

People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated text

Jenna Russell Marzena Karpinska Mohit Iyer



Motivation: Are there people who can reliably detect AI? If so, what clues do they use?

Data & Annotations

WHAT?

150 Human-written news articles



150 Paired AI-generated news articles

GPT-4o
Claude-3.5 Sonnet
o1-Pro

Paraphrasing
Humanization

WHO?



5 experts who use LLMs for writing tasks (e.g., editing)



Annotator #4
content writer,
frequently uses ChatGPT

Annotator's Decision



AI-generated

Confidence



Explanation

... Lots of the quotes felt realistic, but many of the quotes did not need a narration alongside it such as with "He shrugged as if that were the most ordinary idea, then laughed." ... could have been shortened to get more facts in about what people in Alaska face and why they face such limited transportation from the rest of the world. Also, it got sentimental and corny at times too.

In Alaska, a pilot drops turkeys to rural homes for Thanksgiving

A half-dozen villagers in Napakiak, on the Kuskokwim River's west bank, gathered near a gravel airstrip last Thursday to watch a small plane circle overhead. ... This crowd was waiting for a seasoned pilot who had a tradition: dropping Thanksgiving turkeys to homes scattered across miles of tundra and frozen waterways.

The pilot, 47-year-old Alaskan flyer Erik Fosnes, has been doing this for nearly a decade, working with volunteers from a regional nonprofit called Delta North Outreach. "We tried shipping turkeys one year by cargo, but half never made it in time," said Fosnes, running a hand through the frost on his jacket sleeve after landing. "So I said, 'What if I just fly them in myself?'" He shrugged as if that were the most ordinary idea, then laughed. "Folks around here have gotten used to it."

Looks human-written

Looks AI-generated

Results

	GPT-4o	Claude-3.5 Sonnet	GPT-4o Paraphrased	o1-Pro	o1-Pro Humanized	Overall
Expert Majority Vote	100 (0)	100 (0)	100 (0)	96.7 (0)	100 (0)	99.3 (0)
Pangram	100 (0)	100 (3.3)	100 (0)	100 (0)	96.7 (10)	99.3 (2.7)
GPTZero	100 (0)	96.7 (0)	100 (0)	76.7 (0)	46.7 (3.3)	85.3 (0.7)
Fast-DetectGPT	100 (0)	96.7 (3.3)	56.7 (3.3)	86.7 (0)	23.3 (3.3)	80 (2)
Binoculars	100 (0)	93.3 (0)	60 (6.7)	73.3 (0)	6.67 (0)	66.7 (1.3)
GPT-4o + COT + Detection Guide	100 (10)	100 (13.3)	100 (16.7)	86.7 (6.7)	3.3 (3.3)	78 (10.7)

Clues mentioned for correctly labeled articles

	Vocabulary	Sentence Structure	Grammar & Punctuation	Originality	Quotes	Clarity	Formatting	Conclusions	Formality	Names	Tone	Introductions	Factuality	Topics	Other
GPT-4o	69.8	46.8	11.5	19.4	18.7	11.5	14.4	26.6	3.6	18.0	15.8	10.8	1.4	3.6	1.4
Claude	78.0	35.5	11.3	12.1	16.3	17.0	7.1	14.9	3.5	17.0	5.7	7.1	4.3	0.0	0.7
GPT-4o paraphrased	88.0	36.6	10.6	21.8	33.8	17.6	14.8	28.2	0.7	7.0	7.0	7.7	7.0	2.1	0.7
o1-Pro	57.1	48.7	3.4	12.6	12.6	19.3	4.2	12.6	5.0	14.3	3.4	10.1	5.0	1.7	1.7
o1-Pro humanized	42.3	39.6	2.7	10.8	14.4	23.4	8.1	12.6	7.2	18.0	12.6	6.3	6.3	0.0	1.8
Human	34.5	25.1	34.3	28.1	23.3	20.8	15.7	6.9	13.3	7.3	8.6	4.2	9.3	2.3	2.7

What clues do experts use to detect AI-generated text?

Vocab

HUMAN: "Furthermore, I very much doubt AI would have used adventurous adjectives like 'chunky', 'musky' or 'thin' to describe food. Nor would it have used verbs like 'blitzing' or 'bolstering'."

AI (O1-HUMANIZED): "Odd word choices: wheat that 'stores' a lineage; genes that are 'honed.'"

Sentence Structure

HUMAN: "Short choppy sentences and paragraphs."

AI (O1-Pro): "One pattern I've been noticing with AI, and I think I've stated this before, is the comparison of 'it's not just this, it's this' and I'm seeing it here, along with listings of specifically three ideas."

Grammar

HUMAN: "There's a lot of variety in the article's grammar use, with dashes, brackets, quotes intermixed with sentences, and short spurts of comma sections throughout."

AI (GPT-4o PARA.): "there's nothing off about the grammar or syntax in this piece..."

Originality

HUMAN: "it's offset by some great analogies and creative phrasing that works well to convey the topic, such as with 'amateur sleuths', 'catnip for a certain type of Reddit user.'"

AI (O1-Pro): "What happens when AI tries to be creative? Penguins 'stand on their own flippers'."

Quotes

HUMAN: "The quotes being short snippets also makes me think they're real, as the writer had to find a way to fit them into the text, rather than them just perfectly stating either side's views."

AI (GPT-4o): "The quotes also feel fake, every expert speaks the same way and it's too homogenous with the text."

Clarity

HUMAN: "Words like 'meander' are used, but are used sparingly to create better flow of ideas, and its writing style is simplified in the best way possible."

AI (CLAUDE-3.5-SONNET): "The sentences are condensed to provide the best possible precision with its word choice, but the article lacks flow and clarity."

Topics

HUMAN: "Phrases throughout the article, including 'burned to ashes and scattered at sea to prevent the crowd from venerating them as relics' and 'to trample on a brass likeness of Jesus or the Virgin Mary—a blasphemous act.' actively show what happened, the horrors and tragedies that occurred during that time."

AI (GPT-4o): "It spends very little time talking about the horrors of the disease, and instead focuses on future research, hopeful quotes, and potential cures, even referring to it as 'embarking on a new chapter'."

Factuality

HUMAN: "The article is very factual, with details about the dates, materials used, and sales portions."

AI (GPT-4o-Para): "Incorrect information delivered with customary AI confidence: the 'quad axel' is not another name for the backflip."



What are your favorite AI tells?

This time I went to the end of the piece and said: "Hello, AI." There it was in all its glory: the "testament" serving "as a beacon of hope and inspiration" and "demonstrating" to us humans "that anything is possible." Sometimes I feel sorry for AI—it must have a dreary time trying to satisfy its human interlocutor's desire to "showcase" advocacy, social change, inclusivity, gender equality, equity, and representation in an essay of fewer than a thousand words.

— ANNOTATOR 5



READ MORE



READ about AI in NEWS