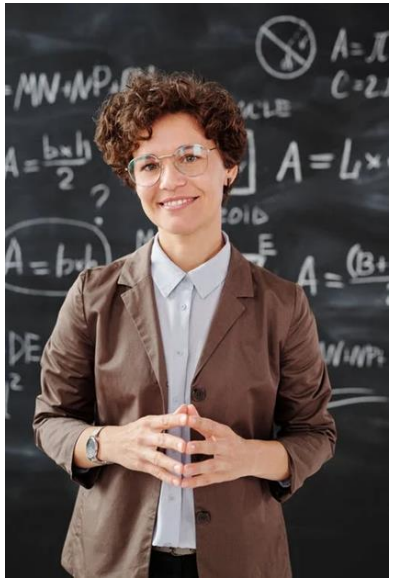
5. Analyse der KI Feedbacks zur Verifizierung der Promptqualität
   - Prompt 1,2 und 3 mit n = 10
6. Bester Prompt wurde Noviz:innen und Expert:innen zum Verfassen von Feedback zur Verfügung gestellt
7. Kodierung aller Feedbacks, Analyse und Vergleich der Feedbackqualität von Noviz:innen, Expert:innen und KI (ChatGPT-4)
   - Prompt 3 n = 30, Noviz:innen n = 30, Expert:innen n = 8



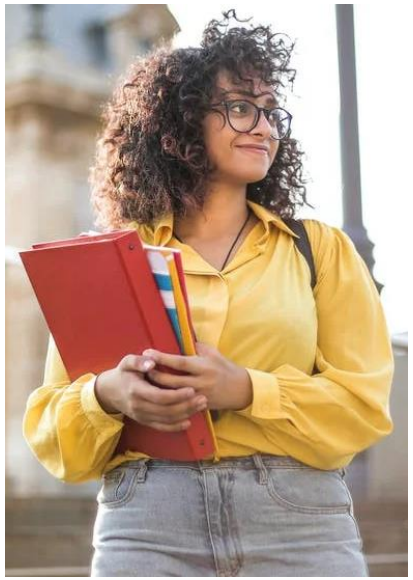
4. ERGEBNISSE

Unterschiede zwischen Noviz:innen und Expert:innen

Expert:innen Feedback **signifikant (p < .001)** besser als **Noviz:innen Feedback**
(siehe auch Prilop et al., 2021; Weber et al., 2019)



>



- Angemessenheit
- Erklärung

→ **Konkret**

- Stil
- Valenz

→ **Empathisch**

- Fragen
- Alternativen
- Erklärung A.

→ **Aktivierend**

- Spezifität
- Fehler

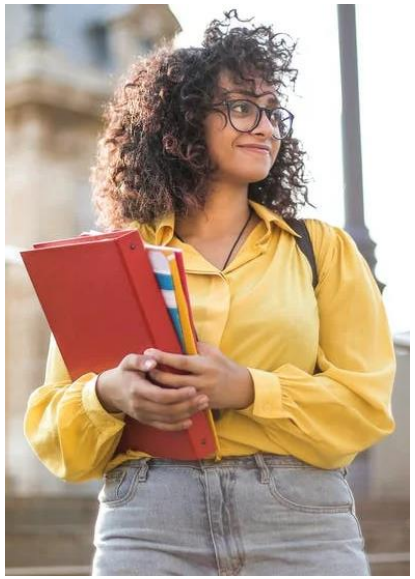
→ **Korrekt**

4. ERGEBNISSE

Unterschiede zwischen Noviz:innen und KI

KI Feedback signifikant (p < .001) besser als **Noviz:innen Feedback**

>



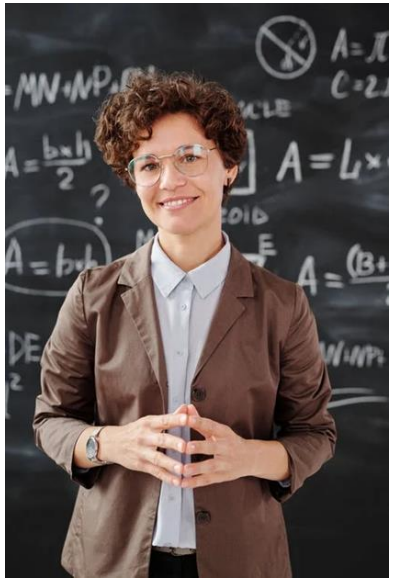
- Angemessenheit
  - Erklärung
  - **Konkret**
- Stil
  - Valenz
  - **Empathisch**
- Fragen
  - Alternativen
  - Erklärung A.
  - **Aktivierend**
- Spezifität
  - Fehler
  - **Korrekt**

4. ERGEBNISSE

Unterschiede zwischen Expert:innen und KI

KI Feedback signifikant ($p < .05$) besser als **Expert:innen Feedback**

>



- Angemessenheit
- Erklärung

→ Konkret

- Stil
- Valenz

→ Empathisch

- Fragen
- Alternativen
- Erklärung A.

→ Aktivierend

- Spezifität
- Fehler

→ Korrekt

- **KI Feedback als Resultat guter Prompts ist zeitökonomischer und hochqualitativer als Expertenfeedback!**

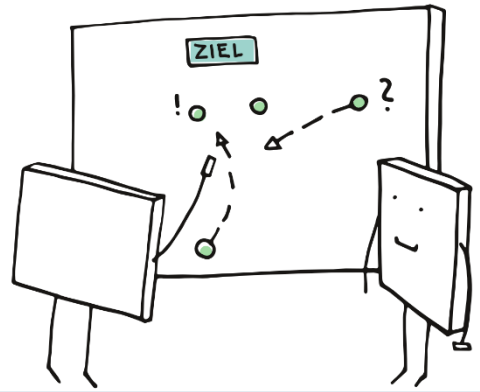
→ Notwendigkeit hochwertiger Prompts

→ Auch beim besten Prompt zeigen sich teilweise Fehler

→ KI-Feedback als Unterstützung; Überprüfung durch Expert:innen notwendig (Azaria et al., 2023)

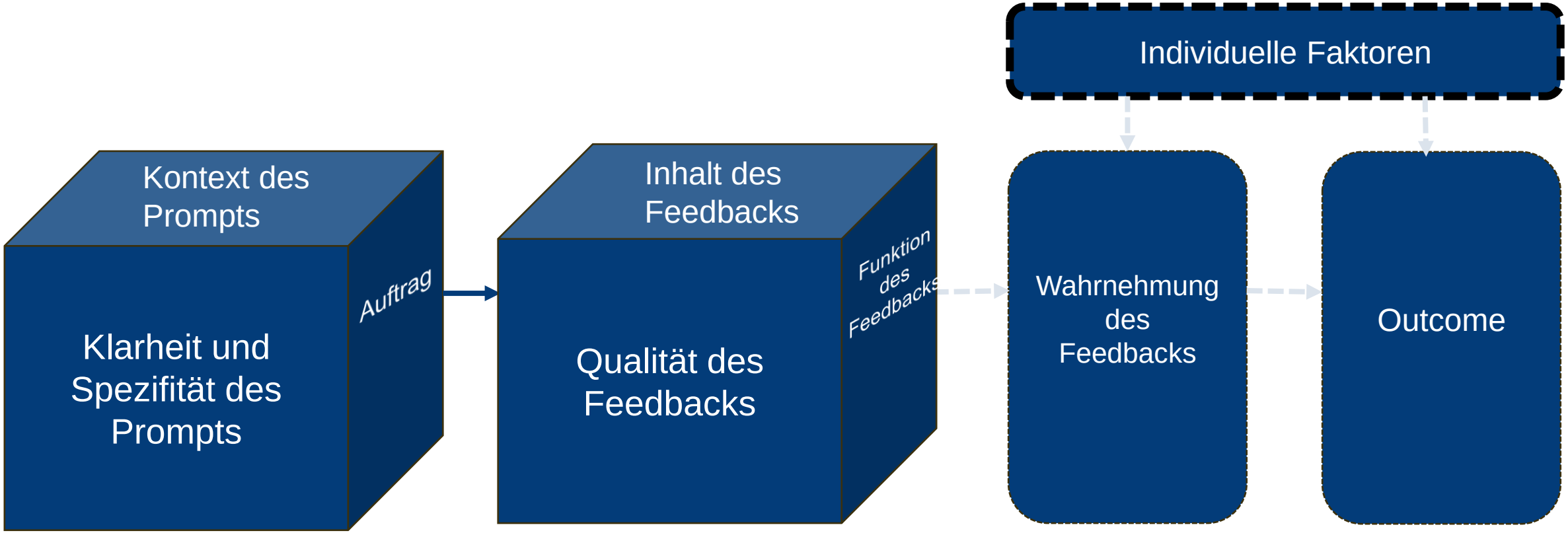
→ Bias durch Trainingsdaten der KI

→ Datenschutz und Ethik



7. AUSBLICK

Wie geht's weiter...



Angelehnt an das *Model of factors and effects of external feedback* (Narciss, 2008)

KONTAKT

Lucas Jasper Jacobsen| Institut für Bildungswissenschaft//
Projekt DigiTaL |
Universitätsallee 1 | 21335 Lüneburg
Fon 04131.677-1629 | lucas.jacobsen@leuphana.de

Jun. Prof. Dr. Kira Elena Weber| Fakultät für
Erziehungswissenschaft// Projekt DigiTaL |
Sedanstraße 19 | 20146 Hamburg
Fon +49 40 42838-2048 | kira.weber@uni-hamburg.de

www.leuphana.de/digital