Der Einsatz von KI-Detektoren zur Überprüfung von Prüfungsleistungen - Eine Stellungnahme

Von Kira Baresel[1], Janine Horn[2] und Susanne Schorer[2]

[1]Universität Vechta, [2] Carl von Ossietzky Universität Oldenburg

Herausgegeben vom „Digitale Lehre Hub Niedersachsen“ (DLHN)

Lizenziert unter CC BY-SA 4.0.

Stand: 24.02.2025

0. Abstract

Diese Stellungnahme diskutiert den Einsatz von KI-Detektoren an Hochschulen und zeigt die Möglichkeiten und Herausforderungen auf. Aufgrund der Unzuverlässigkeit und rechtlichen Herausforderungen bei der Implementierung wird geschlussfolgert, dass von deren Nutzung im Prüfungskontext abzusehen ist und die Ressourcen der Hochschulen besser in die Anpassung der Prüfungskultur und -formate gesteckt werden sollten.

Schlagwörter: Hochschulen, KI, Künstliche Intelligenz, KI-Detektoren, KI-Nutzung, Prüfungskultur, Prüfungsformate, Prüfungsergebnisse, Täuschungsversuch, Datenschutz, Hochrisiko-KI-Anwendung, Informationspflicht, Black-Box-Problematik, KI-Compliance, automatisierte Entscheidungsfindung, Datensicherheit

1. Einleitung

KI-Detektoren sind mit der Hoffnung verbunden, dass Prüfer*innen unkompliziert und verlässlich angezeigt wird, wann KI in einer schriftlichen Arbeit benutzt wurde. In diesem Fall müssten Hochschulen ihre Prüfungsleistungen mit asynchronem Anteil (wie bspw. Haus- und Abschlussarbeiten) trotz der weiten Verbreitung von generativen KI-Modellen nicht anpassen oder verändern, da unrechtmäßige KI-Nutzung erkannt und entsprechend sanktioniert werden könnte. Diese Hoffnung birgt mehrere Probleme auf verschiedenen Ebenen, von denen einige im Folgenden diskutiert werden.

2. Probleme beim Einsatz von KI-Detektoren

2.1 Zuverlässigkeit

KI-Detektoren sind **nicht zuverlässig**, auch wenn die Prozentangaben, mit denen angegeben wird, mit welcher Wahrscheinlichkeit ein Text KI-generiert ist, dies vortäuschen.[1] Sie liefern u. a. **falsch-positive Ergebnisse**, das heißt das Prüflingen KI-Nutzung unterstellt wird, obwohl diese keine KI verwendet haben (siehe dazu (Weber-Wulff et al., 2023) und (Sheinman Orenstrakh et al., 2023)). Insbesondere gut strukturierte Texte können von dieser falsch-positiv Bewertung betroffen sein.

Die Detektion tatsächlichen Fehlverhaltens hängt zusätzlich noch von der Kombination ab, welches KI-Modell zur Textgenerierung genutzt wird und mit wel-

[1] Der Disclamer bei Quillbot lautet bspw.: „Caution: Our AI Detector is advanced, but no detectors are 100% reliable, no matter what their accuracy scores claim. Never use AI detection alone to make decisions that could impact a person's career or academic standing. “ (https://quillbot.com/ai-content-detector). Nicht alle Detektoren gehen so transparent mit diesem Problem um.

chem Detektor die Prüfung des Textes stattfindet. Bestimmte Detektoren erkennen Texte von bestimmten Modellen besser bzw. schlechter (*KI Generatoren im Vergleich mit 10 KI Erkennern - Der Wettlauf um KI Text Erschaffung vs. Erkennung, in der Post-Truth Welt*). Auch sind KI-Detektoren nicht einseitig nur von der Hochschule nutzbar, auch Studierende (insbesondere wenn diese es auf einen Täuschungsversuch abgesehen haben) können ihre Texte mit KI-Detektoren testen und bei Bedarf anpassen, damit die Wahrscheinlichkeit einer positiven Detektion sinkt. Liu et al. (2024) und Sheinman Oranstrakh et al. (2023) veranschaulichen, wie gravierend die Erkennungsraten sinken, wenn KI-generierte Texte umgeschrieben werden.

Auch der Beitrag von Májovský et al. (2024) und die aktuell schnelle Entwicklung und Veröffentlichung immer neuer Sprachmodelle und deren rasante Weiterentwicklung macht wenig Hoffnung darauf, dass KI-Detektoren langfristig in dem Maße besser werden, dass wir zukünftig auf einen zuverlässigen Einsatz hoffen sollten. Prüfergebnisse von KI-Detektoren werden **unzuverlässig** bleiben.

Hinzu kommt, dass anders als bei Plagiatsscannern **keine Originalquelle** danebengelegt werden kann um einen Verdacht zu beweisen. Auch gerichtlich wurden Ergebnisse von KI-Detektoren deshalb bisher nur als ein Indiz verwendet (VG München – M 3 E 23.4371 und VG München – M 3 E 24.1136) und der Schwerpunkt zur Entscheidungsfindung auf **die Einschätzung und das Statement der Prüfer*innen** gelegt.[2]

[2] Den Anscheinsbeweis einer Täuschungshandlung bejaht das VG München aufgrund der Stellungnahme der Prüfer*innen zu folgenden Punkten:

- Vergleich zu den Leistungen anderer Bewerber*innen
- Vergleich der Textpassagen untereinander in Struktur, Inhalt und Ausdruck
- Typische Fehler, Stärken und Ausdrucksformen von KI-Generatoren
- Allgemeine Erfahrungen der Prüfer*innen zu den Fähigkeiten von Bachelorabsolventen bei der Abfassung von Texten
- Vergleich zu bekannten Vorleistungen des Bewerbers (nur in erster Entscheidung)

2.2 Probleme bei der Implementierung in Hochschulen

Unabhängig davon hätten die Hochschulen bei der Implementierung von KI-Detektoren mindestens die gleichen **Herausforderungen** zu bewältigen wie bei der Verwendung von Plagiatsscannern. Als Stichpunkte sind hier die **Bedenken** bei der Freiwilligkeit der Einwilligung, beim Erklären der Technik, beim Anpassen von Ordnungen, bei der Finanzierung der Technik, beim Sicherstellen der Datenhoheit, der Datenschutzkonformität und der KI-Compliance genannt. Im Folgenden werden die Wichtigsten noch einmal ausgeführt.

Prüfungsleistungen der Studierenden sind **personenbezogene Daten**, vgl. EuGH, Urt. v. 20.12.2017 C 434/16. Der Einsatz von KI-Detektoren ist, ebenso wie bei Plagiatsscannern, ohne Rechtsgrundlage in Form einer vorherigen Einwilligung gemäß Art. 6 Abs. 1 lit. a DSGVO oder eine Regelung in der Prüfungsordnung gemäß §17 Abs. 1 NHG **nicht zulässig**. Eine **Einwilligung hat freiwillig und in informierter Weise** zu erfolgen. Die Freiwilligkeit setzt eine echte oder freie Wahl voraus d. h. die Einwilligung muss verweigert werden können, ohne Nachteile zu erleiden. Es müsste also alternativ immer möglich bleiben, die Prüfungsleistung auch **ohne Überprüfung durch einen KI-Detektor** abzugeben.

Hinzu kommt, dass eine einmal **erteilte Einwilligung** vom Prüfling jederzeit ohne Grund für die bis dahin erfolgte Datenverarbeitung **widerrufen** werden kann, ohne Nachteile befürchten zu müssen. Daraus resultiert eine gewisse **Rechtsunsicherheit,** ob während des Bewertungsverfahrens die Rechtsgrundlage für den Einsatz des KI-Detektors entfällt, sofern dieser nicht unmittelbar nach der Abgabe und Einwilligung eingesetzt wird. Zudem sind die Prüflinge über die Datenverarbeitung durch den KI-Detektor und die darauffolgende Entscheidungsfindung **hinreichend zu informieren**. Dies kann aufgrund der Black-

Box-Problematik der Systeme jedoch schwierig sein. Hinzu kommt, dass KI-Detektoren als ein System zur Erkennung von verbotenem Prüfungsverhalten als **Hochrisiko-KI-Anwendungen** i.S.d. Art. 6 Abs. 2 KI-VO, Anhang III Nr. 3 lit. d einzustufen sein werden, was eine ganze **Reihe von zusätzlichen Anforderungen und Pflichten** nach sich zieht.

Zudem ist es nach Art. 22 DSGVO **verboten, eine ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung** zu treffen, die gegenüber der betroffenen Person rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt. Dies ändert sich erst, wenn eine gesetzliche Grundlage existiert oder die betroffene Person in diesen Aspekt der Verarbeitung eingewilligt hat.[3] Dies müsste auch für den Fall der KI-Detektoren gelten, wenn diese die **prüfungsrechtliche Entscheidung** bezüglich eines unzulässigen KI-Einsatzes und die daran knüpfende **Folge des Nichtbestehens** maßgeblich beeinflussen. Gemäß EWG 56 (KI-VO) sollten KI-Systeme, die in der allgemeinen oder beruflichen Bildung eingesetzt werden, als hochriskante KI-Systeme eingestuft werden, **da sie über den Verlauf der Bildung und des Berufslebens einer Person entscheiden** und daher ihre Fähigkeit beeinträchtigen können, ihren Lebensunterhalt zu sichern. Siehe hierzu auch den Fall der amerikanischen Lehramtsstudentin Moira Olmstedt (Davalos & Yin, 2024).

Der **Informationspflicht** nachzukommen, könnte wie oben bereits angesprochen aufgrund der KI-Systemen zugrundeliegenden **Black-Box-Problematik** schwierig werden. Allerdings geht das VG München – M 3 E 24.1136, Rn 34 scheinbar nicht von einer maßgeblichen Beeinflussung der Entscheidung durch

[3] Nach der jüngsten EuGH-Rechtsprechung wurde das Vorliegen einer „automatisierten Entscheidung im Einzelfall" i.S.v. Art. 22 Abs. 1 DSGVO für den SCHUFA-Score bejaht, vgl. EuGH, Urt. v. 7.12.2023 - C-634/21. Begründet wurde dies damit, dass der auf personenbezogene Daten einer Person gestützte Wahrscheinlichkeitswert maßgeblich für die Entscheidung Dritter sei. Scoring ist auch nach dieser Rechtsprechung weiterhin zulässig, setzt jedoch voraus, dass Betroffene eingewilligt und über die involvierte Logik des Algorithmus informiert werden müssen.

KI-Detektoren aus, wenn die Bewertung einer Täuschungshandlung vorrangig auf den nachfolgenden Stellungnahmen der Prüfer*innen erfolgt und die Ergebnisse des KI-Detektors lediglich als Indiz dient. Eine Begründung liefert das Gericht im vorläufigem Rechtsschutzverfahren allerdings nicht. Die Auslegung des Art. 22 DSGVO im Zusammenhang mit KI-Detektoren ist demnach **nicht abschließend geklärt**.

Der Einsatz von KI-Detektoren kann letztlich, auch aufgrund Art. 6 Abs. 3 lit. d KI-VO, **nur vorbereitende Aufgaben** für eine Bewertung übernehmen, welche die Entscheidung über das Nichtbestehen **nicht wesentlich beeinflusst** und somit kein erhebliches Risiko darstellt. Die **Stellungnahme des*der Prüfer*in** bleibt also selbst beim Einsatz von KI-Detektoren **entscheidend**.

Und um Sicherheit über die Weiterverwendung der Daten der Studierenden zu erhalten (Training des KI-Detektors, Marketing etc.) müssten die Hochschulen entweder mit dem Anbieter des KI-Detektors einen **Auftragsverarbeitungsvertrag** gemäß Art. 29 DSGVO abschließen oder ein System **lokal hosten**.

2.3 Sinnhaftigkeit des Einsatzes von KI-Detektoren

Es stellt sich also die Frage, ob sich der beschriebene Aufwand für die Hochschulen lohnen würde, um am Ende ein **unzuverlässiges Indiz** in der Hand zu halten? Dabei sollte auch berücksichtig werden, dass die Gefahr besteht, dass Prüfer*innen ein **falsches Gefühl der Sicherheit** entwickeln könnten, wenn die Bereitstellung einer solche Technik von Seiten der Hochschule erfolgt. Letztlich existierten Probleme, wie bspw. das Ghostwriting, auch bevor generative KI-Modelle zur Verfügung standen und werden durch diese verstärkt, aber, wie gezeigt, nicht gelöst.

3. Fazit

Die Verfügbarkeit von generative KI-Modelle zeigt auf, welche Probleme es schon länger mit den gängigen Prüfungsformaten gab und gibt. Wir sollten dies zum Anlass nehmen, **unsere Prüfungsformate und unsere Prüfungskulturen** insgesamt auf den **Prüfstand** zu stellen und gleichzeitig unsere **Studierenden dazu befähigen,** generative KI-Modelle **sinnvoll** einzusetzen.

In der Gesamtbetrachtung raten wir vom Einsatz von KI-Detektoren an Hochschulen daher ab. Stattdessen möchten wir die Hochschulen dazu ermutigen, Prüfungskultur und -formate neu zu diskutieren mit dem Ziel, **zukunftsgerichtete kohärente Prüfungsformate** zu entwickeln, die valide die Leistungen von Studierenden erfassen und diese gleichzeitig auf eine postdigitale Welt vorbereiten.

4. Literaturverzeichnis

Davalos, J. and L. Yin (2024). AI Detectors Falsely Accuse Students of Cheating—With Big Consequences. Bloomberg Businessweek. November.

KI Generatoren im Vergleich mit 10 KI Erkennern - Der Wettlauf um KI Text Erschaffung vs. Erkennung, in der Post-Truth Welt. Nextcoder Softwareentwicklungs GmbH. https://n3xtcoder.org/de/ai-text-generators-vs-detectors

Liu, J. Q. J., Hui, K. T. K., Al Zoubi, F., Zhou, Z. Z. X., Samartzis, D., Yu, C. C. H.,...Wong, A. Y. L. (2024). The great detectives: humans versus AI detectors in catching large language model-generated medical writing. *International Journal for Educational Integrity*, *20*(1), 8. https://doi.org/10.1007/s40979-024-00155-6

Májovský, M., Černý, M., Netuka, D., & Mikolov, T. (2024). Perfect detection of computer-generated text faces fundamental challenges. *Cell Reports Physical Science*, *5*(1). https://doi.org/10.1016/j.xcrp.2023.101769

Sheinman Orenstrakh, M., Karnalim, O., Suarez, C. A., & Liut, M. (2023). Detecting LLM-Generated Text in Computing Education: A Comparative Study for ChatGPT Cases.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O.,...Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*(1), 26. https://doi.org/10.1007/s40979-023-00146-z

5. Angaben zur Lizenz

Dieser Text entstand im Rahmen des Projektes DLHN. Die Gruppe „AG Prüfungen“ des Teilprojektes „KI in Studium, Lehre und Prüfungen“ hat sich mit der Thematik der KI-Detektoren auseinandergesetzt und gibt daher diese Empfehlung zur Orientierung heraus.

Zitationsempfehlung:

Baresel, Kira; Horn, Janine & Schorer, Susanne (2025). Der Einsatz von KI-Detektoren zur Überprüfung von Prüfungsleistungen - Eine Stellungnahme. Herausgegeben vom „Digitale Lehre Hub Niedersachsen“.
DOI: https://doi.org/10.57961/fjg9-jr89

Lizenziert unter CC BY-SA 4.0.