AI in research and publishing (Sep 2026)

Jason Fantl
Averroes Labs

September 2026

Originally published on LessWrong on 24 September 2026.

epistemic status: I have low confidence in these findings, primarily because my anecdotal experience (I'm a researcher) is that AI usage in research has been changing more quickly in the recent months, so analysis of the last year gives a very fuzzy picture. A lot of the reports rely on Pangram or self-reporting, which is another methodological weakness.

I wanted a better idea of AI usage and impact in the research community, so I spent a few days reading recent articles (mostly published in the last few months, some are a year old) and summarized them here.

Recent AI usage for research

Anthropic claims that 26% of AI R&D work is now being led by AI (still some human oversight). Only 6 months ago they claim researchers had primarily been "collaborating" with AI and AI led research less than 1% of the time. This is a significant change in a short period of time.

OpenAI claims it has built an "automated research intern", an AI agent capable of accomplishing well-scoped problems with some human steering, helping researchers move at increasing rates and solve more complex tasks. They claim 70% of researchers now run 4 or more agents concurrently, that researchers are running 1.6x more experiments each day compared to 2025, and that agents are successfully performing complex research tasks without any intervention 15 percentage points more often compared to 7 months ago. The longer tasks (4-8 hours) which were successful still require at least one intervention over half of the time.

Google analyzed Gemini conversations and conducted a survey and found that 47% of surveyed US/UK scientists use AI daily, and only 7.2% use it rarely or never. Researchers claim they save on average 7 hours a week due to AI, although 46% of those who save time spend over a quarter of the savings verifying AI output. They also find 41% report a growing backlog of untested hypotheses, and 40% say the number of low-quality papers in their field increased.

All of these claims are reports from AI companies, who are incentivized to promote AI usage. We next look at trends in academic publications to hopefully better understand AI's impact on the general research community. Although, even there, many of the results come from Pangram and other reviewing tools, which are incentivized to find more AI usage.

Concerning trends in publishing

The journal of *Organization Science* found that 30% of paper abstracts and 10% of peer reviews were primarily AI written (70%-100% of the total text detected as AI according to Pangram), and only 40% of submitted papers and 65% of peer reviews were primarily human written (0%-15% AI) in early 2026. The quality of submitted papers has been decreasing (by the journal's metrics), and reviews are becoming more homogeneous.

They find the submission volume has risen by 42% since 2022, which they primarily trace to AI combined with publish-or-perish incentives. In their own words, "we view the current system as unsustainable".

The volume of papers has increased dramatically across conferences. The 2026 Association for the Advancement of Artificial Intelligence (AAAI) had to review roughly 23,000 papers, nearly twice as many as the previous year. Many of the new submissions are traced to China and first-time authors. It is unclear how much is enabled by AI. The 2026 International Conference on Learning Representations (ICLR) received 19,525 valid submissions, up from 11,603 in 2025, roughly a 68% increase in one year. Abstract registrations for ICLR 2027 reached roughly 62,000 (estimated from submission IDs), up from 25,649 for 2026.

As a recent anecdote, an editor-in-chief of the Transactions on Machine Learning Research (TMLR) journal found that of 10 sampled authors with papers slated for desk rejection: 3 did not meet and 6 were unable to answer technical questions about their own papers, leaving just 1 of the 10 that could answer technical questions. An interesting proposal is to quiz authors with proctored assessments to verify the authors understand their own papers.

A Pangram case study claims that 21% of peer reviews in 2026 ICLR were entirely AI written, and the more AI usage they detected in a review the higher the score it was likely to give a paper.

ICML ran a randomized study to test the impact of restricting LLM usage in peer review, and observed substantial noncompliance. Policy assignment had near-zero effect on decisions, scores, or reviewer confidence. Anonymized self-reported surveys found 22.5% of the reviewers prohibited from LLM usage still used an LLM, and 36.5% of reviewers under a more permissive policy still reported at least one explicitly disallowed use.

What people are doing about this

Restricting AI use

Conferences such as Neural Information Processing Systems (NeurIPS) are now restricting AI usage in position papers and desk-rejecting papers if a substantial portion is detected to be AI. They desk-rejected 18.4% of papers based on Pangram scores, which caused significant backlash focused on false positives and the sensitivity of Pangram settings. ICLR now desk-rejects papers with hallucinated references, and in 2027 they will rate-limit submissions to 20 papers per author, and to just 1 if this is the first major publication for all authors on a paper.

ArXiv, the primary pre-print platform for AI-related publications, is taking steps to punish unchecked AI output. If a paper contains incontrovertible evidence that the authors did not check LLM output (hallucinated references, leftover chatbot meta-comments, placeholder data), the authors are banned for a year, and afterwards their submissions must first be accepted at a peer-reviewed venue. Review articles and position papers in the CS category must first be accepted at a journal or a conference before being posted on arXiv.

Incorporating AI use

The AAAI conference was able to trial-run AI peer reviews at scale, providing one AI generated review for each paper. Authors mostly rated the AI peer reviews higher than the human reviews for 6 of 9 quality criteria. TMLR is adding one AI review to every submission. The AI review evaluates only soundness, makes no accept or reject recommendation, and is only advisory. NeurIPS 2026 is running a randomized experiment in which reviewers are assigned per paper to no LLM assistance, open-ended LLM assistance, or structured LLM assistance.

HuggingFace hosted a replication hackathon for the papers in ICML 2026. Human-directed agents attempted to replicate 34% of the published papers, and were able to replicate at least one claim for 51% of attempted papers, and falsify or contest a claim in 23%, some of which even found effects opposite to the

paper's claims. Some papers were replicated or falsified multiple times by different users, but some of those reproductions contradicted each other, so replications are still noisy.

New types of scientific structures are also being proposed and trial-run in response to this change:

- Agents4Science, an experimental conference for AI-led papers (they must be the first author) reviewed by AI. Humans were co-authors and provided a second review for papers that AI scored highly.
- The Alignment Journal, a journal being created to allow and incorporate AI into the review process.
- The Agent-Native Research Artifact (ARA) protocol, a proposal to replace the existing PDF format for papers with a more agent-friendly data structure.

Personal thoughts

I expect publication rates will keep increasing, but primarily for lower-quality work. The quality should also increase with model capabilities, but I suspect that to progress at a slower rate. If so, the average quality of submitted research will initially decrease with time, and once the quality from AI catches up to humans, it will increase again.

Journals claim they will not be able to keep up, and are now trying new methods to try and adapt. I expect that more researchers will quickly adopt AI and the number of agents used per researcher will also increase, to the point of breaking the existing review process. I suspect we will need to design a review process whose capacity can scale linearly with the number of agents being used to generate research (volume of submissions), and we will need to do this quickly. This will require incorporating AI into the review process, which introduces its own set of challenges such as: correlated errors, optimizing papers for known reviewers (there are a limited number of models available to use as reviewers), and prompt-injection attacks.

Broad bans on AI use do not seem enforceable, except for the most detectable cases such as hallucinated citations. We should focus on filtering out the bad work, using AI to keep parity with the volume of submissions.

Here are some ideas that come to mind in response to all this:

- Require that papers are replicated as a part of the desk-rejection process (part of the submission cost is replicating random papers, similar to peer review, so it should scale). This would incentivize papers to be replicable and filter out more papers to decrease load on human reviewers. Some challenges include noisy results (AIs are still not on par with humans), correlated failures, papers that cannot be reproduced (too costly, theory-based, position papers).
- Require papers pass peer-review from all SoTA models given a system prompt provided by the journal. These can be gamed since the prompt would be public, but such a requirement can still raise the floor for the quality of papers. The journal prompt could be a living document as failure modes are found.
- Someone builds a peer-review RL environment. This feels close to "just RL research taste bro", but it does not seem like an *impossible* task, and it would be well worth pursuing if possible. It seems like frontier labs are already pursuing this.
- A new category of publication dedicated to deciding on important metrics we want to optimize in the field. Since agents seem especially well-suited to hill-climbing, we should invest more effort into deciding what metrics we want our army of research agents to be aimed at.
- Raise the cost of submissions. If the low-quality work is increasing in volume, then raised submission costs would disproportionately affect the low-quality work. If that additional cost is paid in replications, then we might even be able to turn the volume of submissions into a benefit. *Organization Science* floats the idea of nonlinear submission fees, with respect to the number of papers you submit.