

Impulsvortrag:

Automatisierte Beurteilungen von DSH-Textproduktionen durch KI-Programme

Referent: Prof. Dr. Christian Krekeler, Professor für Deutsch als Fremd- und Fachsprache an der Hochschule Konstanz HTWG

Können KI-Programme Textproduktionen verlässlich beurteilen? Im Vortrag werden Möglichkeiten und Grenzen der automatisierten Textbeurteilung mithilfe von KI-Programmen gezeigt.

Studien zur automatisierten Beurteilung mit KI-Programmen von Texten zeigen, dass die Ergebnisse nicht sehr verlässlich sind, was unter anderem an statistischen Effekten wie der Regression zur Mitte liegt. Auch besonders elaborierte Prompts erhöhen die Zuverlässigkeit der Ergebnisse kaum. Dabei ist zu bedenken, dass menschliche Beurteilungen ebenfalls nicht sehr verlässlich sind. Diese Befunde werden auch an exemplarischen Beurteilungen von DSH-Textproduktionen deutlich.

Für die DSH ist eine menschliche Beurteilung weiterhin unverzichtbar. Die automatisierte Beurteilung von Textproduktionen bietet sich als Feedback-Instrument für die Prüfungsvorbereitung an. Zu überlegen ist, ob KI-Programme die Beurteilung durch einen Zweitgutachter ersetzen können.

Literatur:

Almegren, A.; Mahdi, H. S.; Hazaea, A. N.; Ali, J. K.; Almegren, R. M. (2024). Evaluating the quality of AI feedback: A comparative study of AI and human essay grading. *Innovations in Education and Teaching International*, 1–16. DOI: <https://doi.org/10.1080/14703297.2024.2437122>

Kim Haeun et al. (2024) ChatGPT for Writing Evaluation: Examining the Accuracy and Reliability of AI-Generated Scores Compared to Human Raters. In: C. A. Chapelle, G. H. Beckett, J. Ranalli (Hrsg.) *Exploring artificial intelligence in applied linguistics* (73-95). Iowa State University Digital Press. DOI: <https://doi.org/10.31274/isudp.2024.154.06>

Manning, J., Baldwin, J., & Powell, N. (2025). Human versus machine: The effectiveness of ChatGPT in automated essay scoring. *Innovations in Education and Teaching International*, 1–14. DOI: <https://doi.org/10.1080/14703297.2025.2469089>

Mizumoto, A.; Eguchi, M. (2023) Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics* 2 (2). DOI: <https://doi.org/10.1016/j.rmal.2023.100050>

Seßler, K.; Fürstenberg, M.; Bühler, B.; Kasneci, E. (2024) Can AI grade your essays? A comparative analysis of large language models and teacher ratings in multidimensional essay scoring. arXiv. DOI: [doi:10.48550/ARXIV.2411.16337](https://doi.org/10.48550/ARXIV.2411.16337)