

You Agreed to the Rule. Did You Agree to the Experiment?

ICML 2026 and the missing boundary between review duty and experimental participation

Jooyeol Kim

Independent Researcher | Seoul, Republic of Korea

jooyeol@jooyeolkim.com

Working Paper | Version 1.2 | July 2026 | Not Peer Reviewed

Abstract

This working paper examines a missing boundary in ICML 2026’s policy for LLM use in reviewing. ICML assigned reviewers enforceable AI-use rules while also using some assignments to compare outcomes under Policy A and Policy B. The paper distinguishes agreement to perform review duties under a rule from agreement to have those duties used as experimental conditions. It does not allege a legal violation, an IRB failure, or research misconduct. Instead, it asks whether a reviewer could decline the experimental use of an assignment while remaining in the ordinary reviewing role, particularly where reciprocal-review conduct could affect the reviewer’s own submissions. A fictional idol-audition scenario illustrates how institutions can begin from the point after entry and treat condition A or condition B as the only remaining options, overlooking the possibility of remaining in the underlying process while declining an additional experimental purpose. Comparisons with NeurIPS 2026 and ICML’s 2023 ranking experiment show that ordinary participation, experimental participation, and decision consequences can be separated. The paper concludes that institutions should answer two questions separately: whether a participant agreed to a rule, and whether the participant agreed to an experiment.

Keywords: peer review; artificial intelligence; experimental participation; institutional governance; meaningful refusal; ICML; NeurIPS

Suggested citation

Kim, Jooyeol. 2026. “You Agreed to the Rule. Did You Agree to the Experiment? ICML 2026 and the Missing Boundary Between Review Duty and Experimental Participation.” Working paper, Version 1.2.

AI use disclosure

Generative AI tools were used during drafting and revision to test arguments, improve structure, check internal consistency, and edit language. The author directed the analysis, independently verified factual claims against the cited sources, substantially revised the text, and assumes full responsibility for the final manuscript.

In March 2026, the International Conference on Machine Learning desk-rejected 497 submissions tied to 398 reciprocal reviewers after concluding that Policy A reviews had used LLMs in violation of an accepted no-LLM condition.^[1] ICML presented the episode as enforcement of an agreed rule and emphasized that every flagged review had been manually verified.^[1, 2] Critics of the hidden-prompt method used a different word: trap.^[8]

The dispute therefore settled around a familiar choice. Was this entrapment, or enforcement?

ICML’s own policy contained a third issue. Under the heading “**Randomization, monitoring, experiment,**” the conference said that some reviewers’ policy assignments could be randomized to compare outcomes under Policy A and Policy B.^[2] The same assignments were also rules the reviewers had to obey. Deviation could be treated as an academic-integrity violation and could expose a reviewer’s own submissions to desk rejection.^[2]

Some reviewers broke the rule. That fact matters. It does not answer a different question.

When did agreement to a review rule become agreement to participate in an experiment?

The audition

Imagine a televised idol audition run by a major broadcaster and one of the few agencies capable of launching a career. The contestants are teenage girls and young women who have trained for months or years. Families have paid for lessons and travel. School, friendships, bodies, and daily schedules have been reorganized around the hope that one of these doors might finally open.

The organizer also wants to compare two methods of teaching choreography. One group may use phones and video. Another must learn from written and drawn instructions. The method is folded into the audition rules, and violating the assigned condition can lead to elimination.

A girl in the restricted group falls behind, checks a phone, and is caught. She did break the rule. The evidence may be exact. The organizer may also have a serious reason for comparing the two methods.

But she was never given a way to say:

I will follow the ordinary audition rules. I do not want my attempt to become part of this experiment.

and still remain in the audition.

She can leave, of course. She can abandon the chance for which she and her family have spent years preparing. That is an exit from the opportunity, not a refusal directed at the experiment.

The comparison is not meant to make a machine-learning conference morally or legally identical to an idol audition. It isolates a narrower point. The organizer sees people who have already entered the process and treats the remaining alternatives as obvious: condition A or condition B. The participant entered for something else. She entered to audition.

The missing option is not A or B. It is:

I will remain in the underlying process, but I do not agree to the experimental use layered onto it.

The missing third option

Follow one reviewer through ICML’s process.

First, the reviewer stated whether they preferred a strict no-LLM rule or a limited-use rule. Authors separately decided whether their papers required the strict version. ICML then gave the reviewer one binding rule for all assigned papers.^[2]

For some reviewers, that assignment did another job as well. It helped ICML compare what happened under the two policies. The conference could use some assignments to monitor differences in scoring and to study the effects of Policy A and Policy B.^[2]

The reviewer therefore encountered one assignment in two institutional capacities. It governed how the reviewer had to work, and it also supplied the condition from which ICML hoped to learn.

ICML’s policy begins from the point at which the reviewer is already inside the role. From there, the available alternatives are easy to present as Policy A or Policy B. But that framing starts too late. Before asking which condition a reviewer should receive, the institution had another question to answer: could the reviewer perform the ordinary role without having that role used for the experimental comparison?

That distinction mattered because many reviewers were also authors. ICML described reviewing as giving back to a community whose papers are assessed by peers, and authors had to nominate qualified reciprocal reviewers.^[3] An author who required the strict policy for their own paper also had to be willing to follow it while reviewing.^[2]

In March, the link became concrete. Review conduct led to the rejection of 497 submissions tied to 398 reciprocal reviewers.^[1] People who supplied peer labour in one role saw their own papers affected in another.

The rule was inside the gate. So was the experiment. What the public policy did not separately show was an entrance to the first without automatic inclusion in the second.

ICML had reasons

ICML was responding to real problems. Authors may not want confidential manuscripts entered into LLMs. Reviewers should not delegate critique to a model. A conference handling thousands of papers needs rules it can assign and enforce with more than suspicion. ICML surveyed reviewers and distinguished permitted from prohibited uses.^[4] It disclosed possible randomization in the policy and manually checked every flagged review before acting.^[1, 2]

Those steps explain why the rule existed, how it was assigned, and how ICML decided that it had been broken.

They do not answer whether a reviewer could say:

I will follow the conference’s review rules, but do not use my assignments in the experiment.

The public policy describes agreement to the assigned rule. I found no separately described route for declining the experimental comparison while remaining in the reviewer role.^[2, 4]

A rule can be justified, clearly disclosed, and accurately enforced without thereby settling whether participation under it may also be used as an experimental condition.

This is a governance question, not a finding that ICML violated the Common Rule, bypassed an IRB, or committed research misconduct. The public record does not support those conclusions. The narrower point is enough: ICML placed experiment and enforcement inside one policy architecture without separately explaining how a reviewer could refuse the former and remain subject to the latter.^[2]

Another conference drew the line

Separating ordinary participation from experimental participation may sound easy only because an essay does not have to assign thousands of reviews under deadline. NeurIPS 2026 nevertheless built that separation into a live conference.

Its AI-Assisted Reviewing Experiment included only papers whose authors opted in and reviewers who volunteered. Participating reviewer-paper pairs were then assigned to one of three reviewing conditions. NeurIPS said participation was “entirely voluntary” and would have no bearing on paper decisions or other conference outcomes. It also identified a University of Texas at Austin IRB protocol.^[5]

NeurIPS still expected reviewer-authors to complete their ordinary duties and retained desk rejection as a last-resort sanction for gross negligence after notice and verification.^[6] The conference kept its review obligations and its authority over submissions. It simply did not treat entry into those roles as automatic entry into the experiment.

ICML’s own recent history offers a smaller comparison. The paper documenting its 2023 ranking experiment says authors were told that their survey data would not affect that year’s decisions, and that analysis began only after accept-or-reject decisions had been announced.^[7]

Neither comparison proves wrongdoing in 2026. They show something more limited and more useful. A conference can keep an ordinary gate while marking where experimental participation begins.

Figure 1 condenses the contrast. The missing option is not another experimental condition, but the ability to remain in the underlying role while declining the experiment.

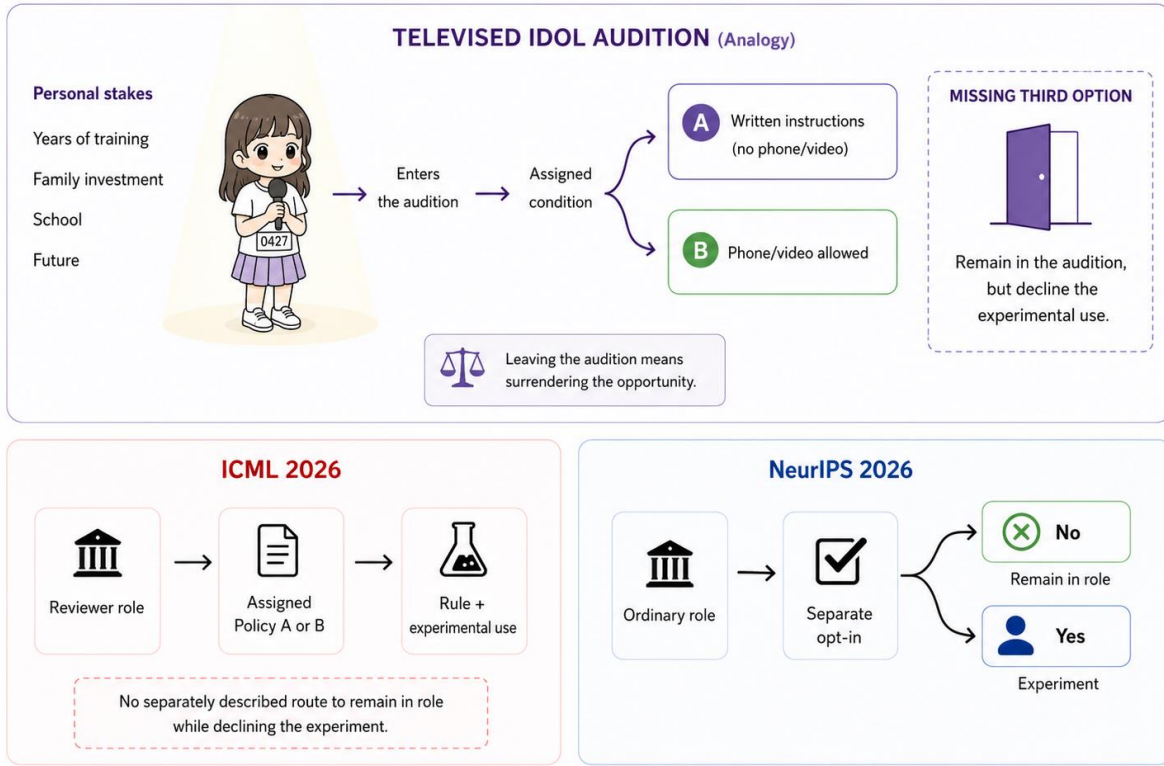


Figure 1. The missing third option. The audition analogy illustrates why leaving an opportunity is not equivalent to declining only an experiment. The lower panels contrast ICML’s publicly described combination of policy assignment, enforcement, and experimental comparison with NeurIPS’s separate opt-in boundary. The diagram is a simplified representation of the relevant participation structures.

From peer to compliance subject

Reciprocity first addressed the reviewer as a peer. Authors were asked to give back because their papers were reviewed by others.^[3] ICML later defended enforcement through the language of community and “our trust in each other.”^[1]

Once violation was detected, the same person entered another institutional grammar. Flagged reviews were removed, submissions tied to reciprocal reviewers were desk-rejected, and reviewers with LLM use detected in more than half of their reviews were removed from the pool.^[1]

Peerhood creates the duty. Compliance classifies the failure.

Both roles can be legitimate. Their sequence still matters. A person can move from peer supplying labour, to reviewer assigned under a policy comparison, to compliance subject, to author exposed to sanction. ICML made the duties and penalties legible. The experimental use of the role was less clearly separated.

Community cannot be only the language used when labour and trust are needed. When peers also become units in an experiment, that role should become visible while they can still direct a meaningful no at it.

What a flagship makes reusable

Prestigious institutions do more than decide who passes. They also produce examples that others can borrow.

In July 2026, *The Transmitter* reported that ICML Scientific Integrity Chair Nihar Shah had shared the conference’s hidden-prompt methodology with the NeurIPS team, which used a similar intervention to detect prohibited AI use.^[8] This was not NeurIPS adopting ICML’s A/B policy experiment. It was one contested enforcement component crossing an institutional boundary.

The distinction matters. So does the movement.

NeurIPS built an explicit opt-in boundary around randomized research in one part of its review system while using a related hidden-prompt intervention in another.^[5, 8] The same institutional ecosystem displayed both a separable experiment and a transferable enforcement technique.

A flagship’s choices become available as precedent before the field has settled the ethical terms of the dispute. Power makes a governance technique easier to run, easier to defend, and easier for others to repeat.

“Everyone does it” is a weak institutional defense.

“ICML did it” is a powerful one.

The boundary that was missing

ICML explained the rule. It documented the breach. It defended the need for enforcement and described the experimental comparison.

What its public materials did not separately explain was how a reviewer could refuse the experimental use of an assignment while remaining in the ordinary reviewing role.

That omission matters because institutional entry is not unlimited consent. A person who enters a review system agrees to perform a role under its rules. That does not automatically settle whether the same role may be used as an experimental condition for a second institutional purpose.

NeurIPS showed that the distinction was operationally possible. It preserved ordinary conference roles while making experimental participation voluntary and neutral to paper decisions.^[5, 6]

The reviewers who violated ICML’s policy remain responsible for that violation. Proof of the breach should not be allowed to answer every other question about the structure surrounding it.

The central question is therefore simpler than a new general theory:

You agreed to the rule. Did you agree to the experiment?

A serious institution should be able to answer both questions separately.

References

- [1] ICML 2026 Program Chairs et al., “On Violations of LLM Review Policies,” *ICML Blog*, 18 March 2026, updated 26 March 2026: <https://blog.icml.cc/2026/03/18/on-violations-of-llm-review-policies/>
- [2] ICML 2026, “Policy for LLM use in reviewing”: <https://icml.cc/Conferences/2026/LLM-Policy>
- [3] ICML 2026 Program Chairs, “What’s New in ICML 2026 Peer Review,” *ICML Blog*, 8 January 2026: <https://blog.icml.cc/2026/01/08/whats-new-in-icml-2026-peer-review/>
- [4] ICML 2026 Program Chairs et al., “Introducing ICML 2026 policy for LLMs in reviews,” 10 December 2025: <https://icml.cc/Conferences/2026/Intro-LLM-Policy>
- [5] NeurIPS 2026, “AI-Assisted Reviewing Experiment”: <https://neurips.cc/Conferences/2026/ai-reviewing-experiment>
- [6] NeurIPS 2026, “Main Track Handbook,” section “Responsible Reviewing Policy for Reviewer-Authors”: <https://neurips.cc/Conferences/2026/MainTrackHandbook>
- [7] Buxin Su et al., “The ICML 2023 Ranking Experiment: Examining Author Self-Assessment in ML/AI Peer Review,” arXiv:2408.13430v3, especially section 2, “Experimental Design and Summary Statistics”: <https://arxiv.org/html/2408.13430v3>
- [8] Dalmeet Singh Chawla, “Scientists decry conference’s use of hidden prompts to snare AI peer reviews,” *The Transmitter*, 1 July 2026, doi:10.53053/FCUI5388: <https://www.thetransmitter.org/publishing/scientists-decry-conferences-use-of-hidden-prompts-to-snare-ai-peer-reviews/>