

LEARNING NUGGET E-LEARNING & KI 15. SEPTEMBER 2026

KI-Detektoren sind kein Beweis

Warum schon ein Prozent Fehllarm viele ehrliche Lernende trifft, was Studien über KI-Detektoren zeigen und was im Unterricht stattdessen hilft.

Von Rico Eberle, E-Learning-Experte für Weiterbildung und Gemeinderat in Dübendorf

RICO EBERLE · LEARNING NUGGET

99 %

Genauigkeit.

Klingt zuverlässig.

Rechnen wir nach.



Video zum Beitrag

Dieses Learning Nugget gibt es als Video mit Transkript (1:15 Min.). QR-Code scannen oder Link öffnen:

ricoeberle.ch/blog/ki-detektoren-sind-kein-beweis/

Seit Sprachmodelle ganze Aufsätze schreiben, wächst der Wunsch nach einem Werkzeug, das auf Knopfdruck sagt: Das war die KI. Anbieter von KI-Detektoren versprechen genau das, oft mit beeindruckenden Prozentzahlen. Das Problem liegt nicht nur darin, dass diese Zahlen in der Praxis selten halten. Es liegt schon in der Rechnung selbst.

Ein Prozent klingt wenig, bis man nachrechnet

Nimm an, ein Detektor liegt in 99 Prozent der Fälle richtig. Deine Schule lässt 1000 Aufsätze prüfen, und alle sind ehrlich geschrieben. Ein Prozent Fehlalarm heisst dann: Zehn Lernende stehen unter Verdacht, obwohl sie nichts falsch gemacht haben.



Schon bei 99 Prozent Genauigkeit geraten von 1000 ehrlichen Aufsätzen zehn unter Verdacht.

Das Entscheidende an dieser Rechnung: Die Fehlalarme wachsen mit der Menge. Je mehr Texte geprüft werden, desto mehr Unschuldige trifft es, auch wenn der Prozentsatz gleich bleibt. Und jeder dieser Fälle ist ein schwieriges Gespräch, bei dem Vertrauen auf dem Spiel steht. Dabei ist 99 Prozent ein bewusst optimistischer Wert.

Was die Forschung zeigt

In der Praxis schneiden die Werkzeuge deutlich schlechter ab. Eine Forschungsgruppe um Weber-Wulff hat 2023 vierzehn Detektoren getestet. Ihr Fazit: Die Tools sind weder genau noch zuverlässig.

In der Praxis schlechter

14 Tools getestet

"weder genau noch zuverlässig"

Weber-Wulff et al., 2023

Nicht in der Muttersprache geschrieben

besonders oft als KI markiert

Liang et al., 2023

OpenAI

eigenen Detektor eingestellt

OpenAI, 2023



Drei Befunde aus der Forschung zu KI-Detektoren.

Eine zweite Studie von Liang und Kollegen zeigt ein Fairness-Problem. Englische Texte von Menschen, die nicht in ihrer Muttersprache schreiben, werden besonders häufig als KI markiert. Untersucht wurden englische Aufsätze, unter anderem aus dem Sprachtest TOEFL; ob das für Deutsch als Zweitsprache genauso gilt, sagt die Studie nicht. Im Video ist das verkürzt. Die Forschenden vermuten, dass die Detektoren Menschen mit eingeschränkterem sprachlichem Ausdruck unbeabsichtigt benachteiligen. Sie raten ausdrücklich davon ab, solche Werkzeuge in Bewertungs- und Bildungssituationen einzusetzen.

Und selbst OpenAI, das Unternehmen hinter ChatGPT, hat seinen eigenen Detektor im Juli 2023 wieder eingestellt, wegen zu geringer Genauigkeit.

Was stattdessen hilft

Ein Detektor-Ergebnis ist kein Beweis. Es kann höchstens ein Anlass sein, genauer hinzuschauen. Was im Alltag mehr bringt, sind meine drei Empfehlungen aus dem Video:

- 1. Den Entstehungsprozess sichtbar machen.** Entwürfe, Notizen und Zwischenstände gehören zur Abgabe. Wer einen Text selbst entwickelt hat, kann diesen Weg zeigen.
- 2. Das Gespräch über den Text suchen.** Nicht als Verhör, sondern als Teil der Beurteilung: Warum dieser Aufbau? Was war schwierig?
- 3. Aufgaben so stellen, dass eigenes Denken gefragt ist.** Zum Beispiel mit Bezug auf den eigenen Unterricht, eigene Erfahrungen oder eine begründete Entscheidung.

Kein Beweis.

WAS STATTDESSEN HILFT

- 1 Entstehungsprozess sichtbar machen: Entwürfe, Notizen
- 2 Das Gespräch über den Text suchen
- 3 Aufgaben stellen, die eigenes Denken verlangen



Drei Wege, die mehr bringen als ein Detektor-Ergebnis.

Eine Übung fürs Kollegium

Mein Vorschlag für die nächste Teamsitzung: Schätzt gemeinsam, wie viele schriftliche Arbeiten an eurer Schule pro Semester abgegeben werden. Rechnet dann mit einem Prozent Fehlalarm, also dem optimistischen Wert. Die Zahl, die dabei herauskommt, sind die Gespräche mit ehrlichen Lernenden, die ihr führen müsstet. Danach lohnt sich die Frage: Welche Aufgaben in unserem Alltag würden einen Detektor überflüssig machen?

Vertrauen entsteht nicht durch Software, sondern durch gute Aufgaben.

Häufige Fragen

Wie zuverlässig sind KI-Detektoren?

Wenig. Eine Forschungsgruppe um Weber-Wulff testete 2023 vierzehn Detektoren und kam zum Schluss, dass sie weder genau noch zuverlässig sind. OpenAI hat seinen eigenen Detektor im Juli 2023 wegen zu geringer Genauigkeit eingestellt.

Ist das Ergebnis eines KI-Detektors ein Beweis?

Nein, höchstens ein Anlass, genauer hinzuschauen. Selbst bei optimistischen 99 Prozent Genauigkeit stünden von 1000 ehrlich geschriebenen Aufsätzen zehn Lernende zu Unrecht unter Verdacht.

Benachteiligen KI-Detektoren Menschen, die nicht in ihrer Muttersprache schreiben?

Für englische Texte ja: Laut einer Studie von Liang und Kollegen (2023) werden englische Texte von Menschen, die nicht in ihrer Muttersprache schreiben, besonders häufig als KI markiert. Für Deutsch als Zweitsprache hat die Studie das nicht untersucht. Die Forschenden raten davon ab, Detektoren in Bewertungs- und Bildungssituationen einzusetzen.

Quellen

1. Weber-Wulff et al.: Testing of detection tools for AI-generated text, International Journal for Educational Integrity (2023)
<https://link.springer.com/article/10.1007/s40979-023-00146-z>
2. Liang et al.: GPT detectors are biased against non-native English writers, Patterns (2023)
<https://arxiv.org/abs/2304.02819>
3. OpenAI: New AI classifier for indicating AI-written text (2023, eingestellt)
<https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/>



ÜBER DEN AUTOR

Rico Eberle

Rico Eberle ist E-Learning-Experte, Betriebsökonom FH und Gemeinderat in Dübendorf. Er präsidiert den Stiftungsrat der Stiftung Weiterbildungskurse Dübendorf (WBK). In den Learning Nuggets erklärt er Forschung zu Lernen, KI und digitaler Souveränität kurz und mit Quellen.

Kontakt: ricoeberle.ch · LinkedIn: [linkedin.com/in/rico-eberle](https://www.linkedin.com/in/rico-eberle)

LIZENZ

Text, Transkript und Video von Rico Eberle stehen unter der Lizenz CC BY 4.0 (ausgenommen Musik und Soundeffekte des Videos). Du darfst die Inhalte teilen und bearbeiten, auch im Unterricht, wenn du die Quelle nennst.
creativecommons.org/licenses/by/4.0/deed.de

ZITIERVORSCHLAG

«KI-Detektoren sind kein Beweis», Rico Eberle, ricoeberle.ch, 2026, <https://ricoeberle.ch/blog/ki-detektoren-sind-kein-beweis/>, CC BY 4.0