Assessment-Vulnerabilität: Overreliance von LLMs in Lernprozessen systematisch erfassen und gestaltungsorientiert begegnen

Tobias Moebert, Evgenia Samoilova
Institut für Informatik und Computational Science, Universität Potsdam

Dieser diskussionsorientierte Beitrag teilt unsere Erfahrungen und Erkenntnisse aus der aktuellen Literatur zur notwendigen Neuausrichtung von Assessment-Praktiken angesichts der Herausforderungen durch die leichte Zugänglichkeit leistungsstarker Large Language Models (LLMs). Ziel ist es, einen direkten Beitrag zur digitalen Transformation der Lehre an den Brandenburger Hochschulen zu leisten.

LLM-Systeme können Prüfungen in vielen Fächern mit hoher Erfolgsquote bestehen [Bo24] und schaffen starke Anreize für ein problematisches Verhalten: *„People often rely too much on AI systems (overreliance), using their suggestions as strong guidelines without seeking and processing more information“* [Va24]. Besonders in Lehrveranstaltungen, wo das Fundament fachlichen Verstehens gelegt wird, erweist sich die oft begrenzte Aufgabenkomplexität als unzureichend, um oberflächliches Copy-Paste-Verhalten zu unterbinden.

Ausgangspunkt unserer Arbeit waren Benchmarking-Studien mit ChatGPT-3.5, das 2023 etwa die Hälfte von 22 Universitätsprüfungen bestand [MSW24]. Neuere Modelle erreichen inzwischen in bestimmten Kursen Erfolgsquoten bis zu 90 %. Offensichtlich lautet die entscheidende Frage nicht länger „Kann die KI die Prüfung bestehen?". Unsere Gespräche mit Lehrenden verschiedener Disziplinen und die Analyse aktueller Forschung verdeutlichen: Das eigentliche Problem liegt in der Frage, **wie die KI-Nutzung das Lernen der Studierenden beeinträchtigt**. Damit rückt die Gestaltung lernförderlicher Mensch-KI-Interaktionen ins Zentrum.

Was bedeutet das konkret für die empirische Erfassung von Assessment-Vulnerabilität? Anstatt Assessments primär darauf auszurichten, LLMs zu "überlisten", sollten wir Assessment-Designs konsequent auf die Förderung menschlicher Lernprozesse ausrichten. Mit dem Publikum möchten wir praxisnah diskutieren, welche Rolle beispielsweise die kognitive Komplexität der Lernaufgaben, die Balance zwischen Prozess- und Produktorientierung, relative Fähigkeitsunterschiede zwischen menschlicher und KI-Performance, aufgabenspezifische Verifikationsanforderungen sowie die angestrebte Rolle der KI im Lernprozess spielen und welche dieser Faktoren in verschiedenen Kontexten besonders relevant sind.

Literaturverzeichnis

[Bo24] Borges, B. et al.: Could ChatGPT get an engineering degree? Evaluating higher education vulnerability to AI assistants. In: *Proceedings of the National Academy of Sciences*, 121 (49), 2024. https://doi.org/10.1073/pnas.2414955121

[MSW24] Moebert, T.; Samoilova, E.; Wiepke, A.: ChatGPT-3.5 im Hochschul-E-Assessment: Prüfungsleistung, akademische Integrität und zukünftige Bewertungsansätze. In: *Informatik Spektrum*, 2024. https://doi.org/10.1007/s00287-024-01575-x

[Va24] Vaccaro, M.; Almaatouq, A.; Malone, T.: When combinations of humans and AI are useful: A systematic review and meta-analysis. In: *Nature Human Behaviour*, 8, 2024, S. 2293–2303. https://doi.org/10.1038/s41562-024-02024-1