Twice-told critiques: evidence that LLM-era peer reviews are repeating each other

22 November 2026 · 6 min read · blogs.iictp.org/llm-peer-review-homogenization-twice-told-critiques

Short answer: “Twice-Told Critiques”, an arXiv preprint dated 19 November 2026 by Ilya Gusev, measures how much the criticisms in reviews of the same paper overlap and reports that overlap is rising at several venues, and is higher for reviews that a text detector flags as LLM-generated. It argues the diversity of critiques may be falling, which matters to anyone relying on LLM judges or reviewers. This is an arXiv preprint. It has not been peer reviewed, and its numbers are the authors’ own claims.

TL;DR The short version

- 'Twice-Told Critiques', an arXiv preprint dated 19 November 2026, reports that criticisms in reviews of the same paper overlap more than before at several venues.
- Overlap is higher for reviews that a detector flags as LLM-generated, suggesting the diversity of critiques may be falling.
- It relies on an imperfect detector, but the lesson holds: if many evaluators lean on the same models, their errors correlate.

Read online, listen to it, or change the text size: https://blogs.iictp.org/llm-peer-review-homogenization-twice-told-critiques

What was measured?

Peer review depends on different readers finding different problems. The paper measures critique overlap at claim level in conference and journal reviews. It reports that at ICLR measured overlap rose from 12.0% in 2026 to 21.3% in 2027 for reviews from the same stage of the process, with year-on-year rises also at NeurIPS, ICML and the F1000 and PLOS journals.

How it works

- The author breaks reviews into claims and counts how many criticisms overlap between reviews of one paper.
- MELD, a detector of LLM-generated text, is applied to all reviews.
- Within a paper, reviews flagged by MELD are compared with unflagged ones on how much they overlap.

Results reported by the author

- A third ICLR review added 5.46 new problems to the other two in 2027, against 6.12 in 2026.

- Two MELD-flagged reviews of one paper overlapped more than two unflagged ones: by 3.0 points at ICLR in 2024–2025, 4.5 points in 2026–2027, and 5.4 points across five journal sources.
- The abstract concludes the diversity of critiques among reviews of a single paper is suggested to be falling.

Limitations and open questions

- Preprint, not peer reviewed; only the abstract was read.
- The study depends on a detector for LLM text. Detectors can make errors, and the abstract says the evidence is that flagged reviews are less distinct, not that any review was written by an LLM.
- Rising overlap could have other causes, such as changes in reviewing guidelines. The abstract says the results “suggest” falling diversity; they do not prove cause.
- The year labels are as given in the preprint.

What it means for LLM practitioners

If many evaluators lean on the same models, their errors correlate and a panel of judges stops being a panel. The same risk applies to using one LLM as judge for another model of the same family. Use different judges, and calibrate against humans; see how to evaluate an LLM application.

Frequently asked questions

- **Does the paper prove LLMs wrote the reviews?** No. It uses a detector and reports an association.
- **Which venues showed rising overlap?** ICLR, NeurIPS, ICML, F1000 and PLOS journals, per the abstract.
- **Why does it matter beyond peer review?** Correlated evaluators weaken any ensemble of judges.

Related reading

- How to evaluate an LLM application
- What is LLM hallucination?
- Responsible AI in practice

Sources and currency

Sources were read on 22 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail.

> Related credential: AI Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Gusev I. Twice-Told Critiques: Claim-Level Homogenization of Peer Review in the LLM Era. arXiv:2611.26286. 19 November 2026. arxiv.org/abs/2611.26286.

Published by IICTP, an independent technical credential body, on 22 November 2026. The current version of this post is at https://blogs.iictp.org/llm-peer-review-homogenization-twice-told-critiques. More guides: blogs.iictp.org.