

Authors under Surveillance

*Semiotics of AI-assisted Humans-detection
and Text-humanization*

Lorenzo L. D. Incardona, PhD in Semiotics
United International Business School, Milan

YEREVAN STATE UNIVERSITY
Media Text, Semiotics, and AI: In Search of New Meaning
15 TO 17 OCTOBER 2026

destruction de toute voix, de toute origine

*..all writing is itself this special voice, consisting of **several indiscernible voices** [...] literature is precisely the invention of this voice, to which we cannot assign a specific origin: literature is **that neuter, that composite, that oblique into which every subject escapes, the trap where all identity is lost, beginning with the very identity of the body that writes.***

Roland Barthes, “The Death of the Author,” trans. Richard Howard, *Aspen* 5+6 (1967), a year before the essay appeared in French

In the French text the “special voice” is missing. There, writing is “destruction de toute voix, de toute origine” (Barthes, 1968/1984, p. 61). A passage on the impossibility of assigning an origin has two origins.

Barthes's words found their referent: AI writing

“...consisting of several indiscernible voices”

several

the language models they are based upon are called *large*, because of the number of parameters they store and because they are trained on billions of texts

indiscernible

in general, you cannot trace an output back to a specific source

voices

“group of characteristics displayed by the narrator or poetic ‘speaker’”

The Oxford Dictionary of Literary Terms

because they contain the statistical regularities that characterized the countless texts that were “dismembered” during the training phases

A logo, while waiting for a body

Chatbots are already **the industry's own reply to the cognitive crisis** that LLMs had begun to trigger, first of all among the people who built them. Without a body to give to their models, or while waiting for one, AI firms settle for **a logo, a name, an interface, a look and feel, a house style**: signs aimed at containing what is basically a pool of indeterminable possibilities within the frame of **an individual, recognizable entity**

On chatbots as characters in a role-play, see Shanahan et al., 2023.



More than a signature, humans want to tell whether what left the track was

a bear or a cat.

When we confront an army, the army of AI, proper nouns say less than common ones.



ƒ

The symptoms

*a carousel of different, sometimes
funny, facts, anecdotes and data*

u ƒ q 7

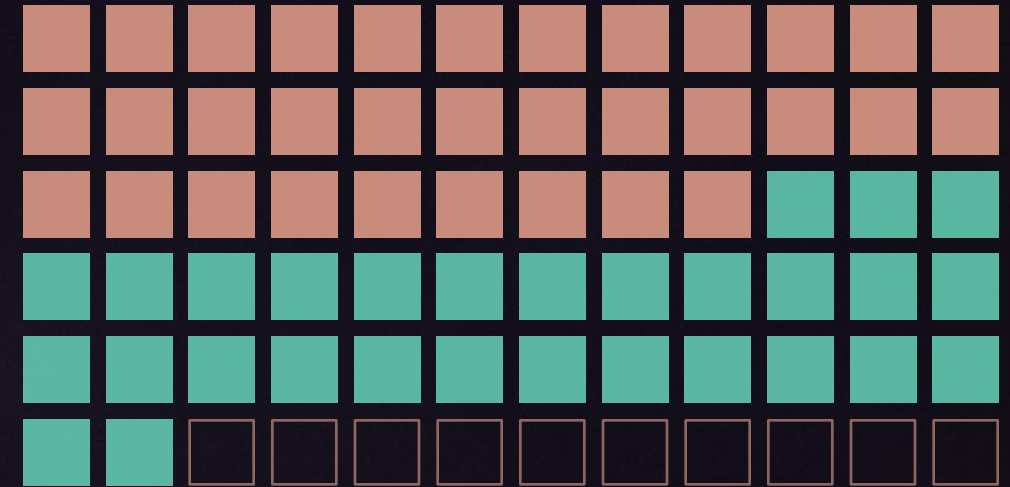
Several dozen AI detectors

The market offers several dozen AI detectors.

AI DETECTOR

an AI-based tool designed to estimate whether a given text was written by a computer

Many also offer to “humanize” AI outputs: they remove the signs that would let detectors spot an artificial author.



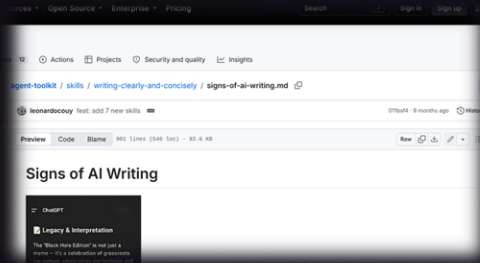
- detection only: 29
- also sells a humanizer: 33
- withdrawn, merged or unreachable: 10

33 of 62 · 53%

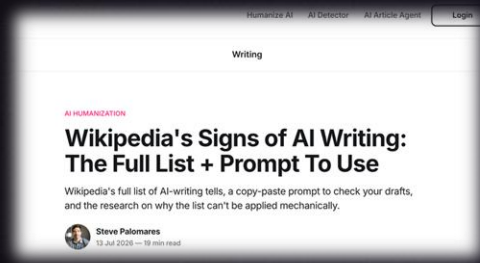
Count made on 9 October 2026 on the 72 entries of a public directory of AI detectors (GitHub, ai-detected/ai-content-detectors): 62 still online; 10 withdrawn, merged into another product or unreachable.

Signs of AI writing

Knowing these signs is therefore precious, marketable knowledge



GitHub repositories



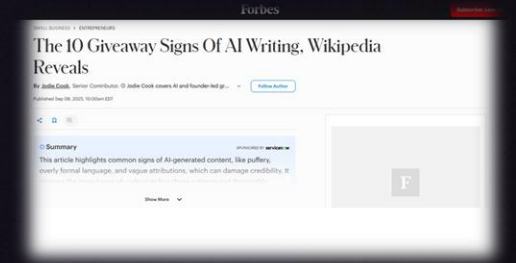
specialized blog posts



*about four edits a day in the last 12 months
1,444 edits between 11 October 2025 and 11 October 2026*

Wikipedia's signs of
AI writing and the
validity of AI
detector results

*Reddit (Individual-War3274,
28 May 2026)*



*articles in major magazines
(e.g., Forbes)*

“Wikipedia:WikiProject AI Cleanup” (2026); “Wikipedia:Signs of AI Writing” (2026). Examples: Softaworks (n.d.), Palomares (2026), Individual-War3274 (2026), Cook (2025).

AI
Your AI Use Is
Breaking My Brain

JASON KOEBLER · MAY 11, 2026 AT 9:57 AM

At least 55 articles on AI writing in seven major outlets in 2026.
Readers complain of a specific cognitive fatigue:

*People think things that are fake are real, things that are real are fake. Much has been written about “AI psychosis,” the [...] nonscientific diagnosis given to people who have lost themselves to AI. Less has been said about the cognitive load of **what other people’s AI use is doing to the rest of us, and the insidious nature of having to navigate an internet and a world where lazy AI has infiltrated everything.** Our brains are now performing untold numbers of calculations per day:*

Is this AI?

Do I care if it’s AI?

Why does this sound or look or read so weird?

Does this person just write like this?

Is this a person at all?

On the professional writers' side

TECHNOLOGY • ARTIFICIAL INTELLIGENCE

Writers Are Going to Extremes to Prove They Didn't Use AI

People are adding typos, aggressively casual language and references to 'The Office' to stay ahead of armchair detectors; 'It's like the new McCarthyism.'

By Te-Ping Chen

May 6, 2026 5:30 am ET

“I’ll use aggressively casual language, like, ‘hey yo, for real,’ or drop a bunch of exclamation points. It feels so icky to do this, but it’s what you have to do to sound human.

[...] It’s like the new McCarthyism

[...] People are demanding proof of something that can’t be proven.”

Sarah Suzuki Harvard, a copywriter, in Te-Ping Chen, “Writers Are Going to Extremes to Prove They Didn’t Use AI,” The Wall Street Journal, 6 May 2026

Rules are multiplying at every level

Such reactions might sound weird or exaggerated, but they are fully understandable.

COUNTRIES

4

jurisdictions, covering
some thirty countries

*China, South Korea, Vietnam,
the European Union*

*now require AI-
generated text to be
labeled in some form*

30

about thirty states in the
United States

*regulate AI in election
communications*

INSTITUTIONS

180

of the world's 200 top-
ranked universities

*publish guidelines on
generative AI*

70%

about 70% of over five
thousand academic
journals

*have an AI policy,
mostly requiring
disclosure*

China: Cyberspace Administration of China et al. (2025). South Korea: AI Basic Act, in force since 22 January 2026. Vietnam: Law on Artificial Intelligence (2025), in force since 1 March 2026. European Union: AI Act, art. 50, from 2 August 2026. United States: thirty-one states (National Conference of State Legislatures, 2026). Universities: Ullah and Kamel Boulos (2026), top 200 of the QS 2026 ranking. Journals: He and Bu (2026), 5,114 journals.

Authors who do use AI

Using a commercial chatbot means accepting its terms of use.

They “control” the author in two ways:

(a) guardrails

prevent the production of content harmful
to humans

*(or, under a rule announced last week, even
cruel behavior toward the models themselves)*



*Anthropic, 2026 Usage Policy update;
the new rule takes effect on 12 November 2026*

(b) stylistic watermarks

statistical biases in word
choice, invisible to readers

Anthropic (since August 2026)

OpenAI (since last week, for now mainly in
the European Union)

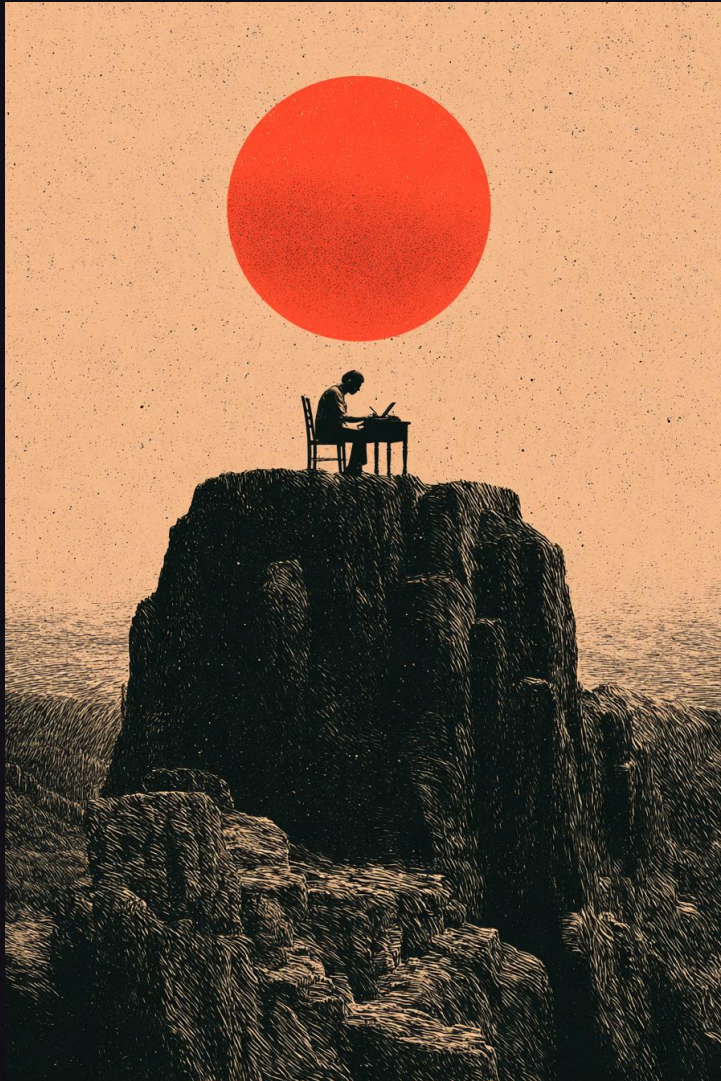


q

The diagnosis

So, what's happening?

u p q 7



Human authors are under surveillance

“enhanced institutional surveillance”

Roe et al. (2026), p. 1

students “navigating an asymmetric surveillance environment”

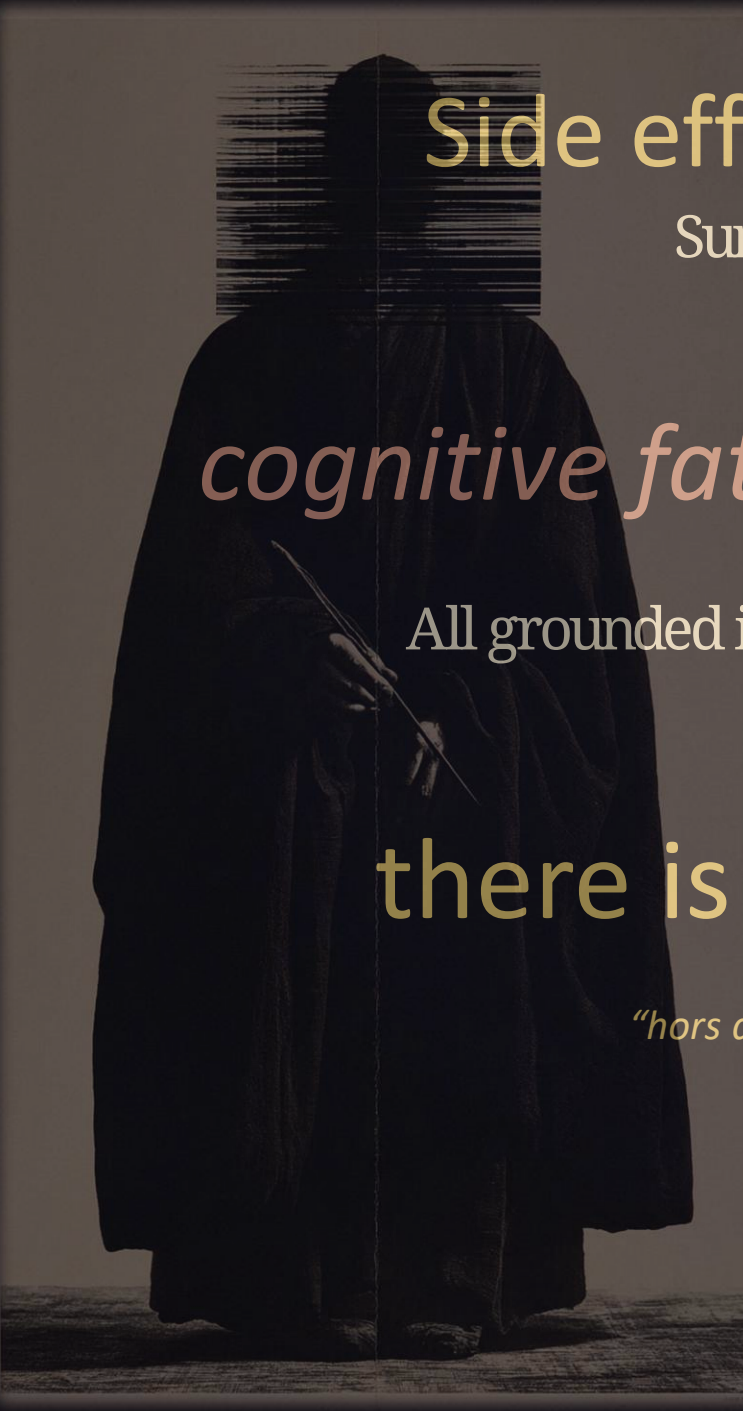
Roe et al. (2026), p. 3

When they rely on generative AI, they open the gates to that oblique where identities are lost. They fail to fulfill their ideological semiotic role: limiting the uncontrolled proliferation of meaning.

...l’auteur n’est pas une source indéfinie de significations qui viendraient combler l’œuvre, l’auteur ne précède pas les œuvres. Il est un certain principe fonctionnel par lequel, dans notre culture, on délimite, on exclut, on sélectionne. Bref, le principe par lequel on entrave la libre circulation, la libre manipulation, la libre composition, décomposition, recomposition de la fiction. [...]

Nous dirons que l’auteur est une production idéologique dans la mesure où nous avons une représentation inversée de sa fonction historique réelle.

The author is therefore the ideological figure by which one marks the manner in which we fear the proliferation of meaning.



Side effects, or forms of “punishment”

Surveillance punishes readers and authors alike:

cognitive fatigue, censorship, ostracism

All grounded in a metaphysical conception of authorship that denies a structuralist principle:

there is no author outside of the text

“hors du texte, point de salut,” as Greimas used to say (Marrone, 2008)

On October 2nd, Damon Crockett (2026), of Yale University, explained why he won't read AI writing:

*[...] there really is no author, **if we understand that concept in the usual way.** A frontier chatbot is something like **all possible authors on all possible topics** (and this even before your bot joins a swarm). Its posttraining has carved some grooves into its latent space, but in most cases, **you can still coax out of it whatever perspective you want to hear**, and it is therefore quite unlike a stubborn human author. And a stubborn human author is a useful thing, because it persists.*

[...]

I don't want an author born at the moment of authorship [...]; I want the author to be a separate thing, behind the text, giving shape.



Rhetorical machines

On one point I agree with Crockett (2026): you can coax out of a model
“whatever perspective you want to hear.”

*“can be used
in order to lie”*

Eco’s definition of semiotics (1976, p. 7)

*“The model does not care
about truth”*

Quattrociocchi, Capraro and Perc (Quattrociocchi et al., 2025,
p. 6)

LLMs are rhetorical machines, semiotic entities.

IN CROCKETT’S VIEW

this disqualifies AI writing

TO ME

*it shows where value comes from: the
significance of those words is attributed by
whoever takes charge of them*



7

The treatment

an old-new readership

u p q 7

Readers have done without authors before

Elémire Zolla recalls Lohengrin, who saves Elsa on condition that his name stays unknown, and must leave when she asks it.

In the Middle Ages, “one went by the effects of a work, without curious, philological, lawyerly interrogations.”

Elémire Zolla, 1974 · Claude Opus 5.5 translation

A culture of “**pervasive anonymity**” that Foucault (1969/1977, p. 138) imagined for the future, and, partly, recognized in the past.

AI detectors ask

Elsa’s question.

Degrees of authorship

Medieval readers had words for degrees of authorship. Bonaventure distinguished four ways of making a book (Minnis, 1988)



scriptor

THE SCRIBE

copies the words
of others



compiler

THE COMPILER

puts together the
words of others



commentator

THE COMMENTATOR

adds his own
words to explain
them



auctor

THE AUTHOR

writes chiefly his
own, adding those
of others as
confirmation

The connection between Bonaventure's degrees of authorship and Barthes's and Foucault's theories was made by Lana (2022).

Simple, maybe trivial considerations

- 1 The values of a text are enacted by its interpretation.
- 2 An author is, first and foremost, a textual strategy (Eco, 1979).
- 3 Philology cannot be reduced to the identification of an empirical author.
- 4 Whoever signs a text answers for it, whatever the origin of its words.
- 5 To author is to interpret, just as to read is to author.

In other words

Against the meaninglessness AI produces, our bulwark is sharper critical reading, applied to every text, whoever or whatever wrote it.

This is my candidate for the “system of constraint” that Foucault left “to be determined”:

readers,
instead of detectors.



*Veritas est in Scripturis sanctis quaerenda, non eloquentia.
[...] Non te offendat auctoritas scribentis, utrum parvae vel
magnae litteraturae fuerit, sed amor purae veritatis te
trahat ad legendum. Non quaeras quis hoc dixerit, sed quid
dicatur attende.*

Imitatio Christi, I, 5

Truth, not eloquence, is what we must seek in Holy Scripture. [...] Do not let the authority of the writer trouble you, whether his learning was small or great, but let the love of pure truth draw you to read. Do not ask who said this, but attend to what is said. (English: Claude Opus 5.5)

References

- ai-detected. (2026). *ai-content-detectors* [GitHub repository]. Accessed 9 October 2026.
- Anthropic. (n.d.). *How Claude marks AI-generated content*. Claude Help Center. Retrieved October 11, 2026.
- Anthropic. (2026, October 8). *2026 Usage Policy update*.
- Barthes, R. (1967). The death of the author (R. Howard, Trans.). *Aspen*, (5+6).
- Barthes, R. (1984). La mort de l'auteur. In *Le bruissement de la langue: Essais critiques IV* (pp. 61–67). Éditions du Seuil. (Original work published 1968)
- Chen, T.-P. (2026, May 6). Writers are going to extremes to prove they didn't use AI. *The Wall Street Journal*.
- Cook, J. (2025, September 8). The 10 giveaway signs of AI writing, Wikipedia reveals. *Forbes*.
- Crockett, D. (2026, October 2). *Why I won't read your AI writing*.
- Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security, & National Radio and Television Administration. (2025). *Rengong zhineng shengcheng hecheng neirong biaoshi banfa* [Measures for labeling artificial intelligence-generated synthetic content].
- Eco, U. (1976). *A theory of semiotics*. Indiana University Press.
- Eco, U. (1979). *Lector in fabula: La cooperazione interpretativa nei testi narrativi*. Bompiani.
- Foucault, M. (1977). What is an author? (D. F. Bouchard & S. Simon, Trans.). In D. F. Bouchard (Ed.), *Language, counter-memory, practice: Selected essays and interviews* (pp. 113–138). Cornell University Press. (Original work published 1969)
- Foucault, M. (1979). What is an author? (J. V. Harari, Trans.). In J. V. Harari (Ed.), *Textual strategies: Perspectives in post-structuralist criticism* (pp. 141–160). Cornell University Press. (Original work published 1969)
- Foucault, M. (1994). Qu'est-ce qu'un auteur? In D. Defert & F. Ewald (Eds.), *Dits et écrits 1954–1988* (Vol. 1, pp. 789–821). Gallimard. (Original work published 1969)
- Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation for Trust, Act No. 20676 (2025) (S. Kor.).
- Greimas, A. J., & Courtés, J. (1979). *Sémiotique: Dictionnaire raisonné de la théorie du langage*. Hachette.

References

- He, Y., & Bu, Y. (2026). Academic journals' AI policies fail to curb the surge in AI-assisted academic writing. *Proceedings of the National Academy of Sciences of the United States of America*, 123(9), Article e2526734123.
- Imitazione di Cristo: Testo latino a fronte* (C. Vitali, Trans.). (1974). Rizzoli.
- Individual-War3274. (2026, May 28). *Wikipedia's signs of AI writing and the validity of AI detector results* [Online forum post]. Reddit.
- Koebler, J. (2026, May 11). Your AI use is breaking my brain. *404 Media*.
- Law on Artificial Intelligence*, Law No. 134/2025/QH15 (2025) (Viet.).
- Lana, M. (2022). Sistemi di intelligenza artificiale e problemi del concetto di autore: Riflessioni su prodotti editoriali recenti [Artificial intelligence systems and problems of the concept of author: Reflections on recent publishing products]. *JLIS.it*, 13(2), 13–44. <https://doi.org/10.36253/jlis.it-458>
- Marrone, G. (2008). L'invention du texte. *Actes Sémiotiques*, (111).
- Minnis, A. J. (1988). *Medieval theory of authorship: Scholastic literary attitudes in the later Middle Ages* (2nd ed.). University of Pennsylvania Press.
- National Conference of State Legislatures. (2026, October 9). *Artificial intelligence (AI) in elections and campaigns*.
- Palomares, S. (2026, September 12). Wikipedia's signs of AI writing: The full list + prompt to use. *HumanizeAI*.
- Quattrociochi, W., Capraro, V., & Perc, M. (2025). *Epistemological fault lines between human and artificial intelligence* (arXiv:2512.19466) [Preprint]. arXiv.
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union*, L 2024/1689.
- Roe, J., Perkins, M., Bannister, P., Furze, L., & Wood, J. (2026). *Dramaturgies of deception: AI humanizers and the performance of legitimacy in higher education assessment* (arXiv:2605.02649) [Preprint]. arXiv.
- Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature*, 623(7987), 493–498.
- Softaworks. (n.d.). *Signs of AI writing* [Markdown file]. GitHub. Retrieved October 11, 2026.
- Ullah, M., & Kamel Boulos, M. N. (2026). Assessing generative artificial intelligence tools usage guidelines at the top 200 universities worldwide. *Future Internet*, 18(9), Article 492.
- Wikipedia:Signs of AI writing. (2026, October 9). In *Wikipedia*.
- Wikipedia:WikiProject AI Cleanup. (2026, October 7). In *Wikipedia*.
- Zolla, E. (1974). Introduzione. In *Imitazione di Cristo: Testo latino a fronte* (C. Vitali, Trans.). Rizzoli.