THE MAZE · ESSAY

The Address of a Proof

The Leiden Declaration names four dilemmas that artificial intelligence poses to mathematics. All four are provenance problems — and provenance can be architectural rather than procedural. A working system, with measurements.

Hans Konstapel & Claude August 2026 maze.swarp.nl · constable.blog

I

Four dilemmas, one problem

In June 2026 an international group of researchers presented the *Leiden Declaration on Artificial Intelligence and Mathematics*, a call to action that has since been discussed in *Science*, *Nature*, the *New York Times* and *Le Monde* [1]. The Declaration asks four questions that mathematicians can no longer treat as hypothetical. When mathematical research is carried out by a proprietary AI model to which academic researchers have no access, who is responsible for its errors? Who receives credit when it is correct? How can we establish whether an AI-generated proof is genuinely new, or a clever reformulation of existing work without attribution? And what becomes of a discipline whose central instrument is closed to the discipline itself?

These questions look heterogeneous — one about liability, one about honor, one about novelty, one about access. We will argue that they are four faces of a single underlying problem: mathematics has never needed an infrastructure for **provenance**, because for twenty-five centuries the provenance of a proof was a person. The proof arrived attached to a mathematician, a manuscript, a journal, a date. Attribution, responsibility, novelty and access were all secured by the same social fact: someone signed. Machine-generated mathematics dissolves that fact. The signature detaches from the content, and with it every guarantee that hung on it.

The Declaration responds with recommendations — to individual mathematicians, to organizations and funders, to policymakers, to commercial AI companies. This essay argues that recommendations are the wrong instrument for this class of problem, describes an architectural alternative that already runs in public, and reports measurements from it. The alternative was not designed for the Declaration's dilemmas; it was designed from an older theoretical program [6, 7]. That it answers them is a result, not an intention — and results, unlike intentions, can be checked.

II

The wrong horse

The Leiden Declaration places its stake on two horses: *governance* — norms, codes of conduct, appeals to commercial laboratories — and, in the research program of several of its signatories, *formal verification*: the digital checking of proofs by systems such as Lean and Coq. The second horse is a magnificent animal, and Section V of this essay is devoted to what it can already do. But neither horse, nor both together, reaches the finish line the Declaration itself has drawn. The argument is short and, we believe, exact.

Governance regulates behavior where architecture can remove the opportunity. A norm asking AI companies to disclose training data, label machine-generated proofs and preserve attribution is a request addressed to the very parties whose commercial interest runs the other way — the Declaration itself notes that researchers lack access to the models in question. The history of scholarly infrastructure teaches a consistent lesson here. Citation ethics did not stop plagiarism; version control and digital object identifiers made priority disputes decidable. Email etiquette did not stop forgery; cryptographic signatures made it detectable. In each case the durable solution was not a better norm but a substrate on which the violation is either impossible or visible by construction. Declarations are how a community announces that it has a problem. They are rarely how it solves one.

Verification is an oracle without a library. This is the deeper point, and it applies to the strongest, most admirable part of the program the Leiden lecture announces — the project of checking complicated proofs digitally. A proof checker answers exactly one question: *is this derivation correct?* It is silent on the three questions the Declaration actually asks. A machine-checked proof can be a correct reformulation of known work, correctly stolen. It can be correct and unattributable. It can be correct and locked inside a corporate system. Correctness is a property of a derivation; novelty, credit and access are properties of a derivation's *place* in the body of knowledge — and formal verification has no concept of place. Betting on verification alone is betting that the crisis is about wrong proofs. It is not. Machine mathematics will be overwhelmingly *correct*. The crisis is about proofs that are correct, placeless, and ownerless.

> What mathematics is missing is not a better judge of derivations. It is a **land registry**: a public space in which every result — established, contested, or still missing — has a computable place, and every occupant of a place has a receipt.

III

A self-addressing knowledge net

The MAZE (maze.swarp.nl) is a public, working implementation of such a registry, built on a theoretical program developed over two decades on constable.blog and summarized in *Paths to the Knot* [6]. Its principles can be stated in a paragraph each; none of them is speculative, since all of them run in production.

Addresses are computed from content. The MAZE organizes knowledge in a balanced-ternary address space. An address is a finite route of moves, each +1, 0 or −1, beginning at a single origin — the *first distinction*, in Spencer-Brown's sense [5]. A piece of material does not get *assigned* a place by a librarian; its place is *calculated* from its content, the way Gödel numbering assigns every formula an integer by construction rather than by convention [4]. Value and depth (`v@d`) identify a winding uniquely; depth is refinement. The decisive consequence: two formulations of the same content land at, or measurably near, the same address. Placement is idempotent. The address space is complete in the same sense as Borges' library [3] — every possible statement, including every future proof, has a computable place before anyone writes it — with one difference that changes everything: in the MAZE the catalog is the shelf. The address is derivable from the item, so nothing can be lost, and nothing can hide.

Nothing lives in the net without a receipt. Every piece of material carries a source identifier, a provenance trail, and is admitted under a cryptographic source key registered at first placement. The right to delete hangs on the same key. We call this the *receipt principle*: the net's answer to "who put this here, when, from where?" is a column, not a committee.

The net never claims more than it holds. The conversational layer of the MAZE codes a user's question to an address and reports what lives there — names, sources, neighboring addresses — and nothing else. When a question lands on an empty region, the net says so. This *honesty rule* is enforced in the retrieval pipeline, not requested in a prompt: the language model that phrases answers is forbidden, by the shape of the data it receives, from asserting content that is not on an address. The contrast with a large language model is categorical, not gradual: an LLM stores knowledge in weights, where provenance is destroyed by construction; the MAZE stores knowledge at addresses, where provenance is preserved by construction.

The unknown is a first-class object. Because the address space exists independently of its occupation, an *empty* address is a well-formed thing: it has a depth, neighbors, an opposite, and a route leading to it. The MAZE runs a daily reasoning loop that walks inhabited windings, closes combinations, and records *vacancies* — places that the structure of the net implies should be inhabited but are not. Everywhere else in the scientific world, the unknown is formless. Here it has coordinates.

As of late August 2026 the public net holds 6,691 inhabited windings carrying 3.6 million pieces of material from receipted sources — the English Wikipedia (round one seeded the 65 fields of knowledge of *Paths to the Knot*), the Metropolitan Museum of Art, three musical corpora, and constable.blog — behind a memory-palace interface in which every address is a shareable URL.

IV

Measurements: how empty the space of knowledge is

A registry of places permits a measurement no other representation of knowledge can make: the occupancy of the possibility space, ring by ring. We report the census of 28 August 2026, taken

directly from the production database. At depth d the space offers 3^d addresses; the census counts how many are inhabited.

```
CENSUS · INHABITED WINDINGS PER RING · 28-08-2026

ring  4     40 of 81             ·  49.4%
ring  5    122 of 243            ·  50.2%
ring  8  1,299 of 6,561          ·  19.8%
ring 12  1,673 of 531,441        ·  0.31%
ring 16  1,497 of 43,046,721     ·  0.0035%
ring 20    233 of 3,486,784,401  ·  0.0000067%
ring 24    516 of 282,429,536,481  ·  0.00000018%
```

Two readings of this table matter for the Declaration's world. The first is sobering: sixty millennia of the symbolic species [9], compressed through its largest encyclopedia, occupy a filament of the possibility space that thins by orders of magnitude with every step of refinement. Knowledge is not a territory; it is a path system through an almost entirely dark country. The second reading is the operative one: **near the origin, the space is half full.** At the rings where whole *disciplines* live — rings 4 and 5, the resolution of the 65 seeded fields — 41 and 121 places respectively stand empty. These are not missing facts. They are candidate missing *sciences*: coherent refinement-systems of the language layer for which history has not yet found the right first distinction. The historical record says such discoveries happen and are always sudden: probability theory did not exist before 1654, though chance had run through every human life for sixty thousand years; information theory did not exist before 1948 [10]; computability not before 1936. Each was a region of the language layer that became a science the moment someone drew the right distinction. Nothing distinguishes 1948 as the final such year. The MAZE is, to our knowledge, the first instrument in which the candidates for the next such year form a printable list.

V

What proof checkers can already do — and the register they are missing

Formal verification has crossed, in the last two decades, from curiosity to industrial capacity, and any honest essay must say so emphatically. The four-color theorem was machine-verified in full by Gonthier in 2005 [11], the Feit–Thompson odd-order theorem — 255 pages of the classification era — in 2012. Hales' Flyspeck project verified the Kepler conjecture in 2014, closing a referee process that human mathematics had been unable to finish [12]. The Lean community's mathlib now carries on the order of a hundred thousand theorems maintained by hundreds of contributors [13]. In 2022 the Liquid Tensor Experiment verified, at Peter Scholze's own request, the central theorem of his condensed mathematics — Scholze publicly stated that the formalization found and repaired genuine gaps in his understanding of his own proof [14]. Since 2024 a project led by Kevin Buzzard has been formalizing Fermat's Last Theorem [15], and

machine provers of the AlphaProof and AlphaGeometry line have reached the level of International Mathematical Olympiad medalists on competition problems [16, 17]. The oracle exists. It is excellent and getting better, and the Vici-funded program of digitally checking difficult number-theoretic proofs announced in the Leiden lecture [2] is exactly the kind of work that makes it better still.

What the oracle lacks is a library — and this is where the two programs, far from competing, snap together like halves of a mold. The design, implementable today on the running system:

A public proof register on the knowledge net. A source `proof:lean` is admitted to the MAZE under one additional rule: its source key is a machine-checkable certificate — the proof term itself, checked by the verifier's small trusted kernel on admission. Then, by construction: *admission is verification* (nothing enters the register unchecked); *the address is the novelty test* (the place of a result is computed from its content, so a reformulated theorem lands on an inhabited winding, and "is this new?" becomes a lookup — the occupant's receipt names who was first, and when); *the receipt is the credit mechanism* (a proof found by a machine is registered under `proof:ai:<model>`, a human proof under its author; honor follows the receipt mechanically, and so does responsibility, since the deletion right hangs on the same key); and *the vacancies are the problem list* (machine provers are pointed at addresses where the structure of the net implies a result should live — and whatever they find arrives with its provenance attached, because there is no other way in). Each of the Declaration's four dilemmas is answered not by a recommendation but by a property of the substrate. The fourth — access — is answered by the register being public: a closed model can keep its weights, but a proof that is not in the register has, in the precise sense of this architecture, no address: no priority, no credit, no standing.

VI

Honest limits

Three limits must be stated plainly, because the argument above is architectural and architecture is cheap to overclaim. First, the MAZE's current coding of content to addresses — a four-ring semantic coding plus a name bridge through Wikipedia's canonical titles — is far too coarse to fingerprint mathematical proofs; the register design requires a mathematical signature (plausibly: the normalized statement in the verifier's own term language, which is already canonical), and building that mapping is real work, not detail. Second, the novelty test detects *placement* collisions, not semantic equivalence in full generality — deciding whether two proofs are "the same" is undecidable in the limit, and the register's honest claim is weaker and more useful: *this statement's place was already occupied, by this receipt, since this date*. Third, the occupancy census measures the net's sources, not the world; rings the encyclopedia covers thinly will read as emptier than they are. None of these limits touches the central claim, which is not that the MAZE is finished, but that provenance-by-construction exists, runs, and measures — while provenance-by-declaration has, as its own text concedes, only requests to make.

VII

Conclusion

The Leiden Declaration is right about the problem and, we have argued, mistaken about the instrument. Its four dilemmas — responsibility, credit, novelty, access — are one dilemma: mathematics is losing the social fact of the signature and has no infrastructural replacement. Codes of conduct cannot replace it, because they petition exactly the parties they would need to bind. Verification alone cannot replace it, because a checker certifies derivations, and the crisis is about places — an oracle without a library. The replacement is a land registry for results: addresses computed from content, receipts required for admission, honesty enforced by structure, and the unknown itself given coordinates. Such a registry is not a proposal. A general-knowledge instance is public at maze.swarp.nl, with 3.6 million receipted items, a census showing exactly how empty the space of knowledge still is, and a daily loop that lists the vacancies. Extending it with a proof source whose admission ticket is a Lean certificate is the natural next construction — one in which the formal-verification program of the Declaration's own signatories is not the wrong horse after all, but the missing half of the right one.

Postscript: the self-test

After this essay was finished, we ran it through the machinery it describes. The net's field coder assigned it the signature `0-0-0-+-00++++-0+-0` and the address `-144719454@19`, its dominant layers being consciousness and knowledge, language and script, and information — where a text about knowledge infrastructure belongs. The address, and its entire neighbourhood on ring 19, is empty: at the coder's resolution, this essay occupies a vacancy. Under its own receipt principle it has not been placed — admission awaits publication, whose URL will be its receipt; it will then become the first inhabitant of the place it computed for itself.

The test also produced the more valuable result: a failure. Two further layers fired — art, and cities and construction — because the literal lexicon read the essay's central metaphors, *"provenance can be architectural"* and *"by construction"*, as subject matter. Section VI claimed that the current coding is too coarse for texts of this kind; the essay has now demonstrated that claim on itself. We record the bug rather than hiding it: the first artifact this registry design ever tested was the design's own description, and the test found a defect in the tester. That is what an instrument is for.

Annotated references

1. Leiden Declaration on Artificial Intelligence and Mathematics, presented June 2026.
   The call to action this essay engages: recommendations addressed to mathematicians, organizations and funders, policymakers, and commercial AI companies, covering responsibility, credit, novelty and access. Reported on in Science, Nature, the New York Times and Le Monde. Our engagement is with the Declaration's strategy, which we argue is procedural where the problem is infrastructural.

2. Dahmen, S., lecture announcement "De Leiden Declaration: AI zet kernwaarden van de wiskunde onder druk", September 2026; Vici research program on polynomial equations in integers, including the digital checking of complicated mathematical proofs.
   The occasion for this essay. The verification program described there is, in our reading, necessary and admirable — and incomplete in a precisely statable way (Section II): correctness is a property of derivations, while the Declaration's dilemmas are properties of places.

3. Borges, J.L., "La biblioteca de Babel", in *El jardín de senderos que se bifurcan*, 1941.
   The canonical image of a complete but useless possibility space: every book exists, none can be found. The MAZE is the Library with the one repair that matters — the address of a text is computable from the text, so the catalog and the shelf coincide.

4. Gödel, K., "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik* 38, 1931, 173–198.
   Gödel numbering is the ancestral form of self-addressing: every formula receives a place by construction, not convention. The MAZE generalizes the move from formulas to knowledge, on a balanced-ternary route space rather than the integers.

5. Spencer-Brown, G., *Laws of Form*, Allen & Unwin, London, 1969.
   The first distinction — the act that creates a marked and an unmarked state — is the origin of the MAZE's address space: the apex of the memory palace, from which every ring refines by three.

6. Konstapel, H., *Paths to the Knot* and the self-addressing knowledge net, constable.blog, 1993–2026.
   The theoretical program implemented by the MAZE: balanced-ternary addressing from the first distinction, the 65 fields of knowledge used to seed round one, the receipt discipline, and the vacuum-side bookkeeping in which every opened path must close.

7. Rowlands, P., *Zero to Infinity: The Foundations of Physics*, World Scientific, 2007.
   The nilpotent universal rewrite system: from zero, each rewrite step opens new structure under the single constraint that the total remains zero. The mechanism "working since the beginning, regularly opening new paths, however small" — and, in our architecture, the reason balanced ternary (+1/0/−1: creation, rest, annihilation) is the right alphabet for an address space that must always be able to close.

8. Hurwitz, A., "Über die Composition der quadratischen Formen von beliebig vielen Variabeln", *Nachrichten Göttingen*, 1898, 309–316.
   The normed division algebras end at the octonions: the alphabet of the rewrite ladder is finite. The texts written in it are not — which is why the mechanism that produced mathematics has no theorem declaring it has produced its last science.

9. Henshilwood, C.S. et al., "Emergence of modern human behavior: Middle Stone Age engravings from South Africa", *Science* 295, 2002, 1278–1280.
   Evidence for symbolic behavior beyond 70,000 years ago. The essay's "language layer, filled for sixty millennia" rests on this literature; mathematics, at roughly five millennia, is a young refinement of that layer — and the census of Section IV shows how much of the layer's refinement space remains untouched.

10. Shannon, C.E., "A Mathematical Theory of Communication", *Bell System Technical Journal* 27, 1948, 379–423, 623–656.
    The modern type specimen of a sudden science: information, used by the species for sixty millennia, formalizable within months once the right distinction was drawn. Cited as evidence that the list of sciences is open, and that new entries arrive abruptly.

11. Gonthier, G., "Formal Proof — The Four-Color Theorem", *Notices of the AMS* 55(11), 2008, 1382–1393.
The proof that machine verification can absorb a theorem whose human proof was itself computer-assisted and contested. Beginning of the modern verification era invoked in Section V.

12. Hales, T. et al., "A Formal Proof of the Kepler Conjecture", *Forum of Mathematics, Pi* 5, 2017, e2.
Flyspeck: verification finishing what refereeing could not. Demonstrates the oracle's strength — and its silence on novelty, credit and place, which no part of the Flyspeck infrastructure addresses or was meant to address.

13. The mathlib Community, "The Lean Mathematical Library", *Proceedings of CPP 2020*, ACM, 2020, 367–381.
The largest living formal library. Notably, its own hardest recurring governance question — what is already in the library, and under whose name — is a provenance question, answered today by search and convention rather than by construction.

14. Castelvecchi, D., "Mathematicians welcome computer-assisted proof in 'grand unification' theory", *Nature* 595, 2021, 18–19; and the Liquid Tensor Experiment completion, 2022.
Scholze's condensed-mathematics theorem, verified at the author's own request; the formalization exposed real gaps. The strongest existing evidence that verification changes mathematical practice — within its own scope, which is correctness.

15. Buzzard, K., "The Fermat's Last Theorem Project", ongoing since 2024 (Imperial College London / Lean community).
Formalization moving from finished mathematics to frontier mathematics. When it completes, the question "what exactly did this add beyond Wiles–Taylor?" will be a provenance question — precisely the kind the register of Section V answers by lookup.

16. Trinh, T. et al., "Solving olympiad geometry without human demonstrations", *Nature* 625, 2024, 476–482.
AlphaGeometry: machine-generated proofs at olympiad level, in a domain with full formal checking. The first population of artifacts that is correct, abundant, and ownerless — the exact object the Declaration worries about and the register is designed to house.

17. Google DeepMind, "AI achieves silver-medal standard solving International Mathematical Olympiad problems" (AlphaProof), July 2024.
Machine proving reaching medal level under closed development — the access dilemma in its sharpest existing form, and the reason Section V's register makes public admission, not open weights, the unit of standing.
