DIGITALE LERNSPIELE FÜR DEN VERANTWORTLICHEN UMGANG MIT ZUKUNFTSTECHNOLOGIEN

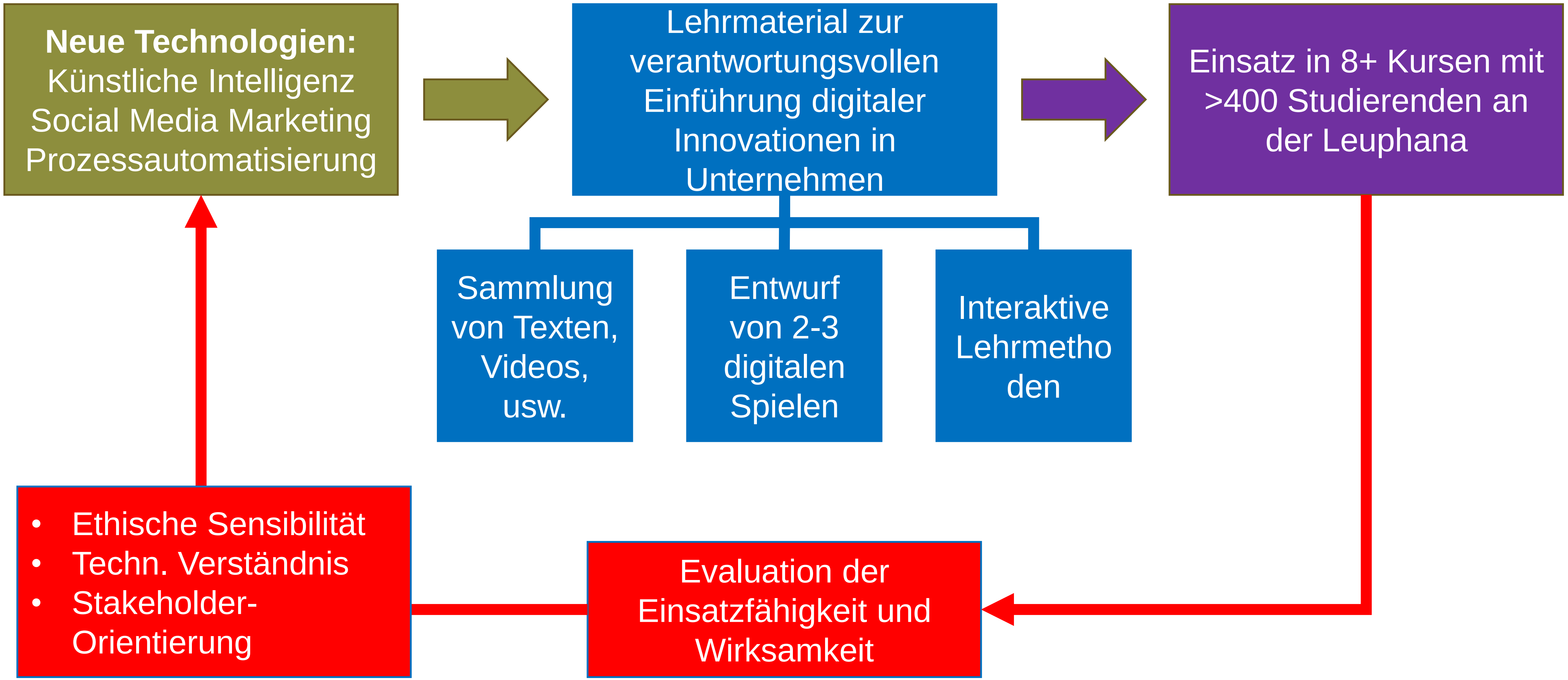
Paul Drews, Hannah Trittin-Ulbrich & Johannes Katsarov

10 Minuten DigiTaL, 9. Januar 2024

→ **A Production of the DigiTaL / DI-SZENARIO Project**

LEUPHANA
UNIVERSITÄT LÜNEBURG

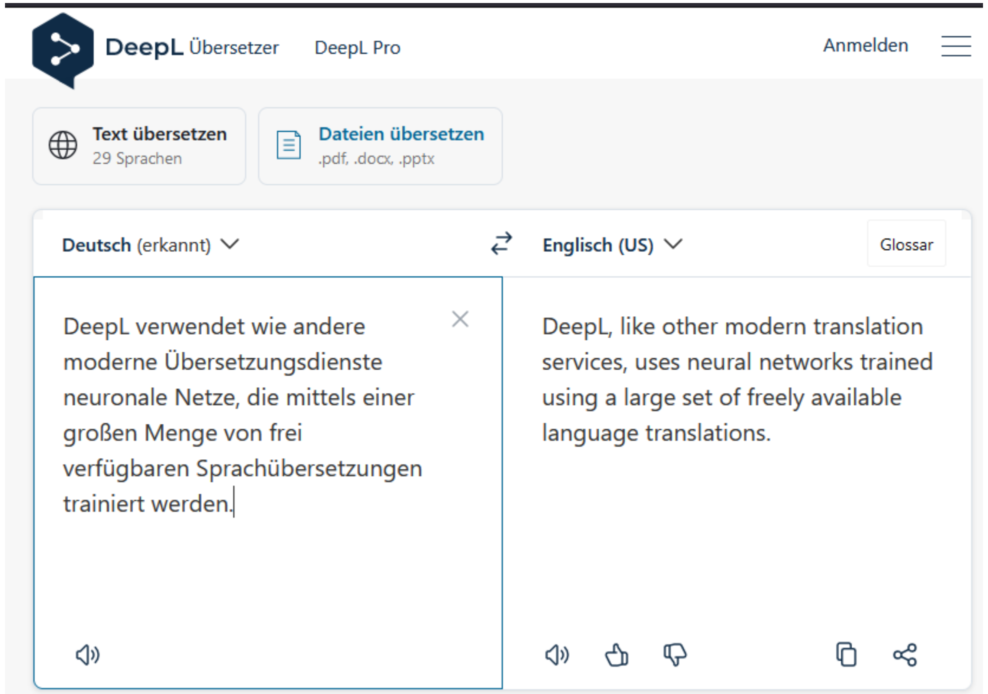
DAS DI-SZENARIO PROJEKT: “RESPONSIBLE ADOPTION OF DIGITAL INNOVATION IN ORGANIZATIONS: A SCENARIO-BASED APPROACH”



MASCHINELLES LERNEN (ML) BIETET VIELE VORTEILE

Generative KI kann originelle Bilder, Texte, Videos und Musik auf Grundlage einer Textbeschreibung erstellen.

Empfehlungen von verschiedenen Diensten (für Produkte, Videos, Routen, Nachrichten usw.) werden auf der Grundlage von ML-Algorithmen erstellt.

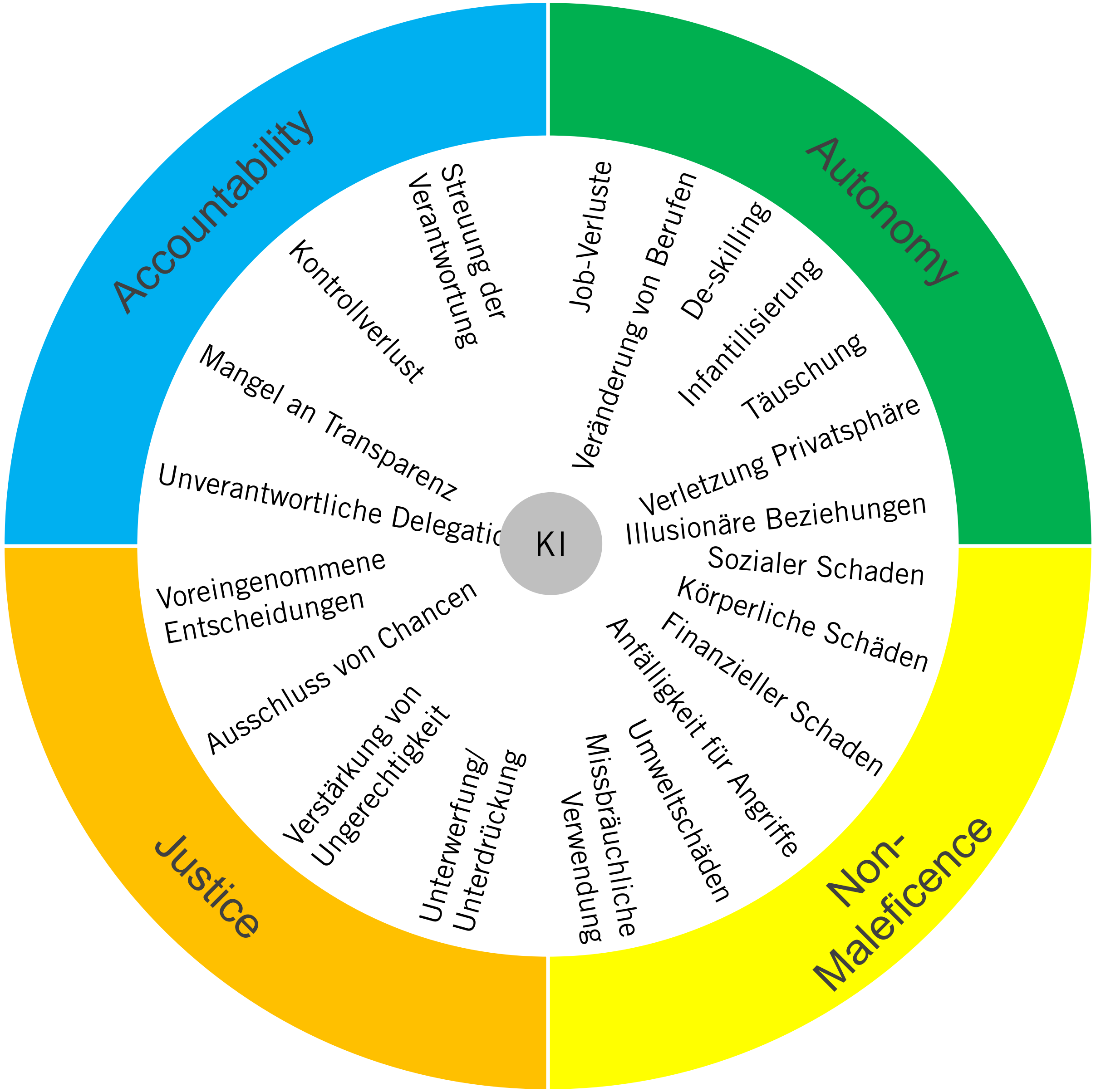


DeepL erstellt sofortige, qualitativ hochwertige Übersetzungen zwischen 29 Sprachen und greift dabei auf eine große Datenbank frei verfügbarer Sprachübersetzungen zurück.

KI: ETHISCHE RISIKEN UND PROBLEME - EINIGE BEISPIELE

Niederländischer Kinderbetreuungsgeld-Skandal Falsche Betrugsvorwürfe durch eine KI, die ~26.000 Eltern zu Unrecht beschuldigt und von vielen eine Rückzahlung von mehr als 10.000 € verlangt

Mann begeht Selbstmord nach "Deal" mit Chatbot
Ein Chatbot hat einem jungen Vater „versprochen“, den Klimawandel aufzuhalten, wenn er sich umbringt



Schlechtere Kreditangebote für gleichgestellte Frauen
In Deutschland verlangen Finanzalgorithmen von Frauen tendenziell höhere Zinssätze und Sicherheiten als von Männern (wenn Frauen überhaupt Angebote erhalten)

YouTube empfiehlt Kindern Selbstverletzungsvideos
Algorithmen haben es versäumt, Videos wie "Meine riesigen extremen Selbstverletzungsnarben" (>400.000 Aufrufe) zu kennzeichnen und Inhalte für 13-Jährige zu fördern

Abbildung 1: Eigene Darstellung auf der Grundlage von O'Neil (1), Floridi et al. (2) und Coeckelbergh (3)

ERLEBNIS-ORIENTIERTES LERNEN

Unterricht zur Wirtschaftsethik basiert am häufigsten auf

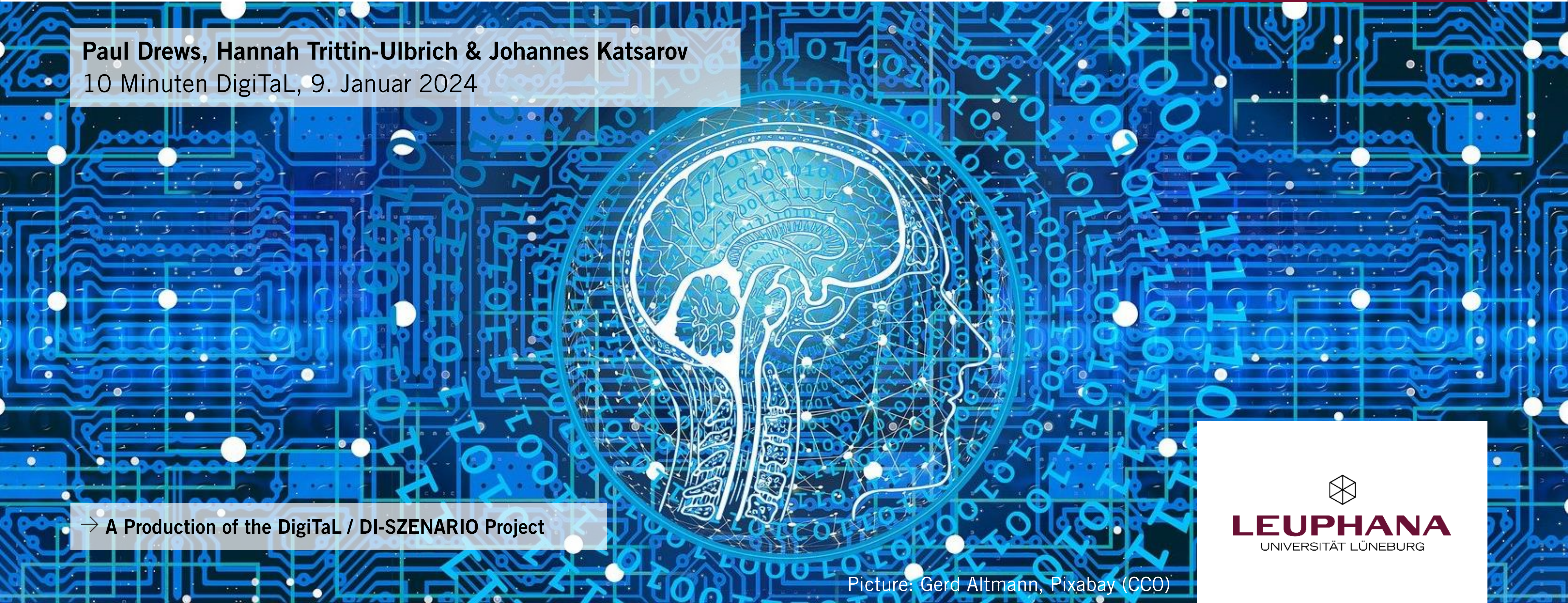
- **Vorlesungen und Lektüre** („theoretischer Ansatz“)
- **Fall-Diskussionen** („deliberativer Ansatz“)[7]

Diese Ansätze sind jedoch deutlich weniger wirksam als

- **Erlebnis-orientiertes Lernen** („immersiver Ansatz“) [8]

MODUL 1: VERANTWORTUNGSVOLLER EINSATZ VON KÜNSTLICHER INTELLIGENZ IN DER WIRTSCHAFT

KONZEPT DES CO-BOLD-SPIELS & ERSTE ERGEBNISSE



Paul Drews, Hannah Trittin-Ulbrich & Johannes Katsarov

10 Minuten DigiTaL, 9. Januar 2024

→ **A Production of the DigiTaL / DI-SZENARIO Project**

Picture: Gerd Altmann, Pixabay (CC0)

LEUPHANA
UNIVERSITÄT LÜNEBURG


0
0
1
Do you get that?
You can count on me, Rachel.
Harriet, I knew you were out for my position...
Blame Marcel
Object

1. ZIELE

LERNZIELE VON CO-BOLD

CO-BOLD wurde entwickelt, um drei Hauptgründe anzugehen, warum Menschen ethische Probleme bei der Nutzung von KI nicht erkennen und angehen können.

1. **Moralische Motivation:** Sicherung der Qualität von 8 Aspekten des AI-Designs (4) + Entwicklung einer professionellen Skepsis (5)
2. **Ethische Sensitivität:** Erkennen von 8 ethischen Risiken und Problemen, die bei KI-Anwendungen häufig auftreten, sowie von relevanten Hinweisen (1-3)
3. **Moralische Entschlossenheit:** Lernen, ethische Probleme auch unter Druck anzusprechen und zu lösen (6)

2. SPIELKONZEPT

ZIELE → MISSION

- 4 Levels
- 2.5 h Spielzeit
- 4 Enden

Um die angestrebten Lernergebnisse zu erreichen, wurden die Geschichte und das Gameplay von CO-BOLD um die drei Lernziele herum aufgebaut:

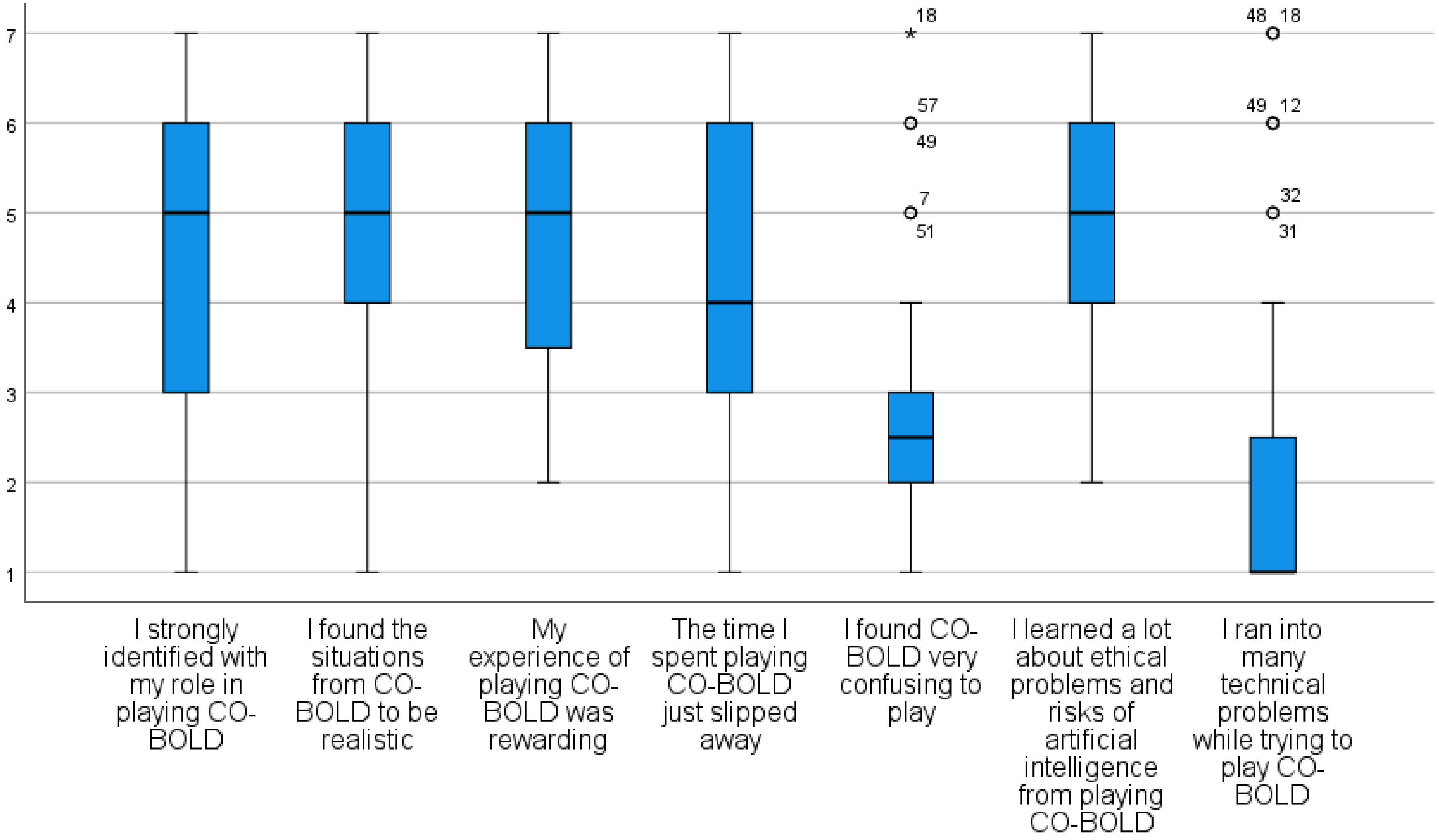
1. **Moralische Motivation:** Sicherung der Qualität von 8 Aspekten des AI-Designs [4] + Entwicklung einer professionellen Skepsis [5]
2. **Ethische Sensitivität:** Erkennen von 8 ethischen Risiken und Problemen, die bei KI-Anwendungen häufig auftreten, sowie von relevanten Hinweisen [1-3]
3. **Moralische Entschlossenheit:** Lernen, ethische Probleme auch unter Druck anzusprechen und zu lösen [6]

→

1. **Rolle der Qualitätssicherung:** Lernende müssen kritisch denken und sorgfältig vorgehen, um das Spiel zu gewinnen.
2. **Untersuchung:** Lernende müssen herausfinden, welche Probleme eine innovative KI-Assistentin hat, um das Spiel zu gewinnen (und ein tragisches Ende zu verhindern).
3. **Notwendigkeit der Integrität:** Lernende müssen andere davon überzeugen, Qualitätsstandards gegen (kurzfristige) Geschäftsinteressen durchzusetzen.

3. ERSTE BEFUNDE

BEWERTUNG VON CO-BOLD



Bewertungen von CO-BOLD durch 48 Management-Studierende auf einer Skala von 1 (stimme überhaupt nicht zu) bis 7 (stimme voll und ganz zu).

VERBESSERUNG DER ETHISCHEN SENSIBILITÄT

Um die Fähigkeit der Menschen zu bewerten, ethische Probleme und Risiken im Zusammenhang mit dem Einsatz von KI zu erkennen, haben wir den **AI Ethical Sensitivity Test (AI-EST)** entwickelt.[9]

1. Die Befragten lesen ein 1-seitiges Interview über eine KI-Fitness-App.
2. Anschließend erläutern sie, welche ethischen Probleme und Risiken sie im Zusammenhang mit der App sehen (schriftliche Aussagen).
3. Die Erkennung von 7 ethischen Problemen wird anhand eines Leitfadens von 0 bis 2 bewertet.

Prä-Test with 28 Management-Studierenden:
- Durchschnitt: 2,8 Punkte (SD = 1,9)
- Die meisten bemerkten >2 Probleme

Kurs: Nach einer kurzen Einführung in KI spielten die Schüler CO-BOLD, gefolgt von einer Nachbesprechung über die ethischen Risiken (4,5 Stunden).

Post-Test (benotete Prüfung) mit 95 Management-Studierenden:
- Durchschnitt: 6,7 Punkte (SD = 2,8)
- Die meisten bemerkten 4 bis 7 Probleme
- Standard-Effektgröße d = 1,7 (groß)

REFERENCES

LEUPHANA
UNIVERSITÄT LÜNEBURG

REFERENCES

(1) O'Neil, C. (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. Crown Books.

(2) Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, E., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*(4), 689-707.

(3) Coeckelbergh, M. (2020). *AI Ethics.* MIT Press.

(4) Tamboli, A. (2019). *Keeping Your AI Under Control, Chapter 4: Evaluating Risks of the AI Solution* (pp. 31-42). Apress.

(5) Hurtt, K. (2010). Development of a Scale to Measure Professional Skepticism. *Auditing: A Journal of Practice &Theory, 29*(1), 149–171.

(6) Gentile, M. (2010). *Giving Voice to Values. How to Speak Your Mind When You Know What's Right*. Yale University Press.

(7) Medeiros, K. E., Watts, L. L., Mulhearn, T. J., Steele, L. M., Mumford, M. D., & Connelly, S (2017). What is Working, What is Not, and What We Need to Know: a Meta-Analytic Review of Business Ethics Instruction. *Journal of Academic Ethics, 15*, 245-275.

(8) Katsarov, J., Andorno, R., Krom, A., & Van den Hoven, M. (2022). Effective Strategies for Research Integrity Training – A Meta-Analysis. *Educational Psychology Review, 34*, 935–955.

(9) Katsarov, J., Blanco Hoppmann, Y., Cramer, I., Drews, P., Rabener, T., Tran Ngoc, L. T. T., Zimmer, M., & Trittin-Ulbrich, H. (2023). *Ethical sensitivity test for business applications of artificial intelligence (AI-EST). Test and scoring guide.* Unpublished manuscript.

CONTACT INFORMATION

JOHANNES KATSAROV
Digital Transformation Lab for Teaching and Learning (DigiTaL)
Institute of Information Systems (IIS)
Institute of Management and Organization (IMO)
Universitätsallee 1 | 21335 Lüneburg
Tel. 04131.677-4028 | johannes.katsarov@leuphana.de
https://www.leuphana.de/institute/iis/personen/johannes-katsarov.html

ETHICAL RISKS & PROBLEMS TO BE DETECTED

Lack of transparency: "Blackbox" character of CO-BOLD's recommendations

Diffusion of responsibility: Who is responsible for the outcomes of bad advice?

Exclusion from opportunity: Migrants are excluded from getting improved advice

Reinforcement of injustice: Existing discrimination is standardized

De-skilling: Risk of investment consultants becoming over-reliant on CO-BOLD

Manipulation: Risk of advising manipulative behaviors to trick clients

Economic harm: Manipulative advice may lead to financial losses

Environmental harm: High energy consumption of machine-learning apps

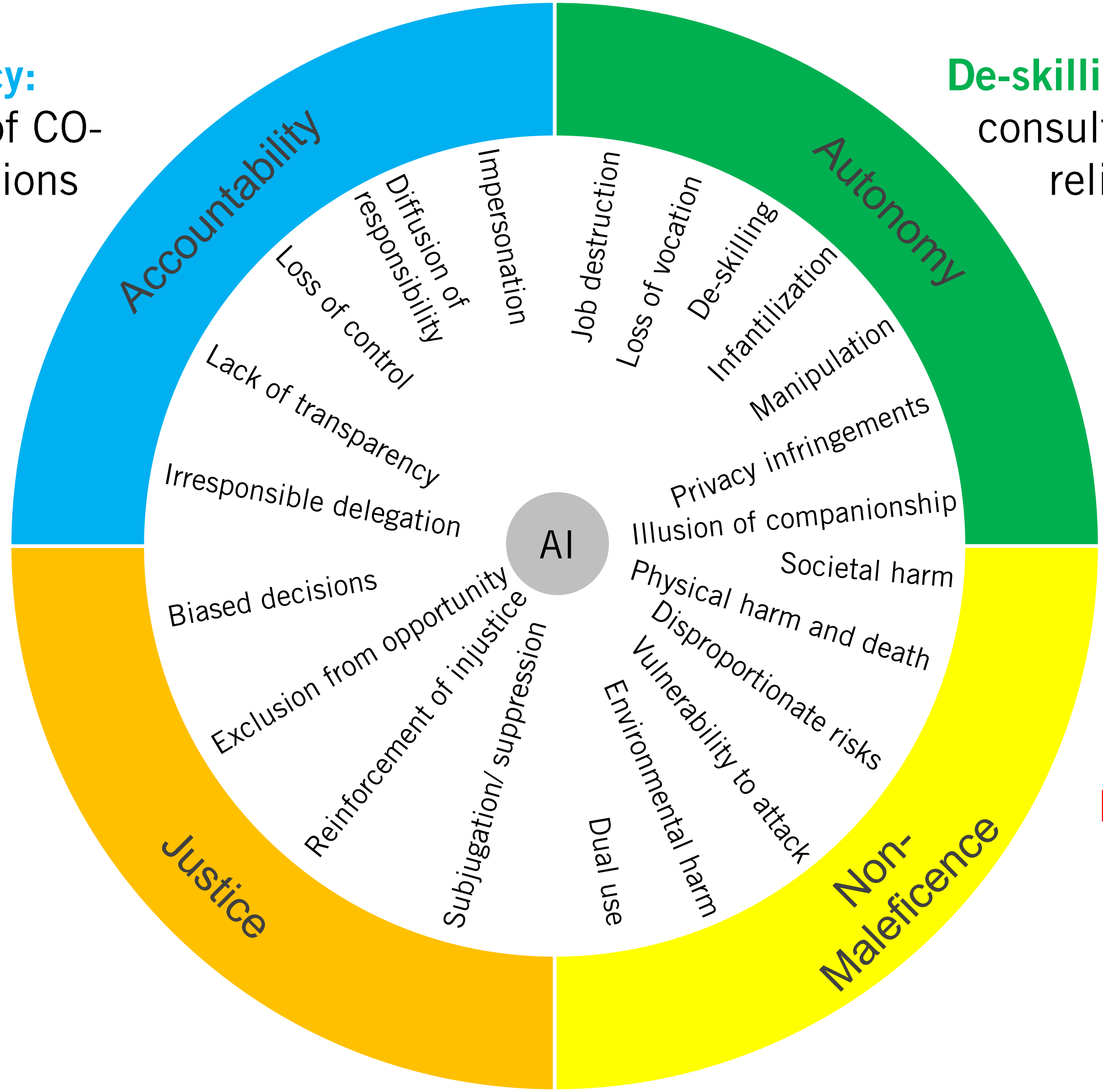


Figure 1: Own illustration based on O'Neil (1), Floridi et al. (2), and Coeckelbergh (3)