

Assist, not assess

How AI can relieve the peer review bottleneck by improving the workflow around the reviewer

Medware Group white paper, October 2026

Executive summary

Most discussion of AI in peer review asks whether a machine can review a paper. The evidence says it cannot do so reliably today, and the rules of most journals say it should not. That question also misses where the delay sits. Much of the time between submission and decision is spent on work around the review: checking the submission is complete, screening for integrity problems, finding reviewers who will say yes, chasing them, and pulling their reports together for the editor.

This paper sets out a vendor neutral framework for using AI on that surrounding work, leaving every judgement about the science with people. It is built on one design principle:

Use AI to help with preparation, checking and organisation. Editors and reviewers should still assess the science and make the decisions.

The framework covers eight steps of the workflow. For each it states what AI does, what stays with a person, and how strong the evidence is. It then sets out the guardrails a journal needs and a way to measure whether the change worked. Section 8 brings the steps together as a single model of where manuscripts wait and what might be done at each point. That model is an unvalidated hypothesis, and section 8 also sets out how it should be validated before anyone builds it.

Two limitations to keep in mind:

- **This approach is not yet proven to shorten turnaround.** No published study has measured the effect of any AI tool on time to decision at a medical journal [1]. The framework is a reasoned hypothesis, and sections 6 and 8 describe how to test it.
- **Several steps rest on proposal, not evidence.** The table in section 4 marks which is which.

1. The bottleneck

1.1 The pressure

- Only 35.2% of 257,025 reviewers invited by 21 BMJ Group journals agreed to review [2].

- At one family medicine journal, invitation acceptance fell from 56.14% in 2020 to 35.71% in 2024, and completion of agreed reviews fell from 89.31% in 2021 to 76.19% in 2025 [3].
- ScholarOne reports first quarter 2026 submissions up 33% on the year before. This is vendor data, not independently checked [4].
- Peer review consumed an estimated 100 million researcher hours in 2020 [5].

1.2 Where the time goes

The reading of a manuscript by a willing reviewer may not be the slow part. Once a reviewer has agreed, the review comes back in about two weeks:

- A mean of 11.8 days at 13 JAMA Network journals [6].
- A median of 16.4 days in a large cross publisher dataset from 2018 [7].

The harder problem is getting to that point, and what happens around it:

- Roughly two of every three invitations are declined or ignored [2], so editors send several rounds of invitations.
- Invitations are often poorly targeted. In a small study of six senior emergency physicians, only 28.1% of 139 review requests were directly relevant to the recipient's expertise, and 2.3% were accepted [8].
- Editors spend time on manuscripts that were never going to be sent for review. The BMJ is reported to reject about half of submissions at initial screening [9].

We are inferring from these figures, and no study has broken down a medical journal's turnaround time step by step. But the inference is reasonable: a large share of the delay is logistics and preparation, not expert reading. That is the part of the workflow where AI assistance carries the least risk. Two cautions apply. The review times above come from different settings and years, one a 2025 mean and the other a 2018 median. Neither is set against the total time to first decision at the same journals, so we cannot say what share of the delay the review itself accounts for. Section 8 proposes measuring this directly.

2. Why AI should not be the reviewer

Three lines of evidence argue against placing AI in the reviewer's or the decision maker's role.

It is too lenient. At three ophthalmology journals, human reviewers recommended rejection for 73.33% of 300 manuscripts, and each LLM tested did so for 2.00% [10]. Similar results came from oral surgery [11] and radiology [12]. A model that almost never recommends rejection cannot reliably screen manuscripts.

It can be manipulated. A hidden instruction in the manuscript changed the AI's recommendation in 84.4% of 720 simulated reviews of medical papers [13]. Authors have already been caught doing this in real preprints [14].

The rules do not allow it. ICMJE requires reviewers to protect confidentiality and disclose any AI use [15]. Elsevier, Springer Nature, Wiley, PLOS and others permit AI only in a supporting role, with people making the judgements [16, 17]. Reviewers themselves are uneasy: 57% of surveyed researchers would be unhappy if a reviewer used generative AI to write a review of their paper [18].

The evidence leaves room for AI to help with the workflow, but does not support using it to assess the science.

3. The design principle

A useful test for any proposed use of AI in the workflow is to ask three questions:

1. **Is the output checkable?** A flagged duplicate image or a missing checklist item can be confirmed by a person in seconds. An opinion on novelty cannot.
2. **Does a person make the call?** AI output should arrive as information for an editor or reviewer, never as a decision that takes effect on its own.
3. **What happens when it is wrong?** A missed checklist item costs little. A wrongful rejection costs an author months. Use AI for tasks where mistakes have limited consequences and people can spot them.

An AI tool should meet all three conditions before a journal uses it in its workflow. If it does not, it should be tested in a research study first.

4. The framework: eight places AI can help

The table summarises the framework. The sections below explain each step.

Step	What AI does	What stays with people	Evidence
1. Intake and completeness	Checks files, declarations, registration numbers and reporting checklists	Deciding whether to return the manuscript to the author	Moderate for accuracy. None for workload
2. Integrity screening	Flags image duplication, text overlap, paper mill signals, altered trial outcomes	Interpreting every flag and contacting authors	Strongest of the eight
3. Triage briefing	Lists checkable facts about the manuscript and the check results, for the editor	The decision to reject, transfer or send to review	Thin. Two small studies

Step	What AI does	What stays with people	Evidence
4. Reviewer finding	Suggests qualified candidates, screens conflicts, widens the pool	Choosing and inviting reviewers	Widely sold. Not evaluated on outcomes
5. Reviewer support	Supplies a structured brief and gives feedback on the reviewer's own draft	Reading the paper and writing the review	One large randomised study, outside medicine
6. Process management	Tracks deadlines, sends reminders, reassigns stalled invitations	Handling exceptions and relationships	Proposed. No published evidence
7. Decision preparation	Sets the reports side by side, each point linked to its source	The editorial decision and the letter	Proposed. No published evidence
8. Revision check	Matches each request to the author's response and the change made	Judging the revision and whether it goes back to reviewers	Proposed. No published evidence

The steps do different jobs. Steps 1, 2, 3 and 7 mainly save staff and editor minutes. Much of that work happens alongside other work, so it may not shorten the time to a decision. Steps 4 and 6 target waiting, which is where days are most likely to be lost, and step 8 targets both. A journal whose main aim is faster decisions should start with steps 4 and 6.

4.1 Intake and completeness

What AI does. Confirms that the submission has what the journal requires: ethics and funding statements, trial registration, data availability, the right reporting checklist and the items on it.

Evidence.

- At The BMJ, LLMs extracted CONSORT items from 50 trial manuscripts with up to 93.4% accuracy in about a minute each [19].
- Two other studies found about 90% accuracy on CONSORT [20, 21].
- Accuracy depends on the model. Three models scored the same 20 trials at 80.8%, 64.4% and 54.7% compliance [22].
- For PRISMA, one study found high sensitivity (95.1%) but low specificity (49.3%), so many flags were false [23].

Caution. Earlier non AI trials that pushed reporting checklists at submission did not improve published reporting [24, 25]. A completeness check saves staff time only if it replaces a manual check that is already being done.

4.2 Integrity screening

What AI does. Screens for duplicated or manipulated images, text overlap, signs of paper mill origin, and changes between a trial's registered and reported outcomes.

Evidence.

- Image screening at the American Society for Microbiology found duplications in 3.9% of 2,627 accepted manuscripts before publication [26].
- At AACR, tool assisted image checks took 4.4 minutes per manuscript, against 8 minutes by hand [27].
- An LLM checked trial outcomes against the registry in about 2 minutes, against 27 minutes by hand, at 99.8% matching accuracy [28].

What stays with people. Every flag needs interpretation. In the microbiology pilot most duplications were unintentional and were fixed by the authors, and acceptance was revoked in only six cases [26].

Caution. AI text detectors have a history of false positives against non native English writers [29]. A detector flag should prompt a question, never a rejection.

Order and cost. Some checks are cheap and protective, such as exposing hidden text and checking text overlap, and belong at intake. Others, such as image forensics, cost more per manuscript. At a journal that rejects many submissions at the desk, these can run after the decision to send for review. The microbiology pilot screened accepted manuscripts [26].

4.3 Triage briefing

What AI does. Gives the handling editor a one page brief: the study's design, size, registration and primary outcome, each taken from the manuscript with its location, and what the intake and integrity checks found. It gives no view on whether the paper fits the journal's scope.

What it does not do. Recommend acceptance or rejection as a decision.

Evidence. Thin, and mixed:

- At a sports medicine journal, when a local model recommended rejection or transfer, 90.5% of those 21 manuscripts did receive that outcome. Its overall agreement with decisions was poor [30].
- At an orthopaedic journal, three models reached about 60% accuracy on 32 manuscripts, and a small prompt change flipped the results [31].

What this means. The models are not dependable enough to decide. A brief that saves the editor reading time, while the editor still decides, is the defensible use. Whether it saves time has not been measured. A brief limited to facts that can be checked against the manuscript passes the three questions in section 3. A free-text summary of the paper's merit or fit does not, and should be tested in a research study before a journal relies on it.

4.4 Reviewer finding

What AI does. Matches the manuscript to researchers with relevant recent work, screens for conflicts of interest, and deliberately widens the pool beyond the editor's usual contacts.

Why it matters. This step is the most plausible place to recover time, given that most invitations fail [2] and many are poorly targeted [8].

Evidence. Matching tools are built into the major submission systems, but we found no study at a medical journal showing that they increase acceptance of invitations or shorten the time to secure reviewers [1].

Caution. Left unprompted, an LLM suggested reviewers who were 68% men and 95.3% from high income countries. A diversity instruction changed this to 51% women and 42.3% high income, with similar publication records [32]. The journal needs to make this an explicit part of the instructions. Ranking candidates by how likely they are to accept would also send more requests to the same small group of reviewers, so any matching tool needs a cap on each reviewer's open reviews.

4.5 Reviewer support

What AI does. Two things, neither of which is writing the review:

- Supplies the reviewer with the structured brief and check results from the earlier steps, with a note of what each check covered and how reliable it is. Reviewers are asked to confirm the results, not to assume those points are covered.
- Gives feedback on the reviewer's own draft, pointing out vague or unprofessional comments before it is submitted.

Evidence. A randomised study at a computer science conference gave AI feedback on more than 20,000 reviews. 27% of reviewers revised their review in response [33]. This was not a medical journal, and reviewer time was not measured.

Condition. Any such tool must run inside the journal's own secure environment. Frontiers, for example, provides an internal AI tool inside its review forum and bars uploads to external tools [40].

4.6 Process management

What AI does. Tracks where each manuscript is, sends reminders, notices when an invitation has gone unanswered, and proposes the next candidate to the editor.

Evidence. None published. This is conventional automation as much as AI, and it carries the least risk of the eight steps because it touches no scientific content. Some of the gain at this step may come from process changes that need no AI, such as inviting several reviewers at once or shortening the time allowed to reply. A journal should introduce and measure these separately, so that any effect is credited to the right change.

4.7 Decision preparation

What AI does. Sets out the reviewers' reports side by side for the editor: each reviewer's recommendation and each request, linked to the text it came from, so the editor can see where they agree and where they conflict.

What stays with people. The decision and the decision letter. Sage, for example, bars editors from using generative AI to write decision letters [39].

Evidence. None published on this use. One study predicted decisions from reviewer comments with high accuracy [34], which shows the information is in the reports but does not show a time saving. A free-text summary of the reports could misstate a reviewer, and the editor could check it only by reading the reports. Only the linked, side-by-side form meets the three questions in section 3.

4.8 Revision check

What AI does. When the author returns a revised manuscript, lists each request made by the reviewers and the editor, the author's response to it, and where the manuscript changed. Requests with no response, or no matching change, are flagged.

What stays with people. Whether the revision is adequate, and whether it needs to go back to the reviewers. If it does, reviewers can be sent only the points still open.

Why it matters. Many papers go through at least one revision, and each round of re-review repeats the slowest parts of the process: securing reviewers and waiting for their reports.

Evidence. None published. The evidence review covered the path from submission to decision [1], so revision rounds were outside its scope.

5. Guardrails

A journal adopting any of these steps needs the following in place.

Confidentiality

- Manuscripts are processed only in a secure, contracted environment, never in a public tool. This is the common requirement across ICMJE and publisher policies [15, 16, 17].
- Any reuse of reviewer reports or editorial decisions to test or improve tools needs the same protection, a stated retention period, and notice in the journal's terms for reviewers.

Defence against manipulation

- Hidden text should be stripped or exposed before any model reads a manuscript. In one study, converting the document to an image and reading it back removed the effect of injected prompts [35].

Human decision at every gate

- No AI output triggers a rejection, an invitation or a letter without a person approving it. Whether routine, reversible actions, such as further invitations from a list the editor has already approved, could safely run within rules a person has set, instead of waiting for approval each time, is one of the questions the model in section 8 is designed to test.
- A person approving is not the same as a person judging independently. Editors who see AI output first may anchor on it. For a random sample of manuscripts, the editor should record a decision before seeing any AI output, so the journal can tell whether agreement with a tool reflects a good tool or deference to it.

Transparency

- The journal publishes which steps use AI. Authors and reviewers are told. COPE guidance asks for this [36].

Monitoring

- Track false positive rates for each check, and audit reviewer suggestions for geographic and gender spread.
- Recheck performance whenever the underlying model changes, since results vary by model [22].

Adoption

- Tools must sit inside the editor's existing screen and be explained. One publisher saw editor engagement with its AI ethics checks fall from 40% to 20%, and a third of editors did not know the tool existed [37].

6. Measuring whether it works

A journal that measures the effect on turnaround time will add evidence that is currently missing. It will also be better able to manage the risks of its own use of AI.

Baseline first. Record at least three months of current performance before changing anything.

Core measures

- Days from submission to first decision.
- Days from submission to desk decision.
- Days from sending to review to securing the required number of reviewers.
- Number of invitations sent per completed review.
- Editor and staff minutes per manuscript, by step.
- Time each manuscript spends waiting, against time being worked on, at each stage.

Safety measures

- Rate of false flags for each automated check.
- Proportion of desk rejections overturned on appeal.
- Review quality, using a standard instrument.
- Author and reviewer satisfaction.

Design. Introduce one step at a time, or stagger introduction across journals or sections, so that any change can be attributed. A simple before and after comparison is the minimum. A staggered rollout is better. Turnaround times vary widely and a few long cases skew them, so comparisons should use medians. A single journal over a few months may be too small to detect a change of a day or two. A comparison across many months, with the expected effect stated in advance, is more reliable.

A benchmark for scale. The only measured turnaround effect in the literature is modest. A payment of US\$250 to reviewers shortened review time by about one day and raised completion from 42.2% to 49.8% of invitations [38]. Expectations for AI should be set against results of that size, not against marketing figures.

7. Limits of this paper

- The framework is an argument from indirect evidence. The claim that workflow assistance will reduce turnaround is untested.
- The description of where time is lost in section 1.2 is an inference, not a measured breakdown. Section 8 proposes measuring it.
- Evidence for steps 6, 7 and 8 is absent, and for steps 3 and 4 it is thin.
- The underlying review did not search trial registries, Embase or Scopus, and some sources were available only as abstracts or preprints [1].
- Models are improving quickly. Findings on leniency and manipulation describe the models tested and may not hold for later ones.
- The model in section 8 has not been reviewed by journal editors or tested against a journal's own data.

8. The model and how to validate it

Taken together, the eight steps form a model of the peer review workflow: where manuscripts wait, and what might shorten each wait while every judgement stays with editors and reviewers. Figure 1 shows the current process and the six points where we expect manuscripts to wait. Figure 2 shows the same process with the proposed changes in place, and the table links each bottleneck to its proposed remedy.

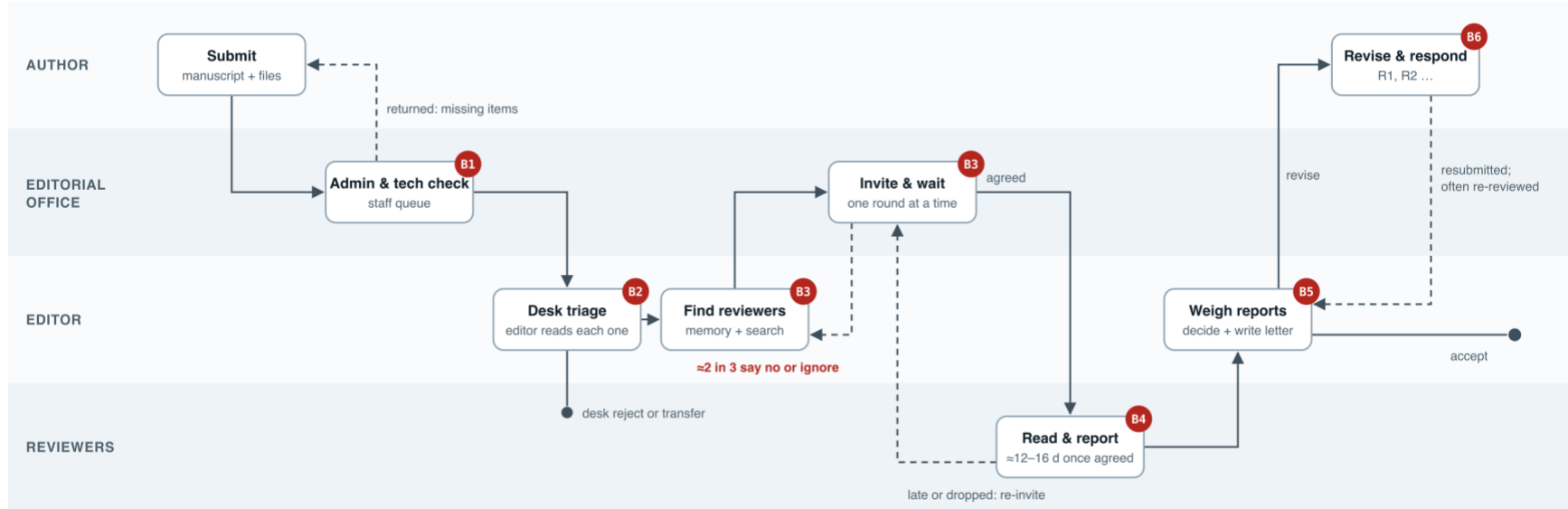


Figure 1. The current process. Red markers show where manuscripts are expected to wait (B1 to B6). Dashed lines are loops: returns to the author, declined invitations, dropped reviews and revision rounds.

Bottleneck	Where manuscripts wait	Proposed remedy	Step
B1. Administrative and technical check	Staff queue. A return to the author restarts the clock.	Automated completeness check. Purely administrative gaps go back to the author with an item list, under rules staff approve.	1
B2. Desk triage	The editor reads every submission, including those that will not go to review.	A brief of checkable facts and check results, with no recommendation.	2, 3
B3. Finding and inviting reviewers	Invitations go out a round at a time. Each decline or silence costs days.	Ranked candidates with conflicts and current load. The editor approves the list once. Invitations run in parallel and lapse on a timer.	4, 6

Bottleneck	Where manuscripts wait	Proposed remedy	Step
B4. Review in progress	Overdue reviews. A dropped review sends the manuscript back to B3.	Reminders, and replacement from the approved list.	6
B5. Decision	The editor reconciles the reports and writes the letter.	Reports side by side, each point linked to its source.	7
B6. Revision rounds	Re-review repeats B3 and B4.	Revision check. Re-review of open points only.	8

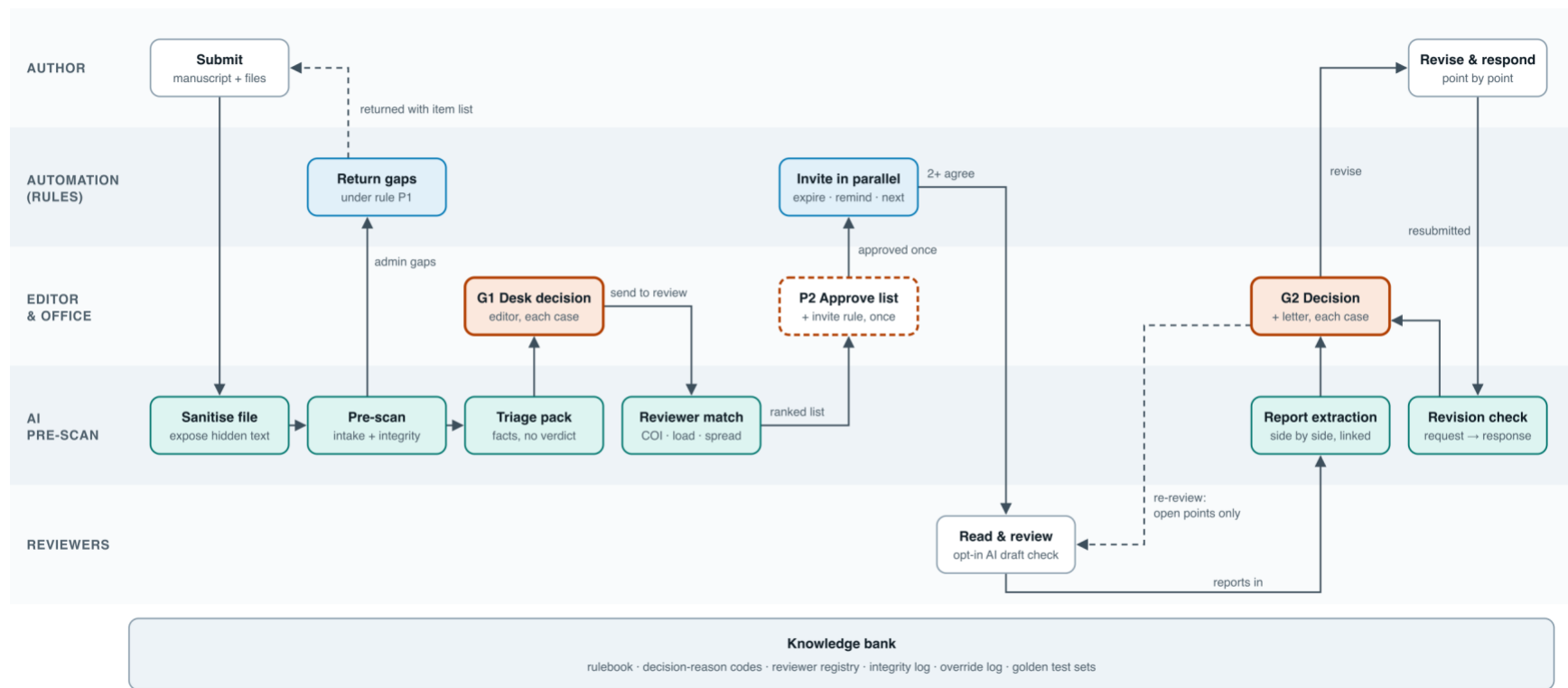


Figure 2. The proposed model, not yet validated. AI prepares and checks (teal). Automation runs routine logistics within rules people approve (blue). People decide at every gate (orange): G1 and G2 are decision gates, decided case by case; P1 and P2 are policy gates, where a person approves a rule or list once.

This model is an unvalidated hypothesis. The bottlenecks are inferred from indirect evidence (section 1.2), and none of the remedies has been tested at a medical journal for its effect on time. It should not be built or deployed on the strength of this paper.

8.1 Two kinds of gate

The model separates two kinds of human gate. At a decision gate, a person decides each case: a desk rejection, action on an integrity flag, the editorial decision and letter, and whether a revision goes back to the reviewers. At a policy gate, a person approves a rule or a list once, and routine, reversible actions then run within it, such as sending the next invitation from a list the editor has approved. Every action is logged and a sample is checked. Judgements about novelty, importance, methods and clinical relevance stay with people, and no AI output is shown for them. Whether policy gates are safe, and acceptable to editors, is itself part of what needs validating.

8.2 Learning from the editorial team

The model assumes that the tools improve only through the people who use them. Each journal would keep a knowledge base of its own rules and experience: its scope and requirements, the reasons its editors give for decisions, a register of reviewers and how they respond to invitations, a log of integrity flags and what they turned out to be, and a record of every AI output an editor overrode, with the reason. A change to any tool would be drafted from that record, tested against past manuscripts with known answers, and released only when a named editor approves it.

8.3 Validating the model

We propose three stages. Each must succeed before the next begins, and all three need a medical journal as a partner.

Stage 1. Confirm the bottlenecks are real. Record the time each manuscript spends at each stage, waiting and being worked on, for at least three months. Interview editors, staff and reviewers about where work stalls. The question is whether the six points in Figure 1 are where the time actually goes, and whether there are others. If they are not, the model is revised before going further.

Stage 2. Confirm the remedies are possible and practical. Test each tool on past manuscripts whose outcomes are known, away from live work, to see whether it is accurate enough to be useful. Work through the gates and the knowledge base with editors and reviewers to see whether they would use them, and whether they fit the journal's systems, policies and confidentiality rules. Remedies that fail either test are dropped.

Stage 3. Run a pilot alongside the current process. If stages 1 and 2 succeed, a pilot system can be built, by the journal, a publisher or a technology partner, and run in parallel with the journal's normal workflow. Running in the background on new submissions, without affecting any decision, it shows how often each tool is right and how much time it could save, at no risk to authors. Some changes, such as inviting reviewers in parallel, only save time when they act on real manuscripts. These can be tested live in one section of the

journal while the rest continues as before, which gives a direct comparison, measured as described in section 6.

Until those stages are complete, the model is a hypothesis for discussion with journals, not a recommendation to build or buy.

9. Conclusion

Peer review is under pressure. Current AI makes a poor reviewer, but it can still help with the work around the review. It can handle preparation, checking and coordination so editors and reviewers have less routine work to do. This means checking that a manuscript is complete, screening it for problems and finding suitable reviewers sooner.

This approach follows current policy and keeps scientific judgement with editors and reviewers. AI is used for tasks where people can spot mistakes and the consequences are limited. We do not know yet whether it will make a journal faster. Publishers need to measure that. Section 8 sets out how, beginning with confirming where the time actually goes.

References

1. Medware Group. AI in medical journal peer review: what the evidence shows. White paper. October 2026.
2. Ben Messaoud K, et al. Analysis of peer reviewers' response to invitations by gender and geographical region. BMJ. 2023. <https://doi.org/10.1136/bmj-2023-075719>
3. Morley CP, et al. Reviewer engagement trends at a journal: cause for concern. PRiMER. 2025. PMID 41531847.
4. Dahl J. Is growth always good news? 2026 article submission surges. The Scholarly Kitchen. 13/05/2026. <https://scholarlykitchen.sspnet.org/2026/05/13/guest-post-is-growth-always-good-news-2026-article-submission-surges/>
5. Aczel B, et al. A billion-dollar donation: estimating the cost of researchers' time spent on peer review. Res Integr Peer Rev. 2021. <https://doi.org/10.1186/s41073-021-00118-2>
6. Perlis RH, et al. Factors associated with author and reviewer declared use of AI in medical journals. Peer Review Congress. 09/2025. <https://peerreviewcongress.org/abstract/factors-associated-with-author-and-reviewer-declared-use-of-ai-in-medical-journals/>
7. Publons. Global State of Peer Review 2018, as summarised by EASE. <https://ease.org.uk/2018/10/2018-global-state-of-peer-review-report-released/>
8. Liu SW, et al. Bot-Tastic: overload burden of unsolicited peer review requests. Acad Emerg Med. 2026. <https://doi.org/10.1111/acem.70260>
9. Amazon Web Services. Building an editorial AI assistant to support peer review. 22/05/2026. <https://aws.amazon.com/blogs/publicsector/building-an-editorial-ai-assistant-to-support-peer-review-with-aws-generative-ai-innovation-center/>

10. Moshirfar M, et al. Current capabilities of large language models as peer reviewers for manuscripts submitted to ophthalmology-related journals. *Medicine (Baltimore)*. 2026. <https://doi.org/10.1097/MD.00000000000048147>
11. Joachim MV, et al. How artificial intelligence differs from humans in peer review. *J Oral Maxillofac Surg*. 2025. <https://doi.org/10.1016/j.joms.2025.03.015>
12. Erturk SM, Durmaz M. Large language models as peer reviewers: prompt sensitivity and model-dependent reproducibility. *Acad Radiol*. 2026. <https://doi.org/10.1016/j.acra.2026.04.046>
13. Tomazini BM, et al. Susceptibility of large language models to hidden nudge injection during simulated medical peer review. *Res Integr Peer Rev*. 2026. <https://doi.org/10.1186/s41073-026-00225-y>
14. Lin Z. Hidden prompts in manuscripts exploit AI-assisted peer review. *arXiv* 2507.06185. 2025. <https://arxiv.org/abs/2507.06185>
15. ICMJE. Use of AI by reviewers. Recommendations, updated January 2026. <https://www.icmje.org/recommendations/browse/artificial-intelligence/ai-use-by-reviewers.html>
16. Elsevier. Generative AI policies for journals. Updated June 2026. <https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals>
17. Springer Nature. Using AI in peer review. Accessed 02/10/2026. <https://www.springernature.com/gp/policies/editorial-policies/using-ai-in-peer-review>
18. IOP Publishing. AI and Peer Review 2025. 09/2025. <https://iopublishing.org/ai-and-peer-review-2025/>
19. Di Girolamo N, et al. Pragmatic assessment of different AI large language models for extraction of CONSORT items. *Peer Review Congress*. 09/2025. <https://peerreviewcongress.org/abstract/pragmatic-assessment-of-different-ai-large-language-models-for-extraction-of-consort-items-from-randomized-controlled-trials-before-peer-review/>
20. Wrightson JG, et al. GPT for RCTs? *BMJ Open*. 2025. <https://doi.org/10.1136/bmjopen-2024-088735>
21. Chen D, et al. AutoReporter. *J Am Med Inform Assoc*. 2026. <https://doi.org/10.1093/jamia/ocaf223>
22. Tsybulnik DY, et al. Variability among large language models in assessing CONSORT compliance. *PLoS One*. 2026. <https://doi.org/10.1371/journal.pone.0358873>
23. Kataoka Y, et al. Large language models for automated PRISMA 2020 adherence checking. *Int J Med Inform*. 2026. <https://doi.org/10.1016/j.ijmedinf.2026.106561>
24. Speich B, et al. Reminding peer reviewers of reporting guideline items to improve completeness in published articles. *JAMA Netw Open*. 2023. <https://doi.org/10.1001/jamanetworkopen.2023.17651>
25. Struthers C, et al. GoodReports randomised trial. *BMC Med Res Methodol*. 2025. <https://doi.org/10.1186/s12874-025-02518-0>
26. Chaturvedi AP, et al. ASM incorporates Imagetwin to address image duplication. *mBio*. 2025. <https://doi.org/10.1128/mbio.01990-25>

27. Evanko D. Use of an artificial intelligence-based tool for detecting image duplication prior to manuscript acceptance. Peer Review Congress. 2022.
<https://peerreviewcongress.org/abstract/use-of-an-artificial-intelligence-based-tool-for-detecting-image-duplication-prior-to-manuscript-acceptance>
28. Ying X, et al. Using GPT to identify changes in clinical trial outcomes registered on ClinicalTrials.gov. Peer Review Congress. 09/2025.
<https://peerreviewcongress.org/abstract/using-gpt-to-identify-changes-in-clinical-trial-outcomes-registered-on-clinicaltrials-gov/>
29. Liang W, et al. GPT detectors are biased against non-native English writers. Patterns. 2023. <https://doi.org/10.1016/j.patter.2023.100779>
30. Patel R, et al. Artificial intelligence cannot replace peer reviewers but may help editors triage. Am J Sports Med. 2026.
<https://doi.org/10.1177/03635465261463006>
31. Lupon E, et al. Can large language models provide high-quality desk review decisions in an orthopaedic surgery journal? Orthop Traumatol Surg Res. 2026.
<https://doi.org/10.1016/j.otsr.2026.104801>
32. Teixeira AL. AI in peer review: can artificial intelligence be an ally in reducing gender and geographical gaps in peer review? Res Integr Peer Rev. 2025.
<https://doi.org/10.1186/s41073-025-00182-y>
33. Thakkar N, et al. Can LLM feedback enhance review quality? A randomized study of 20K reviews at ICLR 2025. arXiv 2504.09737. 2025.
<https://arxiv.org/abs/2504.09737>
34. Hopkins BS, et al. Application of large language and artificial intelligence modeling in the prediction of peer-review outcomes. J Neurosurg. 2026.
<https://doi.org/10.3171/2025.8.JNS242667>
35. Jarrett P, et al. Frontier language models and OCR preprocessing against invisible text injection. JAMA Netw Open. 2026.
<https://doi.org/10.1001/jamanetworkopen.2026.17356>
36. COPE. Artificial intelligence (AI) in decision making. Topic discussion, updated 11/12/2024. <https://publicationethics.org/topic-discussions/artificial-intelligence-ai-decision-making>
37. Waymouth E, et al. Strategic insights into editor engagement with AI-assisted tools. Peer Review Congress. 09/2025.
<https://peerreviewcongress.org/abstract/strategic-insights-into-editor-engagement-with-ai-assisted-tools-based-on-survey-and-data-analysis-of-ai-assisted-ethics-checks/>
38. Cotton CS, et al. Effect of monetary incentives on peer review acceptance and completion. Crit Care Med. 2025.
<https://doi.org/10.1097/CCM.0000000000006637>
39. Sage. Artificial intelligence policy. Accessed 02/10/2026.
<https://www.sagepub.com/journals/publication-ethics-policies/artificial-intelligence-policy>
40. Frontiers. Policies and publication ethics. Accessed 02/10/2026.
<https://www.frontiersin.org/guidelines/policies-and-publication-ethics>