ChatGPT Signed It: A Census of the `utm_source=chatgpt.com` Link Tag in Court Filings, SEC Filings, Scientific Papers, Government Documents and Wikipedia

Fırat Mıhcı, HumanizeMy.ai (a Mıhcı Studios company)

Independent researcher (computational linguistics), HumanizeMy.ai Computational Linguistics Lab. Correspondence: hello@humanizemy.ai

Preprint, November 2026. Prepared for public dissemination via ResearchGate.

Abstract

Every attempt to measure how much of the public record now passes through a chatbot runs into the same wall: detectors are probabilistic, and their verdicts can be argued with. This paper uses a trace that cannot. When ChatGPT hands a user a link, it appends the tracking tag `utm_source=chatgpt.com` to the address. A writer who pastes that link into a document without cleaning it leaves a literal, machine-written signature. We searched six open archives for that exact string on 12 November 2026 and confirmed each hit by reading the document. The tag appears in **298 U.S. federal court cases** across 71 courts, in **133 filings with the U.S. Securities and Exchange Commission** from 67 companies, in **169 peer-reviewed papers** in 110 journals, in **15 U.S. government documents** (Federal Register rules, congressional hearings and reports, court opinions), in **3 UK government publications**, and in **11,702 Wikipedia articles** across 36 language editions. No document in any archive carries the tag before December 2024. In science the rate rose from 0.85 to 15.95 tagged papers per 100,000 full-text papers between the last quarter of 2024 and the third quarter of 2026, about 19 times. Half of the SEC filings are documents used to raise money from investors, and one in five tagged links in those filings points to a market-research vendor. The tag proves only that a link came from a ChatGPT answer, not that any text was machine-written or wrong; it is a floor, not an estimate, because a single keystroke removes it.

Keywords: ChatGPT · provenance · tracking parameters · citations · court filings · SEC filings · scientific publishing · Wikipedia · computational linguistics

1. Introduction

How much of what courts, regulators, scientists and encyclopedias publish has passed through a chatbot? The question matters because these are the records that other people rely on without checking: a judge reads the complaint, an investor reads the prospectus, a doctor reads the paper. The usual way to answer it is to run a detector over text and count the flags. That gives an estimate with an error bar, and every flagged author can reasonably say the detector is wrong.

There is a second kind of evidence that needs no detector. Since late 2024, when ChatGPT gives a user a link to a web page, the address it shows ends with `utm_source=chatgpt.com`. The tag exists so that the receiving website can see, in its analytics, that the visitor came from ChatGPT. It has a side effect its designers probably did not plan: if the user copies the link from the answer and pastes

it into a document, the tag goes with it. The document then contains a string that no human types by hand and that only one product produces.

This is the textual equivalent of a fingerprint on a door handle. It does not tell us what the person did in the room. It does not say the document was written by a machine, or that the cited source is wrong, or that anyone broke a rule. It says one narrow thing with certainty: this link reached the author through a ChatGPT answer.

Editors on Wikipedia noticed the tag early and list it among the signs of chatbot-assisted editing. A legal commentator has pointed to a handful of court opinions and agency publications that carry it (Roberts, 2026). What has been missing is a count: how many documents, in which parts of the public record, since when, and pointing where.

This paper asks: **how many documents in the open public record carry ChatGPT's own link tag, and what do they have in common?** We search six archives for the exact string, verify hits by reading the documents, and report counts, dates, document types and the destinations of the tagged links. Section 2 places the work, Sections 3 and 4 describe the archives and the procedure, Section 5 gives the results, and Sections 6 and 7 say what the count does and does not mean.

2. Related work

Estimating chatbot use from vocabulary. The largest studies of machine-assisted writing in science count words that language models over-use. Kobak et al. (2025) estimated from excess vocabulary that at least 13.5% of 2024 biomedical abstracts were processed with a language model, and Liang et al. (2024) used a population-level word-frequency model to track the rise across fields. These methods estimate a share for a population and cannot point to a single document.

Detectors and their limits. Document-level detectors can point to a document, but they err. Liang et al. (2023) showed that several detectors flagged most essays by non-native English writers as machine-written. Any count built on a detector inherits that dispute.

Fabricated citations. A separate literature follows invented references, from the sanctions in *Mata v. Avianca* (2023) to running lists of court decisions that deal with hallucinated case law. Those cases concern sources that do not exist. The tag studied here marks sources that do exist and were reached through a chatbot.

The tag itself. Wikipedia's volunteer guidance on recognising chatbot-assisted edits lists the `utm_source=chatgpt.com` parameter as a sign. Roberts (2026) described it for lawyers and gave individual examples from court opinions, an agency publication and congressional correspondence. Marketing analytics writers discuss the tag from the receiving side, as a way to count visits from ChatGPT. To our knowledge no study has counted the tag across archives of the public record or analysed where the tagged links lead. That count is the contribution of this paper.

3. Data

We used six archives that are open, offer full-text search over the body of documents, and return results that anyone can re-run.

1. **Europe PMC** (European Bioinformatics Institute): 12,506,278 full-text research articles at the time of reading, mostly peer-reviewed open-access papers in the life sciences, with a small share from other fields.
2. **SEC EDGAR full-text search**: filings made with the U.S. Securities and Exchange Commission from 2001 onward, including annual reports, registration statements, offering circulars and proxy materials.
3. **RECAP Archive (CourtListener, Free Law Project)**: documents filed in U.S. federal courts that have been contributed to the public archive, searchable by text.
4. **govinfo** (U.S. Government Publishing Office): the Federal Register, congressional hearings and reports, published federal court opinions and other government publications.
5. **GOV.UK search**: publications of UK government departments and agencies.
6. **Wikipedia**: the live article text of 36 language editions, searched through the `insource:` operator, which reads the wiki source including link addresses.

All searches and downloads were made on 12 November 2026. The archives differ in what share of their domain they hold. Europe PMC holds open-access papers, not the paywalled literature. RECAP holds the federal filings that users have contributed, a fraction of all filings, and no state courts. The counts are therefore counts within each archive, not totals for science or for the courts.

4. Methods

4.1 Search

In each archive we searched for the exact phrase `"utm_source=chatgpt.com"`. For comparison we ran the same search in Europe PMC and EDGAR for nine other values of the same parameter: `google`, `twitter`, `linkedin`, `newsletter`, `openai`, `perplexity`, `copilot.com`, `gemini` and `claude`.

4.2 Verification

Search engines split text into tokens, so a phrase hit is not proof that the exact string is present. For every hit where the archive offers the document itself, we downloaded it and counted occurrences of the string with a regular expression: the full-text XML for each Europe PMC record, the filed document for each EDGAR hit (text extracted from PDF where needed), the HTML or PDF for each govinfo package, and each HTML attachment for the GOV.UK publications. Because links are often broken across lines in typeset documents, the check for PDFs and government text removed white space before matching. A document counts as **verified** only if the string was found in the downloaded file. For RECAP, where bulk download of the filings is not open, we recorded whether the highlighted search snippet shows the full tag. Wikipedia counts come from an operator that matches the wiki source literally and need no second step.

4.3 What we recorded

For each verified document we recorded its date, its type (journal and publisher; SEC form type and filer; court and nature of suit; government collection) and every distinct tagged link with the domain it points to. Domains were sorted into four groups by fixed rules written before looking at totals: government and intergovernmental bodies, scholarly publishers and repositories, market-research

vendors, and everything else. For papers we computed a rate per 100,000 full-text papers by quarter of first publication, using the archive's own count of full-text papers as the denominator.

4.4 What would have refuted the idea

The premise is that the tag enters documents by copy and paste from ChatGPT. Three results would have undermined it: tagged documents dated well before ChatGPT began linking to the web; similar counts for tags of products that do not append tags; or hits that vanish on verification. We report all three checks.

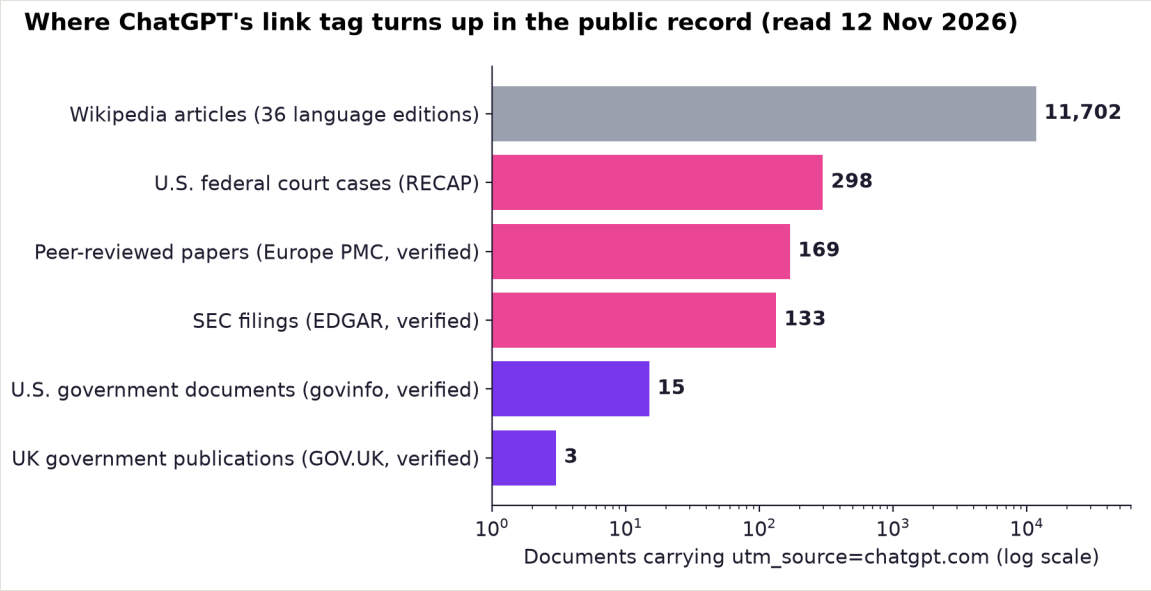
4.5 Reporting rule

We report counts by category. We do not name filers, parties, authors or lawyers. A tag shows how a link was found. It is not evidence of misconduct, and this paper does not treat it as such.

5. Results

5.1 The count

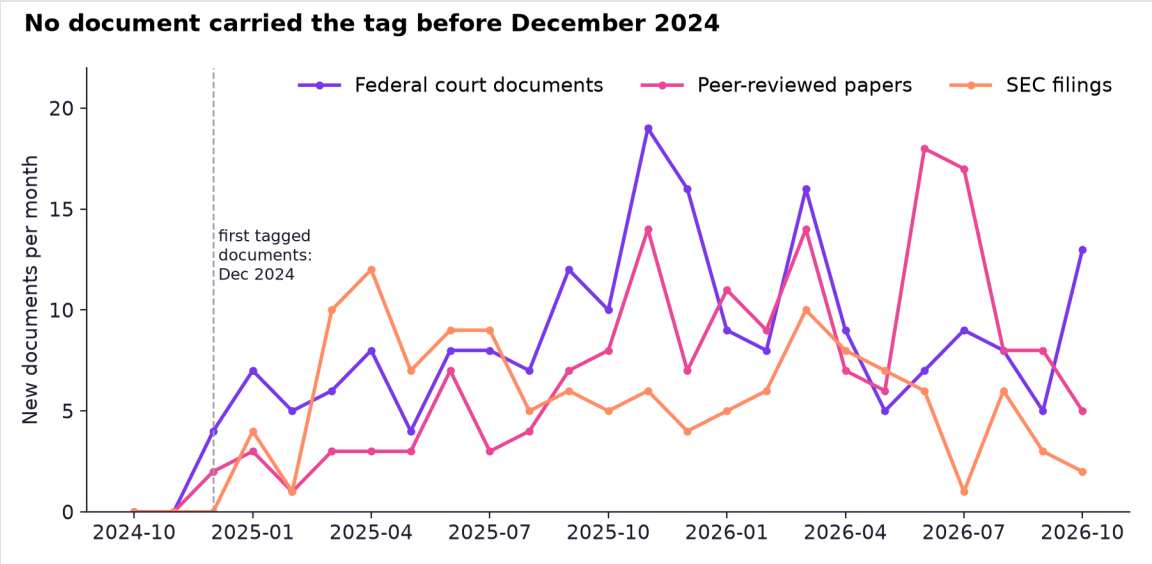
Archive	Search hits	Verified in the document	Spread
U.S. federal court cases (RECAP)	298	295 show the tag in the snippet	71 courts
SEC filings (EDGAR)	135	133	67 companies
Peer-reviewed papers (Europe PMC)	185	169	110 journals
U.S. government documents (govinfo)	21	15	5 collections
UK government publications (GOV.UK)	3	3	3 public bodies
Wikipedia articles	11,702	literal match	36 language editions



Verification removed few hits: 133 of 135 SEC filings and 169 of 185 papers contain the exact string. The 16 unconfirmed papers and 6 unconfirmed government documents are cases where the downloadable file differs from the indexed text or where the link is broken in a way our matcher does not rejoin; we leave them out.

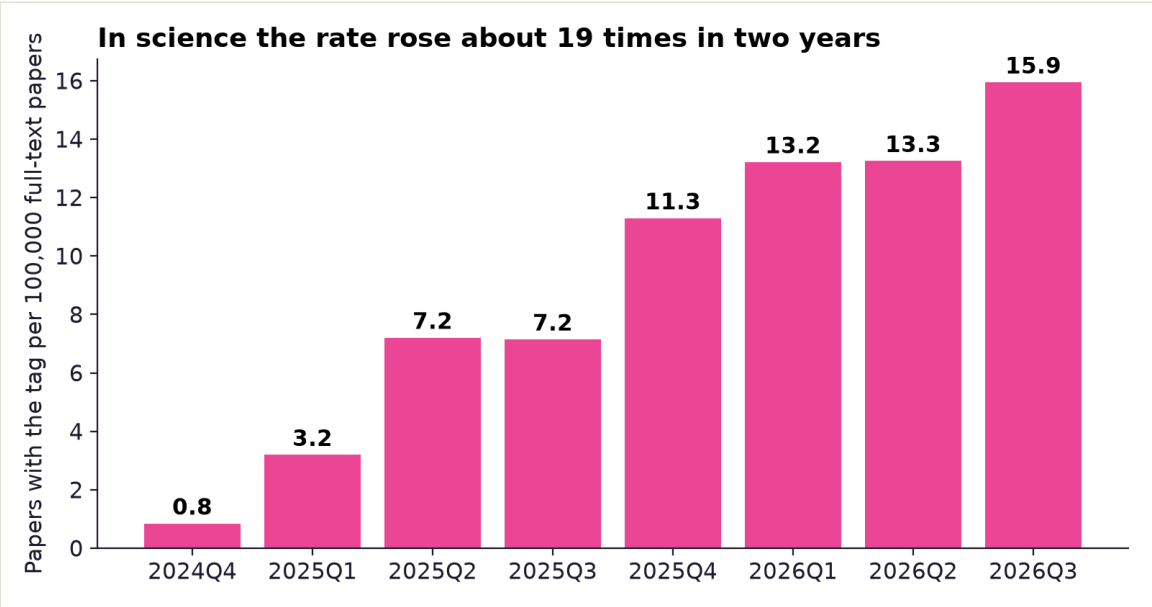
5.2 Nothing before December 2024

The earliest verified paper was published on 13 December 2024. The earliest tagged court documents were entered in December 2024. The earliest SEC filing is dated 28 January 2025. No archive returned a verified document before December 2024, although every archive reaches back many years. A separate search of a technology discussion forum found its first mention of the tag on 13 December 2024. The dates agree with each other and with the period when ChatGPT began to answer with live web links.



5.3 Science: about 19 times in two years

Papers with the tag rose from 2 in the last quarter of 2024 to 35 in the third quarter of 2026. Per 100,000 full-text papers the rate went 0.85, 3.20, 7.21, 7.15, 11.30, 13.22, 13.26 and 15.95 across the eight quarters. The 2026 rate is about one paper in 6,300.



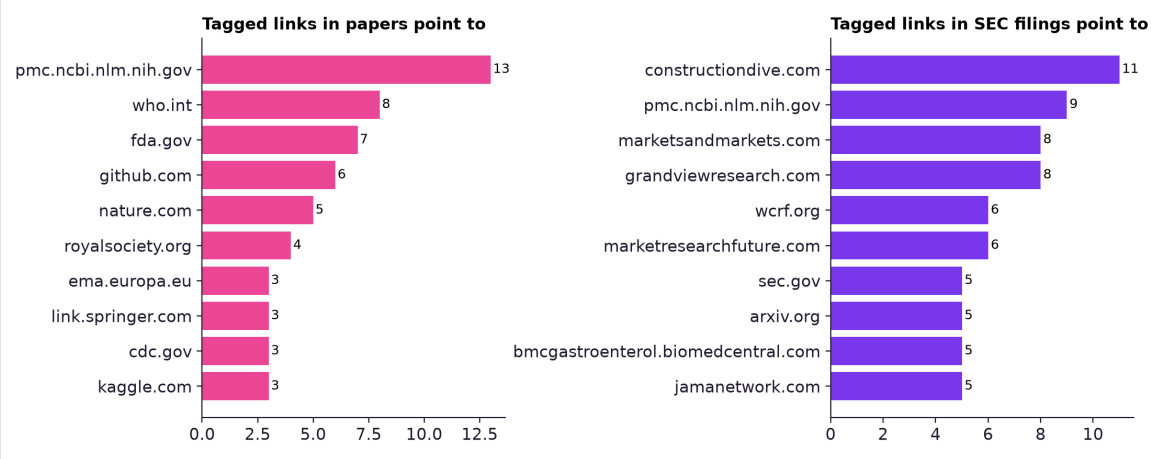
The 169 verified papers come from 110 journals. Their six most frequent publishers are all among the largest open-access publishers and account for 116 of the 169, so the spread follows publishing volume, not a type of journal. The papers hold 272 distinct tagged links to 202 websites. Only 21

papers carry the tag in the reference list; in the rest it sits in the running text, in footnotes, or in data and software statements. The most frequent destinations are a public archive of medical papers, the World Health Organization, the U.S. Food and Drug Administration and a code-hosting site. In total 75 of the 272 links (28%) lead to government or intergovernmental bodies and 23 to scholarly publishers. Eight papers contain ten or more tagged occurrences.

5.4 SEC filings: half are documents that raise money

The 133 verified filings come from 67 companies: 78 were filed in 2025 and 55 in 2026 up to the reading date. Sixty-seven of them (50%) are offering or registration documents, the papers a company files when it asks investors for money (offering circulars, draft and final registration statements, crowdfunding forms, prospectuses). Thirty-four are periodic or current reports and 25 are proxy materials, a category that includes statements written by shareholders.

The filings contain 258 distinct tagged links to 94 websites. **Fifty-two of those links, one in five, point to market-research vendors**, the firms that publish market-size forecasts. Another 32 point to scholarly sources and 28 to government sites. In plain terms, a common use of the chatbot in these filings is to find the figure for how large a market is.



5.5 Federal courts: mostly complaints, in 71 courts

The court archive returned 298 cases in 71 federal courts; in 295 of them the result snippet shows the complete tag. Three districts account for 113 cases: the Southern District of New York (55), the Northern District of California (31) and the Central District of California (27). Where the nature of suit is stated, the largest groups are copyright (30), other statutory actions (28), personal injury (23), civil rights (27), antitrust (15) and patent (15). Among results with a readable description, the tagged document is most often the complaint that opens the case, followed by motions, notices of removal and memoranda. Most cases list counsel, so the trace is not confined to people representing themselves. The archive's search of published opinions returned none, while govinfo's separate collection of court opinions holds two verified ones.

5.6 Governments

In govinfo, 15 documents were verified: five Federal Register rule documents, four congressional hearing records, two congressional committee reports, two federal court opinions and two other government publications. Hearing records include testimony and submissions from outside witnesses, so a tag in a hearing record may come from a witness, not from congressional staff. On

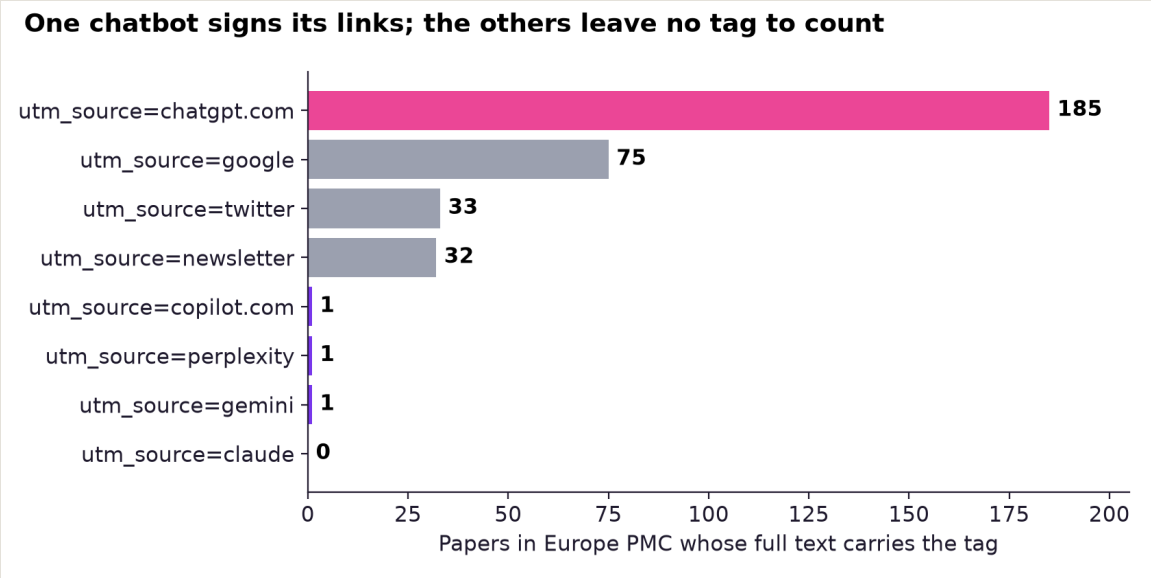
GOV.UK, three publications from three public bodies carry the tag in their attached documents, with 40 tagged occurrences between them; one document alone has 32.

5.7 Wikipedia

The live text of 36 language editions holds 11,702 articles with the tag: 3,166 in English, 1,386 in Spanish, 805 in Japanese, 691 in German, 666 in Italian, 500 in Chinese and 477 in French. English Wikipedia has a further 394 drafts with the tag. Volunteers there remove the tag when they find it, so the live count is what remains after cleaning.

5.8 Other tags

In Europe PMC the ChatGPT tag appears in 185 papers. The tag for Google appears in 75 papers over all years, Twitter in 33 and newsletters in 32: the ChatGPT tag, which is two years old, already outnumbers those three together. Tags for other chatbots are nearly absent: one paper each for Copilot, Perplexity and Gemini, none for Claude. This is not a measure of market share. Most other assistants do not append a tag to the links they give, so they leave nothing to count.



6. Discussion

A floor that cannot be argued with. Every number in this paper is a list of documents that anyone can open and check for a literal string. That is its strength over detector-based estimates. It is also why the numbers are small next to those estimates. The tag is removed by deleting a few characters, by any reference manager that cleans addresses, and by copy editors. What remains are the documents where nobody looked at the link closely enough to notice it. The true number of documents built on ChatGPT-supplied sources is necessarily higher, by a factor this method cannot tell us.

The trace is a proofreading signal. A tagged link in a filed complaint or a published paper means the address was pasted and not read again by the author, the co-authors, the editor, the reviewers or the typesetter. For a journal or a court, the tag is a cheap test of whether that last check happened. It says nothing about the quality of the source: many of the tagged links in papers lead to the World Health Organization or a drug regulator, which are the right sources.

Where the chatbot is used as a search engine. The destinations show what people ask for. In science it is official guidance, datasets and software. In filings that raise money it is, one time in five, a market-size figure from a research vendor. That use deserves attention from the people who read prospectuses, because a language model's choice of which forecast to surface is not a neutral one, and the filer may never have compared it with another.

A signature only one product leaves. The comparison of tags shows that this evidence exists for one chatbot only because of a design decision. If that product stops appending the tag, or if others start, the series will break. The window in which the public record can be read this way may be short.

7. Limitations

The count is a lower bound of unknown tightness, as discussed. The archives are partial: open-access papers, contributed federal filings, published government documents. Search indexes can miss documents whose text was not extracted, and our verification can miss links broken in unusual ways; we excluded everything unverified. Court results were checked from search snippets, not from the filings themselves. A tag shows that a link came from ChatGPT; it does not show who pasted it (an author, a witness, a quoted third party) or that any surrounding text was machine-written. Some documents quote other documents, so a tag can be inherited. Domain groups were assigned by simple rules and the "everything else" group is broad. Wikipedia counts change by the hour. All figures are as read on 12 November 2026.

8. Conclusion

ChatGPT appends its name to the links it hands out, and people paste those links into the public record without looking. On one day in November 2026 the tag stood in 298 federal court cases, 133 SEC filings, 169 peer-reviewed papers, 18 government documents in two countries and 11,702 Wikipedia articles, and in none of them before December 2024. The method needs no detector and accuses no one: it counts a literal string and reads where the links go. As a measure of chatbot use it is only a floor. As a test of whether anyone proofread the links in a document that others will rely on, it is exact.

9. Reproducibility

The search strings, collection scripts, per-document verification results and figure code are in the study repository (`src/europepmc.py`, `src/edgar.py`, `src/wiki_gov.py`, `src/analyze.py`; data in `data/`, tables in `out/results.json`). Every search in this paper can be repeated in a browser by entering `"utm_source=chatgpt.com"` in the public search box of each archive. Counts will have grown since the reading date.

References

Kobak, D., González-Márquez, R., Horvát, E.-Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27).

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Liang, W., Zhang, Y., Wu, Z., et al. (2024). Mapping the increasing use of LLMs in scientific papers. *arXiv:2404.01268*.

Mata v. Avianca, Inc., No. 22-cv-1461 (S.D.N.Y. June 22, 2023) (opinion and order on sanctions).

Roberts, O. (2026, July 7). Detecting ChatGPT in legal documents is easier than you think. *Bloomberg Law*.

Wikipedia contributors. Wikipedia:Signs of AI writing. *Wikipedia, The Free Encyclopedia*. Read 12 November 2026.

Europe PMC, EDGAR full-text search (U.S. Securities and Exchange Commission), RECAP Archive (Free Law Project), govinfo (U.S. Government Publishing Office), GOV.UK search and the Wikimedia search API, all read 12 November 2026.