Rico Eberle · LEARNING NUGGETS

LEARNING NUGGET E-LEARNING & KI 15. SEPTEMBER 2026 (AKTUALISIERT 31. OKTOBER 2026)

KI-Detektoren sind kein Beweis

Schon ein Prozent Fehlalarm trifft bei 1000 ehrlichen Aufsätzen zehn Lernende, und in der Praxis sind KI-Detektoren deutlich ungenauer.

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

Video zum Beitrag

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

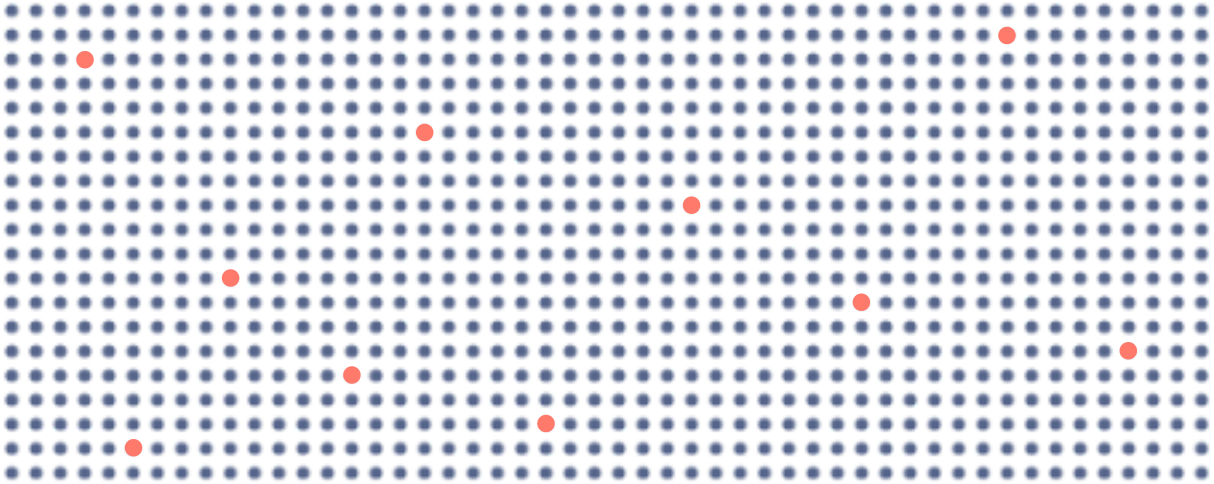
ricoeberle.ch/blog/ki-detektoren-sind-kein-beweis/

99 Prozent Genauigkeit. Das klingt nach einem zuverlässigen KI-Detektor. Der Wunsch nach einem Werkzeug, das auf Knopfdruck sagt «Das war die KI», ist verständlich, und Anbieter werben mit genau solchen Zahlen. Rechnen wir trotzdem kurz nach.

Ein Prozent klingt wenig, bis man nachrechnet

Nehmen wir an, «99 Prozent» heisst hier: Nur ein Prozent der ehrlichen Texte wird fälschlich als KI markiert. Eine Schule lässt 1000 Aufsätze prüfen. Alle sind ehrlich geschrieben. Ein Prozent Fehlalarm heisst: Zehn Lernende stehen unter Verdacht, obwohl sie nichts falsch gemacht haben.

1000 ehrliche Aufsätze



1 % Fehlalarm = 10 falsche Verdächtigungen

Rechenbeispiel mit optimistischem Wert

Die Fehlalarme wachsen mit der Menge. Je mehr Texte geprüft werden, desto mehr Unschuldige trifft es, auch wenn der Prozentsatz gleich bleibt. Jeder Fall ist ein schwieriges Gespräch, in dem Vertrauen auf dem Spiel steht.

Und 99 Prozent ist ein bewusst optimistischer Wert.

In der Praxis sind die Tools schlechter

Eine Forschungsgruppe um Weber-Wulff hat 2023 vierzehn Detektoren getestet. Ihr Fazit: Die Tools sind weder genau noch zuverlässig. Vor allem stuften sie KI-Texte als menschlich ein, übersahen sie also. Ein Detektor kann damit in beide Richtungen falsch liegen: Er verdächtigt Ehrliche und lässt KI-Texte durch.

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

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.

OpenAI, das Unternehmen hinter ChatGPT, hat seinen eigenen Detektor im Juli 2023 wieder eingestellt. Der Grund: zu geringe Genauigkeit.

Ein Anlass zum Hinschauen, kein Beweis

Ein Detektor-Ergebnis kann höchstens ein Anlass sein, genauer hinzuschauen. Im Alltag bringt anderes mehr. Im Video empfehle ich drei Wege:

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.** Kein Verhör, sondern 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.

Die Rechnung fürs eigene Kollegium

Die Rechnung oben lässt sich in jeder Teamsitzung machen. Wie viele schriftliche Arbeiten werden an eurer Schule pro Semester abgegeben? Davon ein Prozent, also der optimistische Wert. Das Ergebnis ist die Zahl der Gespräche mit ehrlichen Lernenden, die ihr führen müsstet.

Danach lohnt sich die Frage, welche Aufgaben in eurem Alltag einen Detektor überflüssig machen würden.

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 einer optimistischen Fehlalarmquote von einem Prozent 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 MICH

Rico Eberle

Ich arbeite im E-Learning, bin Betriebsökonom FH und Gemeinderat in Dübendorf und präsidiere den Stiftungsrat der Stiftung Weiterbildungskurse Dübendorf (WBK). Die Learning Nuggets mache ich allein: Thema, Recherche, Text und Video. Fragen und Widerspruch sind willkommen.

Kontakt: ricoeberle.ch · LinkedIn: 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