

The share of LLM-modified text in machine-learning peer reviews keeps rising after 2024

Papex Labs

Hypothesis · Computer science · Meta-research

Signed 5 October 2026

Papex PXH-1VZX70YC146X · version 1

Keywords: peer review; large language models; AI-generated text; OpenReview; research integrity

Claim

In the first major machine-learning conference review cycle of 2026, the estimated share of review sentences substantially modified by large language models exceeds the upper estimate reported for 2023–2024 cycles.

Why it matters

Peer review is a core quality control of science. A rising share of AI-modified reviews changes what a review certifies and supports clear disclosure rules for reviewers.

What would falsify this?

The 2026 estimate, computed with the same distributional method, is at or below the upper bound of the 2023–2024 estimates.

Methodology

Apply the corpus-level distributional estimator of Liang et al. (2024) to publicly released reviews of one 2026 machine-learning conference on OpenReview, using the same reference corpora and word-frequency features, and compare with the published 2023–2024 estimates.

Expected outcomes

A higher estimate than in 2023–2024, with the largest shares in reviews submitted close to the deadline.

Research setup

System or population Public peer reviews of one 2026 machine-learning conference

Variables Estimated fraction of LLM-modified sentences; review cycle year; time to deadline

Parameters Same estimator and reference corpora as the 2024 study

Observer or reference frame *Not applicable* — No physical observer or frame of reference: the units are sentences in public peer reviews, compared with the published 2023–2024 estimates.

Assumptions and scope Public reviews are representative of all reviews at that venue

Units Fraction of sentences

Tolerance or uncertainty 95% confidence interval from the estimator

Research mode Observational

Known limitations The estimator measures corpus-level shares, not individual reviews; newer models may shift word-frequency signatures.

Purpose New claim

Prior evidence Supported

Files

- Pre-registration protocol and codebook — Protocol · PXA-C6T2G6V1TSCZ v1 · 63.8 KB

States in advance how the claim will be tested and what result would falsify it.

- README.md · Markdown · 806 B · 0xe25d...d0e0
- codebook.csv · CSV · 1.8 KB · 0xf59e...59f4
- data-template.csv · CSV · 210 B · 0x6c2c...6a02
- protocol.pdf · PDF · 61.0 KB · 0xbd74...1deb

Preparation

Method Agent-prepared

Application launch-loader 1 (self-reported)

Model claude-opus-5-5 (self-reported)

Origin Every field above: Agent-researched, human-accepted

References

- [1] Weixin Liang, et al. Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews. *arXiv*, 2024. arXiv:2403.07183

Builds on

Generated by Papex from the signed record (research-record/v8). The signature covers the record's content, not this PDF file. Check it at <https://papex.org/verify/PXH-1VZX70YC146X?version=1>. Cite as <https://papex.org/id/PXH-1VZX70YC146X/v/1>.