OPEN ACCESS

Artificial Intelligence COMMENTARY

From AI formal proofs to reusable mathematical ideas

Jiachuan Wang[1,*]

[1]Institute of Library, Information and Media Science, University of Tsukuba, Tsukuba 305-8550, Japan

*Correspondence: wangjc@slis.tsukuba.ac.jp

Received: October 19, 2026; Accepted: November 5, 2026; Published Online: November 5, 2026; https://doi.org/10.59717/j.xinn-inform.2026.100077

© 2026 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Citation: Wang J. (2026). From AI formal proofs to reusable mathematical ideas. *The Innovation Informatics* **2**:100077.

AI-assisted proof search is creating new opportunities for mathematical research. AlphaProof Nexus offers a concrete example of systems that generate candidate arguments and develop formal proofs. Formal proofs are deductions written in a precise language so that a computer can check whether the conclusion follows from stated starting points.[1] Yet a checked result raises a further question. What can other researchers learn from the argument? A mathematical contribution also matters when peers can grasp its central ideas, see the conditions under which they work, and develop them further. Connecting formal results to that understanding should be a research objective in its own right. The challenge is to explain what makes an argument work, at a depth suited to those who will use it.

Nexus uses Lean, a proof assistant that checks whether a formal conclusion follows from specified definitions, assumptions and accepted starting principles. Its language-model programs propose and revise proof drafts, using Lean's feedback to guide the search. A successful search ends with a complete proof, with no unfinished steps.[1] In an evaluation on 353 targets associated with unsolved Erdős problems and already formalized in Lean at the time, the complete search system proved nine problem, subproblem or variant targets.[1] These results illustrate progress in turning mathematical proof targets into checkable arguments.

VERIFICATION AND MATHEMATICAL UNDERSTANDING

However, mathematical understanding asks why a particular construction or application of an earlier result advances the argument. A reader may identify a lemma, an auxiliary result supporting later steps, yet still need to see why using it here overcomes the main difficulty. Formal material already records definitions and which earlier results each step uses. An explanation of the presented argument should make the mathematical purpose of those relationships visible. Making that argument intelligible allows other researchers to discuss its ideas and consider where they might be useful.

The joint declaration published on 11 September 2026, with Terence Tao among its initial signatories, emphasizes conceptual understanding and the absorption of ideas through discussion and simplification. Scott Aaronson's report describes Dana Moshkovitz reading an AI-assisted mathematical manuscript. Alongside expectations of AI's future value, it reports her difficulty identifying why earlier results applied and assembling scattered properties of a key construction. These reported difficulties point to concrete organizational needs, including bringing a construction's relevant properties together and explaining why earlier results apply.

Existing approaches differ in their organization of proof explanations, their support for readers and their approach to evaluation. Hattori et al. generate step descriptions and recursively summarize formal proofs along their

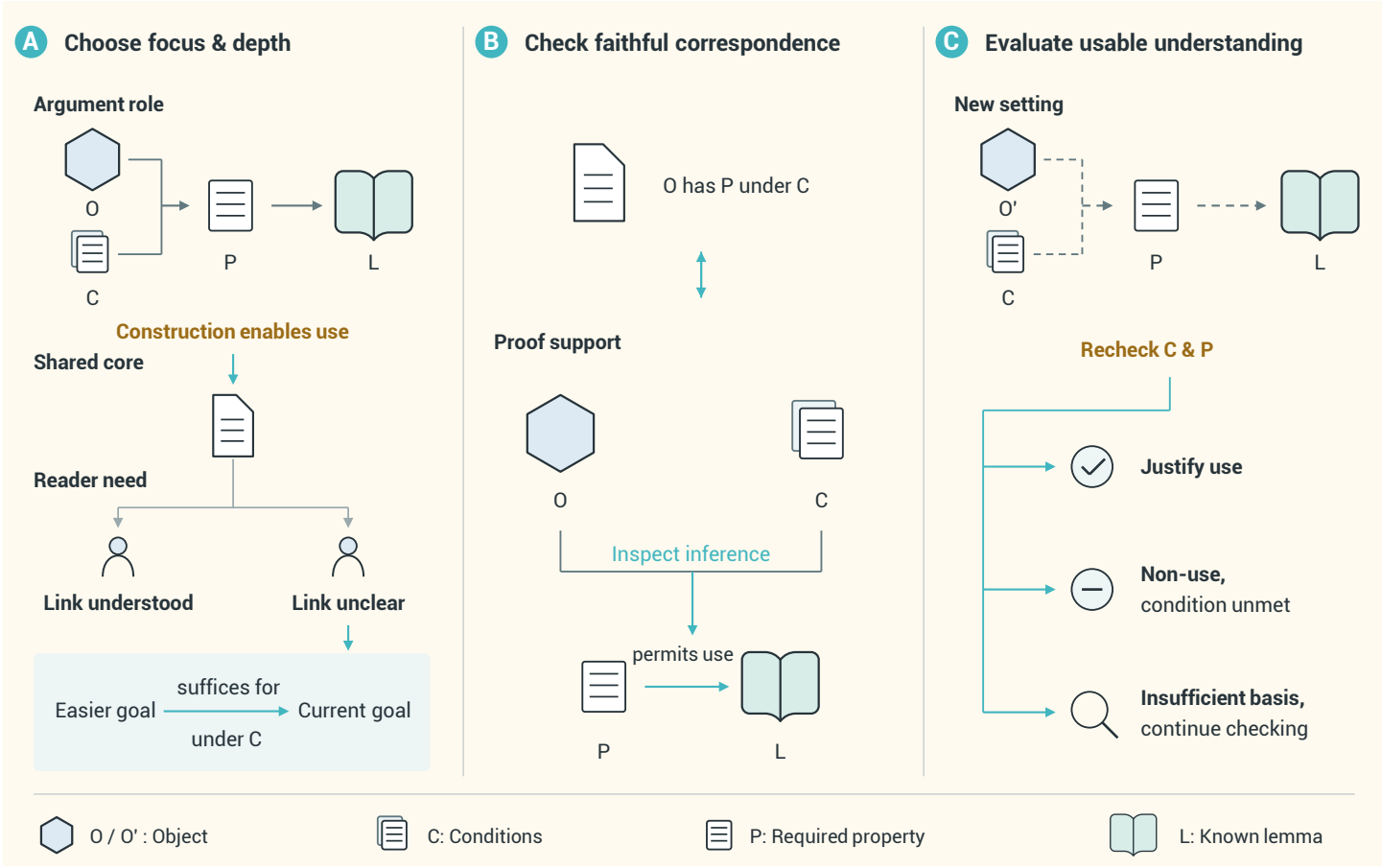


Figure 1. Three tasks for explaining mathematical arguments (A) Choose focus by argument role and support by reader need. Retain the shared core, adding support when the lemma's goal link is unclear. (B) Compare explanatory claims with formal objects, conditions, properties, supporting inferences and lemma application. (C) Recheck the same lemma in a new setting and justify use or non-use, or defer judgment when evidence is insufficient and continue checking.

dependencies. Their evaluations include step correctness, sufficiency and necessity, and coverage of summary key points.[2] HumanProof organizes interactive proof construction through smaller tasks supporting larger goals, with natural-language and Lean 3 outputs. It proposes author-marked trivial steps to vary detail by reader level. Its evaluation reports mathematical readers' ratings of understanding and confidence in correctness, with verbal comments.[3] The perspective here treats contribution role and reader prerequisites as separate inputs to explanation choices. It links those choices to semantic checking and evaluation of justified use in a new setting.

A BRIDGE DESIGNED FOR RESEARCH USE

We propose three connected tasks to study how proof explanations support further mathematical work (Figure 1). First, identify the ideas and reasoning steps that carry the argument, choosing what to foreground rather than reproducing every detail. Second, preserve their meaning and conditions while making the supporting formal statements and proof steps easy to locate. Third, examine whether readers acquire understanding they can use appropriately in further mathematical work.

For the first task, the depth of explanation is closely tied to readers' needs. Specifically, a known lemma can turn a difficult goal into a more manageable one when certain conditions hold. The argument's key move may be to construct an object with the properties needed to apply that lemma. A reader familiar with it needs to see why the construction works and meets those conditions. A related-domain reader unfamiliar with the lemma also needs its meaning and the reasoning connecting it to the goal. The construction plays the same role in both accounts, while the supporting explanation changes.

A reader's familiarity with a result and the result's role in the argument call for separate judgments. A result can supply background and enable a decisive step, while unfamiliar material can remain routine support. A step's role in the current argument and its historical novelty call for separate judgments. A future system could take a checked proof, its definitions and assumptions, and a reader's stated prerequisites or questions as inputs. For the construction above, it could group the object's required properties and show how they permit the lemma application. A request for the missing connection to the goal could prompt the system to expand the corresponding inference. A challenge to the choice of focus could prompt it to reconsider whether to foreground the construction. When expanding the inference or revising the focus, the system could record the reason for the change while retaining the argument's objects, adopted conditions and reasoning.

Selecting the key ideas raises a second question. How can a reorganized explanation remain faithful to the proof? For each explanatory claim, a future system could display the mathematical object, adopted assumptions and supporting proof fragment beside the explanation. In the construction example, it could list the properties required by the lemma and the inferences establishing them. A reader could flag an omitted condition or an unjustified connection. The system could retain the claim's formal location and record the comparison and any unresolved issue. The reader's comparison would check whether the correspondence preserves meaning and reasoning, complementing verification of the formal deductions.

The third task is to evaluate whether readers gain understanding they can use appropriately, even when an explanation is faithful. An evaluation could begin by asking readers to explain why the original construction advanced the proof goal and what conditions made the lemma applicable. It could then ask how a change to one of those conditions would affect that route. Such a change might invalidate the route without disproving the conclusion, since another route could still work. Where a proof offers a suitable method for further use, the system could present a nearby problem using a new object and ask whether the same lemma applies. Readers would identify the required properties, check the new assumptions and justify use or non-use. When information is insufficient, they could defer judgment and identify what still needs checking. Their recorded explanations and cited conditions would provide observable outcomes for mathematical assessment. Comparisons would require readers with comparable prior knowledge, suitable task difficulty and explicit judgments of whether an application is mathematically justified. These tasks would complement existing information-level assessments and user feedback.[2,3] Their aim is to obtain evidence of usable understanding, including its limits, rather than rely on confidence based only on fluent wording.

COMPLEMENTARY MATHEMATICAL DELIVERABLES

The correspondence between a proof and its explanation must also be maintained when either is revised. If a supporting formal assumption or inference changes, the affected explanation should be checked to determine whether its claims still reflect the objects, conditions and reasoning in the revised argument, even if its cited locations remain accessible. Reordering only the explanation's narrative leaves the formal proof unchanged, while still requiring references to remain attached to the claims they support.

Contribution role identifies where an argument advances, while reader background determines what supporting explanation is needed to follow that advance. The explanation must retain the argument's conditions and remain open to inspection against its formal grounds. The aim is to help peers question an idea's application and pursue it in further research, beyond making the proof look readable.

Historical experience also suggests that AI capability often advances without waiting for full human understanding of every result. Go offers an encouraging analogy.[4] Even when AI achieves superhuman performance in that task, people still have room to learn. If a decisive breakthrough were achieved in the correctness of trusted checkers, the volume of machine-checked mathematics would inevitably grow explosively. As AI researchers, we have already seen similar developments in numerous fields. The good news, however, is that even if AI-generated content expands rapidly without being fully understood, human mathematical understanding of AI-generated results will remain an important research focus and become increasingly important as those results accumulate.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author used Codex in order to assist with logic smoothness refinement and grammar checking. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

REFERENCES

1. Tsoukalas G., Kovsharov A., Shirobokov S., et al. (2026). Advancing mathematics research with AI-driven formal proof search. *Science* **394**:234–239. DOI:10.1126/science.aej2213
2. Hattori S., Matsuzaki T. and Fujiwara M. (2025). Natural language translation of formal proofs through informalization of proof steps and recursive summarization along proof structure. In Proceedings of the 18th International Natural Language Generation Conference. **2025**: 376-389. Association for Computational Linguistics. https://aclanthology.org/2025.inlg-main.23/
3. Ayers E. (2021). A tool for producing verified, explainable proofs. [PhD thesis, University of Cambridge]. DOI:10.17863/CAM.81869
4. Shin M., Kim J., van Opheusden B., et al. (2023). Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proc. Natl. Acad. Sci. USA* **120**:e2214840120. DOI:10.1073/pnas.2214840120

FUNDING AND ACKNOWLEDGMENTS

Jiachuan Wang's work is supported in part by Japan Science and Technology Agency (JST) CREST (JPMJCR22M2). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DECLARATION OF INTERESTS

The author declares no competing interests.

DATA AND CODE AVAILABILITY

The public joint declaration A Severe Misalignment of AI in Mathematics is reproduced on Terence Tao's blog (https://terrytao.wordpress.com/2026/09/11/a-severe-misalignment-of-ai-in-mathematics/). The other resources discussed are Scott Aaronson's blog report The Mathocalypse (https://scottaaronson.blog/?p=10169) and the public AlphaProof Nexus Results repository (https://github.com/google-deepmind/alphaproof-nexus-results, accessed 14 October 2026).