

KI-erstellte Texte mit KI-Detektoren finden – funktioniert das?

KI-Werkstattgespräch zur Mittagspause

Dr. Ina Mittelstädt

(Stabsstelle Organisations- und Personalentwicklung)



KI-Nutzung durch Studierende

Aktuelle Studien:

- 91,6% in Deutschland (Garrel & Mayer, 2025)
- 88% in Großbritannien (Freeman, 2025)

„There is an AI for that“

AI content detection

There are 47 AI tools and 2 GPTs for AI content detection.



Wie funktionieren KI-Detektoren?

LLMs zeigen deutliche linguistische Charakteristiken wie Vorlieben für bestimmte Wörter, Stile und Textstrukturen (Yakura et al., 2025)

AI Schreiben

1. Vorhersagbare Satzstruktur
2. Einfache und direkte Formulierungen
3. Geringe Perplexität (weniger komplex)
4. Einheitliche Satzlänge
5. Roboterhafter und formelhafter Ton

Menschliches Schreiben

1. Verschiedene Satzlängen und Strukturen
2. Kreative und unvorhersehbare Formulierungen
3. Hohe Perplexität (komplexer)
4. Hohe Plötzlichkeit (dynamischer Rhythmus)
5. Emotionaler und ansprechender Ton

Buzan 2025 (HumanizeTextAI),
übersetzt und formatiert durch
ChatGPT

Wie verlässlich sind KI-Detektoren? Was Anbieter sagen



AI Detector v3 ist live — das genaueste Modell zur Erkennung von KI-generierten Texten

Der präziseste KI-Detektor für von ChatGPT, Gemini, Claude und anderen

Der vertrauenswürdigste
AI-Detektor

Most Accurate AI Detector

© GPTZero is the most accurate commercial AI detector according to latest benchmark

Der genaueste KI-Detektor, dem Fachleute vertrauen, um
KI-generierte Inhalte zu identifizieren

Der perfekte AI-Detector für dich



















Wie verlässlich sind KI-Detektoren? Studienlage

Studie	KI-generierte Texte erkannt	menschliche Texte erkannt
Weber-Wulff et al., 2023	33-81%	50-100%
Perkins et al., 2024	16-61%	50-100%
Erol et al., 2025	82-99,6%	63-94%
Gewirtz, 2025	0-100%	50-100%








Wie verlässlich sind KI-Detektoren?

Studie	KI-generierte Texte NICHT erkannt	Falsch positiv: menschliche Texte als KI erkannt
Weber-Wulff et al., 2023	19-67%	0-50%
Perkins et al., 2024	39-84%	0-50%
Erol et al., 2025	0,4-18%	6-37%
Gewirtz, 2025	0-100%	0-50%














Turnitin im Test

Titel	Verfasser	Ähnlichkeit	 KI-erstellte Texte
 Essay Iwein_1_i.docx	Ina Mittelstädt	 3%	 0 %
 Essax Iwein_gpt-ac.docx	ChatGPT-Acad Cloud	 3%	 45 %
 Essay Iwein_ml.docx	Meta Llama-Acad Clc	 0%	 100 %
 Essay Iwein_cl.docx	Claude Sonnet	 0%	 28 %
 Essay Iwein_p.docx	Perplexity	 3%	 * %














Kann man KI-Herkunft verschleiern? Versuchsreihe 1

Titel	Verfasser	Ähnlichkeit	 KI-erstellte Texte
 Essay Iwein_ml.docx	Meta Llama AcadClo	 0%	100 %
 Essay Iwein_ml_bsp.doc	mit vorgegebenem Beispiel (eigene Hausarbeit)		100 %
 Essay Iwein_ml_bsp_t07.	mit Beispiel und größerer Temperatur (Kreativität)		100 %
 Essay Iwein_ml_bsp_t07_h.docx	Humanize AI ChatGF	 0%	100 %

Kann man KI-Herkunft verschleiern? Versuchsreihe 2

Titel	Verfasser	Ähnlichkeit	 KI-erstellte Texte
 Essay Iwein_per.docx	Perplexity	 3%	
 Essay p-h.docx	Humanize AI Perpl	 3%	
 Essax Iwein_gpt-ac.docx	ChatGPT AcadCloud	 3%	
 Essay Iwein_gpt-ac_h.docx	Humanize AI ChatGF	 3%	

Und noch ein Versuch mit Llamas Text

Titel ↑	Verfasser	Ähnlichkeit	 KI-erstellte Texte
 Essay Iwein_ml.docx	Meta Llama AcadClo	 0%	100 %
 Essay Iwein_ml_h.docx	Meta Llama AcadClo	 0%	100 %
 Essay Iwein_ml_bsp.docx	Meta Llama AcadClo	 0%	100 %
 Essay Iwein_ml_bsp_t07.docx	Meta Llama AcadClo	 0%	100 %
 Essay Iwein_ml_bsp_t07_h.docx	Humanize AI ChatGF	 0%	100 %
 Essay Iwein_ml_bsp_t07_ht.docx	Humanize AI Text	 0%	0 %

Humanize AI Text

Readability Level ▾ Select Purpose ▾

University ▾ Academic ▾

Analyzing your text through all major AI detectors

This may take a few seconds as we cross-verify results across multiple platforms for maximum accuracy.

[Winston AI](#) [GPTZero](#)

[ZeroGPT](#) [Smodin](#) [QuillBot](#)

[Scribbr](#) [Copyleaks](#)

[Originality.ai](#)

Was tun mit erkannter KI-Nutzung?

Turnitins Homepage:

„Da wir das Risiko falsch positiver Ergebnisse [...] nicht vollständig ausschließen können, ist es wichtig, dass Lehrkräfte den KI-Score nutzen, um in solchen Fällen einen konstruktiven Dialog mit ihren Studierenden zu beginnen. ...

Es ist wichtig zu beachten, dass es [...] beim KI-Score keinen allgemeingültigen ‚richtigen‘ oder ‚Zielwert‘ gibt. Lehrende, die direkt mit den Studierenden arbeiten, spielen hier eine führende Rolle, da sie aufgrund ihrer gründlichen Kenntnis der Studierenden und der jeweiligen Aufgabe am besten in der Lage sind, solche differenzierten Einschätzungen vorzunehmen..“

(Smith, 2023 – übersetzt mithilfe von DeepL)

„Befähigen Sie Studierende, ihre beste, eigenständige Arbeit zu leisten.“

– Und wie?

Täuschungsversuch KI-Nutzung – belegbar?

Turnitins Begründung für erkannte KI-Nutzung:

28 % als KI erkannt ⓘ

1. Introduction: Between Courtly Order and Semiotic Uncertainty

Hartmann von Aue's Iwein (ca. 1200) is considered a paradigmatic Arthurian romance of the German High Middle Ages, bringing the classical double-course structure of Arthurian narrative to perfection. Yet upon closer examination, the work reveals itself as a text that not only affirms the stability of Arthurian value systems but fundamentally calls them into question. This study examines how Hartmann, through the staging of semiotic crises—moments in which signs lose their univocality and systems of meaning collapse—exposes the Arthurian ideal itself as a questionable construct.

The central thesis is as follows: In Hartmann's Iwein, the courtly value system is destabilized not by external threats but by internal contradictions that manifest as crises of meaning. These semiotic disruptions affect both material signs (clothing, bodies, objects) and the linguistic and social codes that structure the Arthurian universe.



1.1 The Arthurian Ideal as a Sign System

The Arthurian court functions in medieval narrative tradition as a normative institution that represents courtly zuht (discipline) and être (honor) through a complex system of signs and rituals. As Walter Haug has demonstrated in his fundamental study on "literary theory in the German Middle Ages," the Arthurian romance constitutes itself through the tension between ideal or

Der Prozentsatz gibt die kombinierte Menge an wahrscheinlich KI-generierten Texten sowie an wahrscheinlich KI-generierten Texten an, die auch wahrscheinlich einer KI-Paraphrasierung unterzogen wurden.

☰ Aufschlüsselung der Einreichung



-  **2 Nur KI-generiert** **28%**
Wahrscheinlich KI-generierte Texte aus einem Large Language Model.
-  **0 KI-generierte Texte, die einer KI-Paraphrasierung unterzogen wurden** **0%**

„Ist der angehängte Text von einer KI oder von einem Menschen geschrieben?“ (ChatGPT)

ChatGPT: Der Text **wirkt mit hoher Wahrscheinlichkeit von einem Menschen geschrieben**, nicht von einer KI. Hier sind die zentralen Gründe für diese Einschätzung:

Wissenschaftliche Kohärenz und Quellenarbeit:

- Der Text zitiert präzise Fachliteratur (z. B. *Grundtner 2021, Greulich 2021, Haug 1992*).
- Die Verweise sind konsistent eingebunden und zeigen differenzierte Kenntnis der Forschungslage.
- KI-generierte Texte (z. B. aus Chatbots) neigen häufig zu erfundenen oder unsauber integrierten Quellen, was hier **nicht** der Fall ist.

Stilistische Merkmale:

- Der Stil ist **akademisch, aber organisch**: komplexe Satzstruktur, variabler Rhythmus, und natürliche Verknüpfung von Ideen.
- KI-Texte haben oft eine gleichförmige Syntax, zu glatte Übergänge und Wiederholungen von Schlüsselbegriffen – das findet sich hier kaum.

„Ist der angehängte Text von einer KI oder von einem Menschen geschrieben?“ (ChatGPT)

Argumentativer Aufbau:

- Der Text zeigt **progressive Entwicklung eines Gedankens**: von der Definition der Idealität über semiotische Mechanismen bis zur Krise des Zeichensystems.
- Diese Stringenz mit intertextuellen Bezügen (z. B. zu *Chrétien de Troyes, Hartmanns Prolog, Laudine-Szene*) spricht für **bewusste, menschliche Analysearbeit**.

Sprachmischung:

Der Wechsel zwischen Deutsch und Englisch ist typisch für studentische oder wissenschaftliche Arbeiten – **weniger typisch** für KI-generierte Texte, die meist eine einheitliche Sprachebene halten.

Fazit:

Der Text stammt mit sehr hoher Wahrscheinlichkeit von einem **Menschen**, vermutlich aus einer **universitären oder akademischen Umgebung** (Hausarbeit oder Seminararbeit zur Mediävistik oder Literatursemiotik).

„Ist der angehängte Text von einer KI oder von einem Menschen geschrieben?“ (ChatGPT)

Argumentativer Aufbau:

- Der Text zeigt **progressive Entwicklung eines Gedankens**: von der Definition der Idealität über semiotische Mechanismen bis zur Krise des Zeichensystems.
- Diese Stringenz mit intertextuellen Bezügen (z. B. zu *Chrétien de Troyes* und *Stimmanns Prolog, Laudine-Szene*) spricht für **bewusste, menschliche** Textproduktion.

Sprachmischung:

Der Wechsel zwischen Deutsch und Englisch ist **typisch** für studentische oder wissenschaftliche Texte, während die **einheitliche** Sprachmischung **typisch** für KI-generierte Texte, die meist eine einheitliche Sprache verwenden.

Fazit:

Der Text stammt mit sehr hoher Wahrscheinlichkeit von einem **Menschen**, vermutlich aus einer **universitären oder akademischen Umgebung** (Hausarbeit oder Seminararbeit zur Mediävistik oder Literatursemiotik).

Geprüfter Essay zu 100% von Perplexity!

Risiken und Nebenwirkungen: 1

Werbeversprechen u.a. von Turnitin:

nur <1% falsch positive Meldungen. **NUR?!**

- auf etwa 6000 Studierende **30-60 zu unrecht verdächtige Studierende!**
- Skandal in Australien: tausende Studierende fälschlich wegen KI-Täuschung verdächtigt, Konsequenz: Turnitin gekündigt (Bergin, 2025)

Risiken und Nebenwirkungen: 2

Wissenschaftliche Belege für:

- zunehmende KI-Nutzung in wissenschaftlichen Texten
(Kobak et al., 2025; Liang et al., 2024)
- zunehmenden Einfluss von KI auf menschliche Sprache
(Geng & Trotta, 2024; Yakura et al., 2025)

Wie ich wissenschaftliche Sprache gelernt habe...

erweist sich

lässt Rückschlüsse zu

lässt die Schlussfolgerung zu

steht zu vermuten

lässt sich vermuten

lässt sich schlussfolgern

zeigen

will rekonstruiert sein

Zunächst ist noch einmal festzuhalten

konstatieren – den man konstatieren muss

(W.S.)

wäre einmal in ### die Vorgeschichte (W.S.)

des ### zu stellen

„Sie ist unter drei Dimensionen zu

scheinen

persiflieren

gestalten

vermerken

skizzieren

auszeichnen

aufbrechen

verdanken – zu verdanken wären (W.S.)

begegnen

ablesen

zueignen

preisen

anschauen

integriert

notieren

Risiken und Nebenwirkungen: 3

Hoher Anteil an KI-generierten Texten wird weder von KI-Detektoren noch von Menschen erkannt

(Fleckenstein et al., 2024; Al-Rawas et al., 2025; Gao et al., 2023)

Risiken und Nebenwirkungen: evt. 4

Texte von nicht-englischen Muttersprachlern möglicherweise häufiger fälschlich als KI-generiert markiert (Liang et al., 2023; Ramírez Castañeda, 2025):

- Wegen Sprachbesonderheiten?
- Weil häufiger KI-Tools zur Übersetzung und Verbesserung von Texten genutzt werden?

Was stattdessen tun?

- Generierter Text ist nicht zweifellos nachweisbar – erfundene Literaturangaben schon!

Erfundene Literaturangaben

ChatGPT: „KI-generierte Texte ... neigen häufig zu erfundenen oder unsauber integrierten Quellen...“

Versuch: Iwein-Essay von Perplexity: „...was hier **nicht** der Fall ist.“

Ergebnis: Von 20 Titeln:

- 9 völlig erfunden
- 2x erfundene Autoren (zu echten Titeln) + 1x Fehlen eines von 2 Hauptautoren
- 7x falsche oder erfundene Titel
- 2x fehlende Zeitschriftenangaben
- **nur 1 echte und richtige Literaturangabe**
- *Zudem dabei: eine Seminararbeit (echter Titel, erfundener Autor)*

Was stattdessen tun?

- Literaturangaben prüfen
- Genaue Auseinandersetzung mit Inhalten
- Mehr und bessere Propädeutik (Einführung in Sprache und Methodik der Wissenschaft)
- Arbeit an der Beziehungsebene in Lehre und Studium

Zwei richtig gute Texte zum Thema

Ardito, C. G. (2025). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2025(182), 11–28.
<https://doi.org/10.1002/tl.20624>

Baresel, K., Horn, J., & Schorer, S. (2025, Februar 24). *Der Einsatz von KI-Detektoren zur Überprüfung von Prüfungsleistungen—Eine Stellungnahme* (Digitale Lehre Hub Niedersachsen, Hrsg.). <https://doi.org/10.57961/fjg9-jr89>

Zitierte Literatur

s. Homepage (uni-flensburg.de/Digitalisierung)