

Fundamental Mathematics in the Age of AI

The Residue, the Journey, and the Ecology

Benjamin COLLAS

Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606-8502, Japan

bcollas@kurims.kyoto-u.ac.jp · August 13, 2026

Abstract. Large language models have begun refuting long-standing conjectures and, for a few thousand dollars of tokens, solving long-open problems (OpenAI, August 2026). The introspection this has prompted about the future of mathematical discovery is overdue, and the anxiety accompanying it legitimate – but both are attached to the wrong loss.

What machines now produce is the countable part of mathematics – theorems, proofs, refutations – which was always the *residue* of the work, not its product. The product is human understanding: not a stock of results but a collective, hard-won way of deciphering the world and acting upon it^[A]. The two are arcs of a single loop – looking produces the residue; taking it up again, one journey at a time, is what rebuilds the shared understanding. Machines are strong on the countable arc, absent from the one that feeds it. The peril is to leave the loop open. AI did not create the confusion between residue and product; it has called a bluff long on the books, driving the cost of the residue towards zero and making the scarce thing visible at last.

The pressing questions are therefore institutional: who can check the claims of AI companies, what the work becomes for the next generation of researchers, and whether the one thing that cannot be mass-produced – the journey that nourish a shared understanding – continues to be funded. Mathematics, we argue, is uniquely placed among the sciences on the first question – a proof answers to no one’s permission – and uniquely exposed on the last: the journey has never had a price our institution knew how to pay. The decision is ours.

MSC (2020): 01A80 (Primary) 68T01, 68V20, 00A35 (Secondary)

Forewords. *These notes set down a reflection of some three years’ standing. Their writing was occasioned by a conversation with Parmy OLSON (Bloomberg) at the end of July 2026, and occupied the 27th and the 31st; they can be read as a companion paper to [OLS26]. The OpenAI announcement of 1 August arrived as they were being finished, and, to my surprise, fitted the argument without altering it. They draw on personal experience as a researcher, and on exchanges with mathematical communities, private companies, and representatives of governing agencies. Reference to economic principles should not be read as a claim of expertise, but as an attempt to reach across to those who have it.*

Why a working mathematician should write this at all, rather than get back to work? *The seventh recommendation of the Leiden Declaration makes it a duty: mathematicians “have a responsibility to support serious science journalism and to engage in public discourse to explain and contextualize artificial intelligence-assisted methods and results” [Lei26]. The questions addressed in this note will be settled by whoever turns up to discuss them.*

CONTENTS

1	The economy of the residue	3
1.1	A conservative field, moving fast	3
1.2	What is being measured	3
	§ 1.2.1 Counterexamples, and the constructive half, p. 4 §1.2.2 What \$2,000 really buys, p. 4	
	§1.2.3 Capability rests on context, p. 5.	
1.3	The economics were already wrong	5
	§ 1.3.1 Calling the bluff, p. 5 §1.3.2 The externality, and the non-rivalry of the corpus, p. 6.	
2	The audit and the structures	7
2.1	The asymmetry, and the audit	7
	§ 2.1.1 Verification without the producer’s permission, p. 7 §2.1.2 Our asset is their raw material, p. 8	
	§2.1.3 Why do we cite? p. 8.	
2.2	Four questions for an announced result	8
	§ 2.2.1 The third question, institutionalised, p. 9 §2.2.2 The standard already exists, p. 9.	
2.3	The structures are the story	9
	§ 2.3.1 Interfaces in both directions, p. 9 §2.3.2 The unit to fund is the structure, p. 10.	
3	The ecology of the journey	10
3.1	The loop is the profession	10
	§ 3.1.1 The two cuts, p. 10 §3.1.2 The open loop, p. 11 §3.1.3 The loop kept whole: where to look, p. 11.	
3.2	Compression is not acceleration	12
	§ 3.2.1 From proof scarcity to proof abundance, p. 12 §3.2.2 A mismatch of clocks, p. 13.	
3.3	The decision is ours	13
	§ 3.3.1 Priority, and the journey, p. 13 §3.3.2 Nothing is lost, p. 13 §3.3.3 What is actually at stake, p. 13.	
	Notes	15

Acknowledgements. The author thanks Parmy OLSON, whose questions occasioned these notes; Julien CORNEBISE, whose good-humoured and productive disagreement with an earlier version sharpened the arguments that follow; Rodrigo OCHIGAME for warm exchanges on AI, Lean, and understanding in May 2026; and Pierre DÈBES for comments on a previous version of these notes. He also thanks Jesse HAN (MathInc) and Patrick SHAFTO (expMath) for checking the passages describing their work. Thanks also to colleagues in several countries, within the AHGT project and beyond, for many conversations on the future of scientific discovery.

1 The economy of the residue

Three claims, in order: that machines now produce a real but narrow part of mathematics; that this part was always the residue of the work rather than its product; and that the economics of the field had already mistaken the one for the other, long before any model was trained.

1.1 A conservative field, moving fast

Thurston · human understanding · Leiden Declaration · ICM 2026 · means of production

Academia – and fundamental mathematics in particular – was built as a place for producing sound knowledge in the service of human understanding¹: not a stock of results, but a way of deciphering the world and of acting upon it. This is the position on Mathematics of Thurston (Fields Medal 1982): he opens by refusing the obvious question – not “how do mathematicians prove theorems?”, but “*how do mathematicians advance human understanding of mathematics?*” (§ 1) – and closes where this essay starts:

[The proposed division of roles into “speculation” and “proving”] only perpetuates the myth that our progress is measured in units of standard theorems deduced. This is a bit like the fallacy of the person who makes a printout of the first 10,000 primes. What we are producing is human understanding. We have many different ways to understand and many different processes that contribute to our understanding. [Thu94, § 5]

That position makes us, researchers in Mathematics, conservative in adopting new tools and new ways of working. It is therefore worth registering how fast the field is currently moving.

The most valuable tool of a researcher in fundamental mathematics is their voice, so the explosion of discussion is itself the measure. Over the past year researchers strongly engage with AI tools:

- (1) *In practice*: writing, investigating hypotheses, producing examples;
- (2) *In community*: the Leiden Declaration [Lei26] was drawn up in under a year;
- (3) *In public*: the ICM 2026 programme includes a public lecture by Terence Tao (Fields Medal 2006), *Mathematics in the Age of AI*[ICM26];
- (4) *In Funding*: establishment of dedicated funding structures such as DARPA’s *expMath* [expM], which explicitly aims at accelerating progress in pure mathematics².

A field that changes on the scale of two generations has reorganised its conversation in a semester.

Absorbing new modes of producing knowledge is, in one sense, ordinary work for researchers in mathematics: we did it with computers in the 1990s, and we are doing it with mass collaboration and formalisation since *circa* 2020. *What is not ordinary this time is that the means of production belong to someone else.*

1.2 What is being measured

Counterexample search · choice of definitions · ten problems for \$2,000 · Goodhart’s law · what a benchmark measures

Mathematics is often presented as the ideal measure of AI progress: a proof is right or wrong, where prose is a matter of taste. The measurement is real, but it is worth asking what it measures.

¹**Premise.** Throughout, “understanding” means human understanding, and “the field” the human community that reads, checks and teaches. Whether a machine could hold understanding of its own, or constitute an alternative ecosystem with its own criteria, is a real question and not this essay’s.

²expMath oversees multiple projects to shape the interface between AI and a broad range of mathematical disciplines – from number theory to partial differential equations and homotopy theory – combining fundamental research with societal impact.

1.2.1 Counterexamples, and the constructive half. Machines have so far produced results of two kinds: counterexamples to standing conjectures, and – more recently, and beyond refutation – solutions to open problems (§ 1.2.2). We take them in turn.

The first kind is counterexample search – killing conjectures, most recently, and among many, *the Jacobian conjecture*, by Levent Alpöge on July 20, 2026 [RHB26]. This is real, and we do not minimise it: counterexample search is the region of mathematics most amenable to machine methods. It is also the least representative of a mathematician’s work, and it already illustrates, structurally, the missing interface between AI companies and academia (§ 3.1).

Disproving a conjecture is like pruning a branch that grew in the wrong direction. It does not tell us in which direction to look; it is not an end result. A counterexample, like any result, is worth the horizon it opens.

The other half of the work – the constructive half – is deciding which definitions are worth making and which objects deserve names. That half is what makes a field rather than a pile of results: *a theorem is only as good as the notions it is stated in*. As of today, no system yet performs that choosing³. Pruning accelerates a field only if someone is still choosing which branches were worth growing⁴.

1.2.2 What \$2,000 really buys. The second kind goes further. OpenAI has recently reported solutions to ten long-open problems, obtained for roughly \$2,000 of tokens [OAI26]⁵, and – the part worth registering – published them in good order: manuscripts prepared with human help, arguments shipped with a Lean certificate, and, in accordance with the Leiden Declaration [Lei26], an explicit refusal to claim human authorship. The same announcement concedes the limit of operating apart from academia: these questions “cannot be answered by a technology company alone.”

Good order, however, is not yet good practice⁶. The manuscript – a narrative of the discovery – names no one: no co-author to write to, none of the integrators and writers behind the “human help”. The letter of the standard is met while its point – a human interface on which the field can build – is left empty; we return to this *open loop* in § 3.1.

It is worth being exact about what the price tag measures.

What \$2,000 buys is not a theorem; it is evidence of capability at the tasks the benchmark selected – and that evidence holds only for as long as the community that certifies it still functions.

Note that the claim is about *independence*, not about the indispensability of researcher in mathematics – see § 1.3.2. What is mainly at issue is not the correctness of the proof – which is or isn’t independently of the reader – but the inference from them to a capability.

³And the field is moving fast. Beyond this question, what machine’s developing interaction with fundamental mathematics do create is a new frontier to investigate: the shape of the line defining the area where they perform.

⁴Interestingly, it is precisely how the conversation has evolved once the counter-example to the Jacobian conjecture has been obtained: “what does the counter-example tell us on the statement?”

⁵Ten successes at roughly \$2,000 apiece – but out of how many attempts? What about the \$200–500 million development cost for each version of a frontier model with a six-month life expectancy [SCH26]? The announcement does not say: the figure prices a success, a marginal cost, not the search that found it.

⁶OpenAI’s solution to 3. *Non-sofic groups* (existence) is a case in point: the narration rests on an unrefereed preprint (cited as “[Kun19]”) containing typographic errors, and on what experts regard as a strong hypothesis. The existence of a non-sofic group was, moreover, widely-accepted (see [FFF26]).

This is Goodhart’s law⁷, and here it is not a metaphor: the measure is itself the product being sold.

Generative systems are especially exposed to this issue: they are ungrounded, and the party with the financial incentive to optimise the measure is the party producing it.

1.2.3 Capability rests on context. On capability itself, our assessment is plain. A year ago, LLMs behaved in mathematics like badly educated students. They have since been very well trained, and can now identify wrong directions of thought and good direction for problem solving. But their contribution still relies – with a narrow scope⁸ – on context previously built – the definitions, the libraries, the accumulated understanding. Whether that limit is conceptual or merely a matter of engineering, we do not know, and we would not bet against scale: nothing we know of forbids a system from proposing its own definitions – see Footnote 1.

The concession costs the argument nothing. *If machines come to choose the notions, what changes is who chooses – not that choosing is the scarce act.* Someone must still recognise a good choice as good⁹, and it is that recognition, not the generating, which this essay is about.

Whether or not they one day generate their own context, the question will remain that of the human–machine interface: its nature, and its quality.

1.3 The economics were already wrong

Residue and product · public good · externality · the unfunded ecology · mathematical capitalism

1.3.1 Calling the bluff. The economic story about mathematics was wrong before any of this. We are funded as if theorems were the product, and we sell teaching as the justification. Both are residue. The product is human understanding – and no one ever worked out how to buy it: The distinction is not that the theorem is worthless – it is the instrument of transmission – but that what it transmits is *a capacity to think that must be rebuilt in each reader*; the writing, the seminar, the refereeing are that rebuilding, and none of it is what the theorem states.

The membrane between what we understand and the economy has always been close to impermeable.¹⁰

AI has not created this problem. It has called the bluff, because it can mass-produce the residue (Table 1). What it cannot produce is the reason anyone wanted the residue in the first place.

The bluff was, in fact, already on public record. Japan’s ministries of economy and education published a joint report in 2019 under the remarkable title *The Coming Era of Mathematical Capitalism* [METI19]: it names the country’s top three science priorities as “mathematics, mathematics, and mathematics,” observes that the corporate race in AI “requires human resources with a higher knowledge of mathematics,” and asks – explicitly, and without answering – what “management or social system” would suit such an

⁷“When a measure becomes a target, it ceases to be a good measure” [Goo75]. Mathematics has watched this happen before, from the inside: publication counts and impact factors became targets, and ceased to measure what they once tracked. The novelty is not the mechanism but the scale, and that the party optimising the measure now sells it.

⁸“[C]ompared to human papers, LLM ideas are disproportionately concentrated around bridge-like opportunities and synthesis methods [...] with a narrower range” [CZC26].

⁹One could reach for a trained notion of “goodness”, but a trained ranker optimises an already existent proxy, while mathematics favours a notion on which one can potentially build – a fact not available at ranking time.

¹⁰**Orders of economic magnitude** – with the caveat that such studies measure the economic activity mathematics *enables*, not the value it produces. In the United Kingdom, the mathematical sciences were credited with £208 billion of gross value added and 2.8 million jobs for 2010 – *about 16% of GVA and 10% of employment* [Del12] – and with £495 billion and 4.2 million jobs, 13% of UK employment, for 2023 [AMS24]. In France, the corresponding study puts the share of *GDP impacted by mathematics at 18% and the jobs at 3.3 million, 13% of salaried employment*, on 2019 data, up from about 16% and 2.4 million on 2012 data [CNRS22]. The trend is the interesting part; so is the fact that *none of these studies attempts to price what this essay calls the product*.

Tab. 1. Residue and product. *The claim of this essay is the last row: AI drives the cost of the left-hand column towards zero, and thereby makes the right-hand column visible, for the first time, as the scarce thing.*

	The residue	The product
What it is	Theorems, proofs, counterexamples; courses and credentials ^a	Understanding: the choice of notions, judgement, the horizon ^a
Countable	Yes	No ^b
Reproducible	Always was – by any trained mathematician	Never – remade, not copied
Mass-producible by machine	Yes, and increasingly so	No
What pays for it	Grants, publication counts, teaching	Nothing, directly
What AI changes	<i>Its cost falls towards zero</i>	<i>Its scarcity becomes visible</i>

^a A journey is what deposits both columns at once: a countable proof and the horizon it opens.

^b Counting the people who have understood a notion counts instances of transmission, not the understanding itself: the collective grasp of a notion is not the sum of individual graspings.

era. The question of our social contract is thus not a private anxiety of mathematicians; it is an open item on the books of governments. What has changed with AI is that the membrane – between what we understand and the economy – has become permeable, in both directions – mathematics is drawn into the economy as a priced input, and economic demand now reaches back into what mathematicians are funded to do. That is, the traffic is priced.

From a broader societal perspective – with the economy as one of its parts – what crosses the membrane is the portion of the good that carries a price; what still stays behind is the rest – a public able to read and integrate what its machines produce, the slow formation of people, a cultural shift. When does the greater part begin to cross?

1.3.2 The externality, and the non-rivalry of the corpus. The new arrangement gives the old imbalance a precise name.

Responsibility for correctness is the cheap half, and it is now machine-dischargeable; the expensive half – reading, situating, repairing, teaching, turning a correct proof into something that can be built upon – is left to us, researchers, and takes years.

That is an *externality* in the ordinary economic sense: extremely well-capitalised consumers; an unfunded ecology of researchers who keep understanding alive (“maintainers”, says the software idiom – a word that fits the libraries and undersells both the people and the collective intelligence), still needed, by OpenAI’s own admission (§ 1.2.2); and infrastructure that rots quietly. An externality is only an argument if it names an instrument; ours would be structural rather than fiscal, and we return to it in § 2.3.

Two things in particular are consumed and neither is paid for. *And here the economics is, unusually, on the argument’s side:* what is known of the returns to basic research places them in people and their accumulated understanding, not in the results they publish – the residue and the product, under an older name; non-rivalry is why the corpus, not the theorem count, is the asset.^[B] *The first is the corpus:* not the literature, but the digested, exposit, taught mathematics the models are trained on, which took centuries to build and which does not deplete by being read – what depletes is the pipeline that keeps producing and digesting it.

The distinction is not academic, and the companies discovered it empirically: Practitioners report that training on the published literature and on forums like MathOverflow underperformed – though both are already digested for human readers – and AI companies found they had to commission new corpora, written by mathematicians for the purpose. For LLMs as well, the digested layer is not a by-product of publication; it has to be made, by people, deliberately.

The second is the measure – what \$2,000 buys is evidence of capability, and that evidence is only worth something while the community certifying it still functions (§ 1.2). This is not a guild claiming to be indispensable: *it is the ordinary requirement of any science that a measure be certified independently of the party it measures*. Kernels check proofs – imperfectly, and only proofs: whether a statement says what was meant (§ 2.1), and whether ten instances warrant a claim about a system, no machine certifies.

Which raises the question to put to the ecosystem rather than to the models: can it produce understanding, economically and culturally, with no researcher in the loop, when its targets are whatever can be measured?

An oracle answers; a science explains – one should not trade the second for the first without noticing.

2 The audit and the structures

From what is produced to what can be required. Mathematics can verify without asking anyone's permission, which is rarer than it sounds, and it can therefore put questions to whoever announces a result. What it cannot do alone is build and fund the structures on which that capacity depends.

2.1 The asymmetry, and the audit

Verification without permission · formalisation · Lean and Mathlib · training data · attribution

2.1.1 Verification without the producer's permission. A decisive source of tension: this time the means of knowledge production belong to someone else. It is not only a question of what kind of interface is being built, but of scale – *an asymmetry of manpower and of compute that no university can match*. This is serious for a field with a culture of openness, which has always been the cheapest science there is: a person, some paper, access to a library¹¹... some mathematics!

Mathematics, however, is the one field where that asymmetry has a technical answer rather than a negotiated one. Every other discipline that wants to check an AI company's claims must ask the company for access; the weights and the compute are theirs, and access can be withdrawn quietly, without anyone breaching an agreement. We do not have to ask (Table 2).

A proof – written or formal – can, and must, be checked independently of the prover: it has always been the rule of the field. No API key, no compute parity, no permission.

The volume of machine production is a genuine difficulty, and this is where formalisation enters: machine-checked proofs, open source, carried by communities such as the one behind Lean and its library Mathlib [ML20]. It is the one form of refereeing that scales with machine output, and it makes hallucination detectable – at the level of the proof. *But refereeing was never only error-detection, nor the certification of some absolute truth:* it is also how a field digests a result – how one person's argument is read, situated, and made part of what everyone can build on. A kernel scales the first office and not the second.

¹¹*And some coffee!* Following the polymath Paul Erdős' apocryphal "A mathematician is a machine for turning coffee into theorems."

Tab. 2. Who can check whom. *Mathematics is the only row whose verification does not require the producer’s cooperation, and the only one with a form of refereeing that scales with machine output. The “yes” holds at the level of the proof; at the level of the statement it does not.*

	A claim rests on	Who holds the means of verification	Checkable without the producer’s permission?	What scales with machine output
Fundamental mathematics	A proof	Anyone holding the statement and the argument	Yes	Machine-checked formalisation (Lean, Mathlib)
Empirical sciences	Data and an experiment	Whoever holds the apparatus, the sample, the cohort	In principle – at the cost of replication	Nothing comparable: replication does not scale
Machine learning itself	A benchmark score	The company: weights, data, compute	No – access is granted, and can be withdrawn quietly	Nothing: evaluation is itself model-mediated

At the level of the statement, matters are different: no kernel can check that a formal statement says what the mathematics meant – see the accident report of [LS25, § 5]: the authors relate discovering, mid-project, that a kernel-accepted statement was not the theorem they had meant to prove – a Lean evaluation convention (functions silently returning “junk” values outside their domain of definition) had changed the content of the statement, while leaving every proof valid.

2.1.2 Our asset is their raw material. There is an irony here worth stating plainly. The property that makes formalised mathematics auditable – that a proof can be checked by a kernel, with no human in the loop – is exactly the property that makes it valuable as training material for LLMs.

The Leiden Declaration says so without euphemism: what makes mathematics attractive to general-purpose AI development is that “the correctness of formalized proofs can be checked automatically, without the need for human oversight,” which yields “an effectively unlimited source of feedback for training” [Lei26]. *The same property that lets us audit them lets them mine us.* This is not an argument against formalisation. It is an argument for noticing that our defensive asset is also somebody’s raw material, and for asking on what terms the library is being read.

2.1.3 Why do we cite? Attribution is a different matter, and *deeper than an acknowledgement of debt*. Why do we cite? Not to pay persons. We cite prior works as lighthouses: they mark where understanding advanced – the progress, the techniques, the context that cast light, and meaning – over a landscape still being charted. Models are trained on the published commons and, as the same Declaration observes, “frequently return outputs that do not properly cite the human works they synthesize” – on data often obtained by “systematically exploiting licenses and access arrangements that were not made with artificial intelligence in mind” [Lei26]. No kernel adjudicates attribution, and here we are exactly as exposed as every other discipline.

What is lost is not credit: *an output without lighthouses flattens the landscape*, and a flattened landscape is barren ground for building future understanding. Three of the questions below establish whether a claim is sound; only the fourth asks whether the field can build on it.

2.2 Four questions for an announced result

Disclosure · who chose the problem · independent evaluation · First Proof · EMS Code of Practice

Hence the questions we should put to any AI company announcing a mathematical result, written or formal:

- (1) *Is the complete argument public – the understanding, not only the residue?*
- (2) *Who chose the problem, and the terms in which it is stated – and when?*
- (3) *What independent checking process has been applied?*
- (4) *Who closes the loop – is there a named interlocutor, and terms under which the field can read, question, attribute, and build on the result?*

In the formal case¹² this becomes precise: *who wrote the formal statement, and was it fixed before the model started?* If one system produced both statement and proof, the verification establishes mostly *internal consistency* – which is ground for suspicion rather than reassurance about the claim itself, and which leaves the loop open (§ 3.1): nobody stands behind the statement to be asked what it was meant to say.

2.2.1 The third question, institutionalised. Something answering to the third question already exists. First Proof [FP26] evaluates research-level problems under controlled conditions, with expert refereeing for correctness *and* for exposition: of ten problems, seven were solved at publication-level quality, at a compute cost between \$10 and \$1,000 apiece. The score is not the interesting output. What matters is that the evaluation was designed and run by the people qualified to referee it, rather than reported by the party being evaluated – *the third question, institutionalised*.

2.2.2 The standard already exists. None of this is a demand invented for the occasion. The European Mathematical Society’s Code of Practice [EMS] already instructs mathematicians not to make public claims of new theorems unless full details can be provided in a timely manner. And the extension of that standard beyond the profession is no longer a private view: the Leiden Declaration calls on collaborations between mathematicians and industry to abide “at minimum, by the standards we expect of our colleagues” [Lei26]. *We are only asking the companies to meet the standard mathematicians bind themselves to.*

Questions and standards, however, only bind where the two sides actually meet (the fourth question). What remains is to build, and to fund, the places where they do – the structures, to which we now turn.

2.3 The structures are the story

Interfaces · dependency at scale · evaluation and corpus · sovereignty · fund structures, not projects

2.3.1 Interfaces in both directions. Interfaces are now being built from three sides, and they will outlast any individual result: (i) From the companies: OpenAI’s *ChatGPT for Academic Researchers* offers free model access to a hundred thousand scientists [OAI26]; (ii) From academia: First Proof, expMath, Mathlib, and a growing set of instruments built by the community rather than borrowed from outside¹³ – academia building its own instruments, and not only consuming theirs;

Between the two, a *third kind of structure is emerging: interface-structures*, neither corporate laboratory nor university department – MathInc among them, a *company* in the older sense of the word¹⁴ rather than in the sense of these pages have so far had to give it. They hire doctoral students, early-career and senior researchers; they build new interfaces of knowledge, integrating AI capabilities, for sound and fast progress in understanding. What they need from academia is to be developed, funded, and accepted as partners: an alternative to sending our people to the companies, not an outpost of them.

¹²It may be of interest to note that mathlib Lean library is not only a technical instrument but a culture with its own norms: the reference library’s own paper presents “the social organisation that has led us here” [ML20].

¹³For example: <https://www.erdosproblems.com>; Optimization Constants in Mathematics; Competitions of the SAIR type, which “explores the frontier of AI and science through open challenges, benchmarks, and community collaboration”.

¹⁴A group of people with a shared interest who mean to make something happen in the world.

The two sides are not symmetric, and the asymmetry is worth naming twice. *Free access at scale is dependency at scale, not a research partnership*: access is granted, and can be withdrawn without anyone breaching anything (§ 2.1), only now with a hundred thousand people downstream of the decision. In the other direction, two things sit on our side of the table and not on theirs. The first is the evaluation:

A capability claim is worth what its certification is worth, and the certification is ours – designed, run and refereed by the community, see § 2.1.

The second is the corpus, which academia does not merely hold but produces and keeps legible – the definitions, the libraries, the expositions, and the people who write them. *Each side is short of exactly what the other holds.*

That is a negotiating position, and a better one than the current mood suggests, which incite to close (symbiotic?) collaboration – provided we notice that it is the structures, and not the individual results, that are being settled now.

2.3.2 The unit to fund is the structure. If there is an instrument here, it is structural rather than fiscal, and it is not for a mathematician to design it. But the direction can be stated. The oldest interface between researchers and society is education, which produces the value of the next generation rather than of the next quarter – including the capacity to use these tools well – and it needs funding that does not follow the cycle.

The newer interface, where academia and AI companies genuinely meet rather than transact, has still to be built, and it is *the natural place for the externality of § 1.3 to be internalised*: the ecology funded where the corpus and the evaluation are actually produced – the people, not only the libraries they maintain. Both are questions of sovereignty and of culture as much as of economics – a country that stops training the people who can read and integrate what the machines produce has outsourced rather more than its compute.

The unit to fund is the structure, not the project and not the individual: projects and individuals are countable, which is precisely why they are what we fund today.

3 The ecology of the journey

What all of this does to the work itself, and to the people who do it: the loop that constitutes the profession, the mismatch between the speed at which results are produced and the time understanding takes to form, and the one thing in the field that was never reproducible. These are questions of morale as it is of production.

3.1 The loop is the profession

The discovery loop · judgment · human-machine interface · open loop · doctoral training

3.1.1 The two cuts. A widely shared fear – vividly put in a recent essay by a young mathematician [Ham26] – is that the profession survives only as a taste-labelling layer: paid “to optimize the weights for ‘interesting’, ‘beautiful’, anything you like,” permitted to comment and appraise, forbidden to create. The fear can be turned into a mechanical answer.

Discovery and the weighting of “interesting–beautiful” are not two jobs but two moments of a single loop: understanding tells us where to look, and looking is what produces understanding. *The loop is the profession*. Cut it, and the damage arrives at both ends at once – at the *output*, checking and cleaning up after the machine; at the *input*, labelling and weighting with no discovery of one’s own (Figure 1).

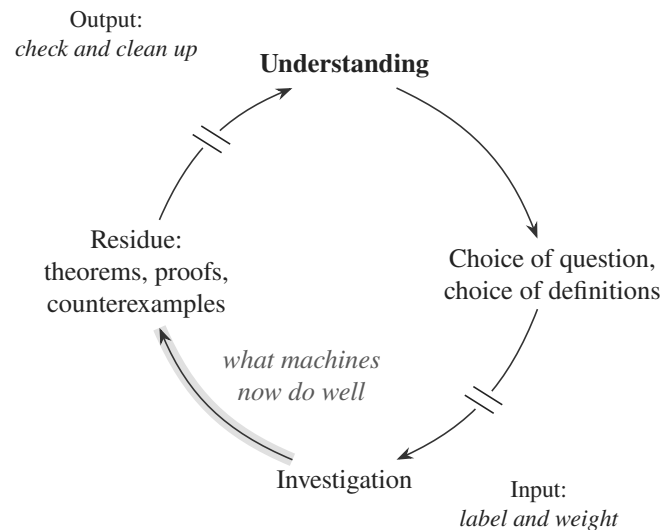


Fig. 1. The loop is the profession. *Understanding* tells us where to look; looking is what produces understanding. Machines are now strong over one quarter of the circumference – and not over the quarter that feeds it. The taste-labelling layer is this same loop cut at the two points where the human meets the machine.

How does one label where to look, once detached from the understanding that alone tells us where to look?

3.1.2 The open loop. The ten problems of § 1.2 are a live specimen of that cut, and of what falls through it. At the input, there is no co-author. There is a published narration of the model’s reasoning, which one can read but not question: nobody to write to, nobody to invite to a seminar, nobody who can be asked why this direction and not another. A narration is a record of a path, not a partner on it.

At the output, the contribution to understanding is real but in its weakest form: what Thurston calls a collection of “answers”, where what people wanted was understanding [Thu94, § 1]. The result is correct, documented – and still to be converted. That conversion, which is reading, situating, exposing and teaching, is years of work; it is ours; and nothing in the announcement funds it. So the proofs are correct and *the loop is left open* – and an open loop is not a slower loop, it is a different object.

3.1.3 The loop kept whole: where to look. The same loop, kept whole, points the other way. AI is a formidable trainer of intuition for reaching new research frontiers: a doctoral student – also any established researcher investigating outside their original field of expertise – can test an intuition and build new understanding in afternoons where it used to cost months. Learning that one asked the wrong question was the most expensive lesson in a mathematical education; it is now nearly free.

What practice maintains, and a labelling layer cannot, is taste – whose deep form, in fundamental mathematics, is the choosing of definitions (§ 1.2).

“Knowing where to look” is not a stock that can be harvested; it is maintained by practice.

Severed from practice, the layer degrades – so the arrangement the fear imagines fails on its own terms.

The danger is not the tool but the routing: the next generation of researchers pushed toward the countable ends of the pipeline, because those are easy to fund and easy to measure – and are precisely the parts that train no judgment.

Until now, nobody funded question-asking spontaneously; it produces nothing you can count. AI may change that: once the countable part costs nothing, the scarce part – the question – becomes visible, and acquires a value our socio-economic system never assigned it. *The bluff, called.*

3.2 Compression is not acceleration

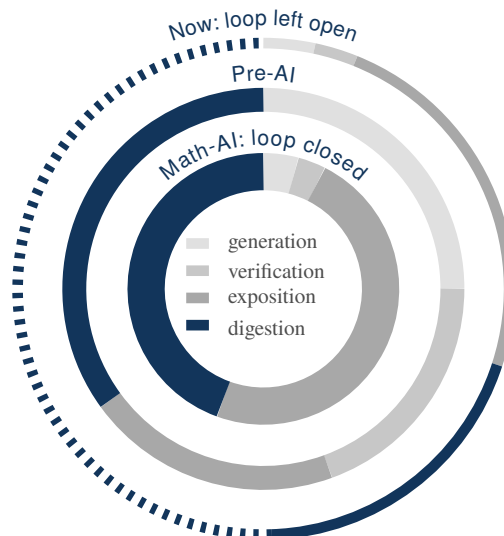
Rate problem · proof abundance · company time vs. generation time · Thurston on foliations

3.2.1 From proof scarcity to proof abundance. AI compresses the stages of the work that can be counted – from an open problem to an argument on the page. It leaves untouched the stage that follows: the one in which a result is read, situated, exposed, taught, and eventually becomes part of what the field simply knows – the closing of the loop¹⁵.

Compressing the countable stages does not accelerate understanding; it widens the distance between what exists and what is understood.

Fig. 2. Time compression of generation is not, by itself, acceleration of understanding (*schematic*). One turn of the loop, from an open problem to a result one can build on. The angle is each stage's share of the turn; the radius is how long the turn takes, so nearer the centre is faster; the thicker the ring, the more broadly the understanding is held.

Machines compress the early stages in every case alike – but only where the closing stage is carried out does the turn shorten. Left open, it never closes at all: compression of the early stages alone cannot reach understanding.



A field can move faster and understand less, at the same time. Economists have a name for the missing quantity: *absorptive capacity*, the ability to use knowledge produced elsewhere, built only by doing research oneself [CL90] – which is why free access to results confers little on a community that has stopped producing the people who can read and integrate them (§ 2.3).

Properly grounded, and held to the code of practice of the mathematical sciences (§ 2.2), *AI can however raise that capacity* – the rate at which frontier work can be taken up – by enhancing individual practice. The distinction matters: canonicalisation is slow because it requires broad deliberative consensus, and AI does not speed deliberation. What it can shorten is *each participant's climb to the point of being able to take part*.

As Terence Tao (Fields Medal, 2006) states it plainly, if no “suitable policy and cultural changes” are made:

[W]e will transition from an era of proof scarcity to an era of proof abundance.
– T. Tao [ICM26, slide 44]

Abundance is not, by itself, progress. The fact of a result travels cheaply; the horizon it opens does not. A thousand results nobody has travelled to will open little more than the fact of themselves.

¹⁵Researchers are now doing exactly this for one of OpenAI's ten solutions: turning the narration into scientific progress (see [MO26] and Footnote 6).

3.2.2 A mismatch of clocks. Underneath is a mismatch of clocks. *Company time is release-cycle time; understanding takes root on the time of a generation.* The two cannot be brought into phase by working harder at the fast end – which is the only end anyone is currently paying to speed up.

Nor is this a machine problem in origin. The precedent is at human scale and required no AI: Thurston’s own account of his first subject, “foliation theory”, where his flood of results emptied the field within a couple of years: colleagues were giving or receiving advice not to go into foliations – they were saying that Thurston was cleaning it out”, and told him, not as a complaint, but as a compliment”, that he “was killing the field”. He is explicit that the territory was not intellectually exhausted; what *the rate of production destroyed were the ecological conditions of the subject*, the ones under which anybody would want to keep thinking in it^[C].

A rate of production can destroy the conditions of understanding. It is a rate problem before it is a machine problem – and it now has a very much larger engine attached to it.

3.3 The decision is ours

Priority vs. journey · Library of Babel · reproducibility · authorship · what cannot be mass-produced

3.3.1 Priority, and the journey. A worry one hears from the next generation of researchers [Ham26] has the shape of the Library of Babel – Borges’ library that contains every possible book, hence, somewhere on its shelves, every theorem, every proof, every refutation, already written. But this situation is already our condition: *every researcher assumes that any result they obtain could be obtained by another trained mathematician.*

Is the experience of discovery being foreclosed? To answer, one must split “discovery” in two. One sense is priority – first place in the world’s ledger. That, machines will indeed make scarcer and noisier. The other is the journey – the personal passage from confusion to understanding – and it is indexed to a person, not to the ledger.

3.3.2 Nothing is lost. The residue was always reproducible; *the journey never was.* Here the essayist quoted above puts it exactly right: “human mathematics is Talmudic” [Ham26] – the charge of learning comes from being conversant, across centuries, with the person who discovered. But then the conclusion inverts: if meaning travels through human authorship, the accumulated corpus – the thing we spend our lives learning and teaching – loses nothing to anything a machine proves afterwards.

Each of us adds a little to a great deal that was already there: almost everything one will ever understand was discovered by someone else, and that has never once made the understanding less ours. For understanding is collective only in the way it is kept alive: not stored but re-walked – each journey rebuilds a piece of the shared corpus in another mind¹⁶.

That is why a machine’s proof, which enters only the ledger, adds nothing to what the field actually understands: *the shared understanding lives only where someone has made the passage again.*

3.3.3 What is actually at stake. The machine takes nothing from what the field understands; it enlarges the residue and leaves the product untouched. What is genuinely at stake is not on the ledger at all – and it is a decision, not a fate.

¹⁶A philosophically inclined reader will recognise a Hegelian movement here: mathematical thought reaching a reflexive stage at which the *Geist* recognises itself in its own productions – transcending any particular individual, medium or epoch, while subsisting nowhere but in those who take it up again.

The journey cannot be mass-produced – and a collective understanding lives only where the journey is made. But it can be routed away, defunded, traded for what can be counted. That decision is ours, not the machine's.

A decision of that kind shows itself in budget lines, in what a committee agrees to count, in the positions offered to people starting out – and it is announced, plainly enough, in the collaborations struck with the new interface-structures (§ 2.3).

The channel of discovery is not being sealed. It has become crowded and noisy – not a bad problem to have, and one we already build for: curated libraries, large collaborative interfaces, machine-assisted formalisation.

The journey, then, is not a private consolation; it is the piece the rest of this essay defends: what closes the loop (§ 3.1), what the rate endangers (§ 3.2), and what the audit and the structures of § 2 exist to keep funded.

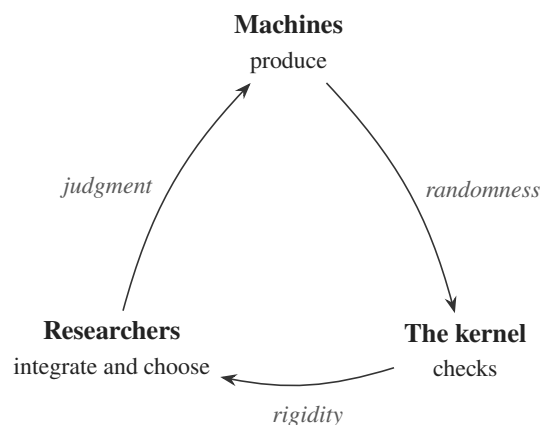


Fig. 3. The virtuous loop. *People choose which questions are worth the asking; machines produce at a rate no community could sustain; the kernel makes what they produce checkable without its author having to be trusted, which opens the door to whoever's argument checks. Researchers integrate what arrives into the understanding – and only that arc, the return to judgment, closes the loop.*

Coda

The models will keep producing mathematics, and mathematics will keep absorbing new modes of production, as it always has. The open questions do not belong to the machines: who checks, who trains, who pays – and whether we continue to organise, and to fund, the one experience that was never reproducible in the first place.

A virtuous loop nonetheless becomes available: machines propose at a rate no community could sustain; formalisation makes the proposals checkable without their authors having to be trusted; and, the material now being machine-tractable, *the same tools can be turned back on it* – questioned, replayed, explained – so that expertise becomes something a newcomer builds rather than a barrier already cleared (Figure 3). And once the corpus once grounded by researchers, those tools let far more of them read, question and answer one another than any seminar could hold. Read against Thurston's case, *the virtuous loop lowers the entry cost of high-level knowledge for a more inclusive and diverse practice.*

None of that is automatic: it holds only where the practice is built for it. Such tools can be turned outward or inward: they can prepare more people to take part in a conversation, or they can answer in its place.

Which happens is settled by the structures, not by the model. The “randomness” from the LLMs, the “rigidity” from the kernel, and the judgment from people – and the loop closes only on the third.

Benjamin COLLAS

Researcher at the Research Institute for Mathematical Sciences of Kyoto University (RIMS), working in arithmetic homotopy geometry – a field that reconciles the rigidity of numbers with the flexibility of geometry. He leads the CNRS International Research Network AHGT, which links RIMS, the Université de Lille and the ENS-PSL, and brings together some ~120 researchers in ~15 countries.

Email: bcollas@kurims.kyoto-u.ac.jp

Web: <https://collas.perso.math.cnrs.fr>

Declaration on the use of AI

The arguments, judgments and factual claims in this essay are the author’s own. An AI assistant (Claude Fable 5 and Opus 4.8, Anthropic, accessed via Claude Code, July 2026) was used to draft and restructure English prose from the author’s own writing and notes, for copy-editing, and editorial critique.

The author reviewed and revised all text, verified the sources, and takes full responsibility for the content.

Notes

- [A] From this perspective, it should be no surprise that *the book of nature “is written in the language of mathematics”* (Galileo 1623) – or closer to us, that the advanced theory of topos is applied to AI (2025). For a debate on the end game of fundamental mathematics or its utilitarianism, we refer to a 1830 comment from Carl Gustav Jacobi (result of Humboldt’s mathematics as *reine wissenschaft* as a source of freedom and personal realization) to Joseph Fourier (the Polytechnicien of the French revolution, for which Mathematics as a state instrument is required) [JCor75]

Fourier avait l’opinion que le but principal des mathématiques était l’utilité publique et l’explication des phénomènes naturels ; mais un philosophe comme lui aurait dû savoir que le but unique de la science, c’est l’honneur de l’esprit humain, et que sous ce titre, une question de nombres vaut autant qu’une question du système du monde.

The maxim comes a few months before the opening of the Liverpool-Manchester steam railway (Sept. 1830), which set off the Railway Mania of the 1840s, and, with it, the globalisation of the industrial revolution.

- [B] Economists have treated basic research as a public good since Nelson [Nel59] and Arrow [Arr62]: non-rival, non-excludable, and therefore privately under-supplied – the standard argument for funding it publicly. *Non-rivalry is also the engine of endogenous growth theory*, in which the accumulation of ideas rather than of capital sustains growth [Rom90] – which puts the corpus, and not the theorem count, at the centre of the economic story. The review literature adds something more useful here. Surveying the evidence, Salter and Martin [SM01] find that the benefits do not travel mainly through the published result: they sort them into *six channels, of which (1) codified knowledge is only one*, the others being (2) trained people, (3) new instruments and methods, (4) networks, (5) problem-solving capacity, and (6) new firms.

That is, in an economist’s vocabulary, the distinction this essay draws between the residue and the product, the limitation of problem-solving results, and the requirement for funding new integrated structures. *A non-specialist’s question follows*: if most of the return travels through people rather than through results, what becomes of the return when the results are free and the people are not funded?

- [C] Thurston’s account, in [Thu94, § 6], deserves quoting at length. Having “fairly rapidly proved some dramatic theorems” as a graduate student:

An interesting phenomenon occurred. Within a couple of years, a dramatic evacuation of the field started to take place. I heard from a number of mathematicians that they were giving or receiving advice not to go into foliations – they were saying that Thurston was cleaning it out.

People told me (not as a complaint, but as a compliment) that I was killing the field. Graduate students stopped studying foliations, and fairly soon, I turned to other interests as well.

On the cause: “I do not think that the evacuation occurred because the territory was intellectually exhausted – there were (and still are) many interesting questions that remain and that are probably approachable.” He names instead two “ecological effects”: a high entry barrier, his results being “documented in a conventional, formidable mathematician’s style” that “depended heavily on readers who shared certain background and certain insights”; and the absence of anything left in it for anyone else – “More than the knowledge, people want personal understanding. And in our credit-driven system, they also want and need theorem-credits.”

The remedy he adopted afterwards is the one this section is arguing for: “In reaction to my experience with foliations [...] I concentrated most of my attention on developing and presenting the infrastructure”, publishing definitions and ways of thinking while remaining “slow in stating or in publishing proofs of all the ‘theorems’ I knew how to prove”, so as to “[leave] room for many other people to pick up credit.”

References

-
- [AMS24] Academy for the Mathematical Sciences, analysis of the economic contribution of the mathematical sciences in the UK, October 2024: £495 billion of gross value added and 4.2 million people directly employed in mathematical science occupations (13% of UK employment), on 2023 data. <https://www.acadmathsci.org.uk/publications/mathsci-supercharges-uk-econ-growth/>
 - [CNRS22] CNRS, Institut national des sciences mathématiques et de leurs interactions (INSMI), *Étude de l’impact économique des mathématiques en France*, 2022: 18% of GDP and 3.3 million salaried jobs impacted by mathematics in France.
 - [Arr62] K. J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER, Princeton University Press (1962), 609–626.
 - [CZC26] Ziyu Chen, Yilun Zhao, and Arman Cohan. *Measuring the Gap Between Human and LLM Research Ideas*. arXiv preprint arXiv:2607.01233, 2026. <https://arxiv.org/abs/2607.01233>.
 - [CL90] W. M. Cohen and D. A. Levinthal, *Absorptive Capacity: A New Perspective on Learning and Innovation*, *Administrative Science Quarterly* **35** (1990), 128–152.
 - [Del12] Deloitte, for the EPSRC and the Council for the Mathematical Sciences, *Measuring the Economic Benefits of Mathematical Science Research in the UK*, June 2013: £208 billion of gross value added (about 16% of GVA) and 2.8 million jobs (about 10% of employment), on 2010 data. Available from the London Mathematical Society, <https://www.lms.ac.uk/>
 - [EMS] European Mathematical Society, EMS Executive Committee, *EMS Code of Practice for Mathematical Publication* (2025), on public claims of new theorems and timely provision of full details. <https://euromathsoc.org/code-of-practice>
 - [expM] P. Shafto (Rutgers University), *expMath* (2025–28), a DARPA programme to accelerate the rate of progress in pure mathematics, <https://www.darpa.mil/research/programs/expmath-exponential-mathematics>
 - [FFF26] Francesco Fournier-Facio, *A torsion-free non-sofic group*, Available at: <https://drive.google.com/file/d/1F1hBzrlGwX42z9h10cd8SkosqlJcHYGt/view> (Accessed: 5 August 2026).
 - [FP26] First Proof, independent evaluation of AI systems on research-level mathematical problems under controlled conditions, with expert refereeing for correctness and exposition; seven of ten problems solved at publication-level quality, at \$10–\$1,000 of compute per problem. <https://1stproof.org> (2026).
 - [Goo75] C. A. E. Goodhart, *Problems of Monetary Management: The U.K. Experience* (1975).
 - [Ham26] K. Hampshire, *The Dark Night of Mathematics – What are we really doing?*, Substack, July 25, 2026. <https://kirwinhampshire.substack.com/p/the-dark-night-of-mathematics>
 - [ICM26] T. Tao, *Mathematics in the age of AI*, public lecture, International Congress of Mathematicians, 2026. <https://teorth.github.io/tao-web/slides/age-of-ai-icm-2026.pdf>
 - [Lei26] J. Portegies *et al.* (working group of sixteen), *Leiden Declaration on Artificial Intelligence and Mathematics*, June 2026. Drafted in the eight months following the conference *Mechanization and Mathematical Research*, Lorentz Center, Leiden University, September 2025. DOI: <https://doi.org/10.5281/zenodo.20302944>.
 - [LS25] D. Loeffler and M. Stoll, *Formalizing zeta and L-functions in Lean*, *Annals of Formalized Mathematics* **1** (2025), 43–56. DOI: <https://doi.org/10.46298/afm.15328>.

- [ML20] The mathlib Community, *The Lean Mathematical Library*, in Proc. 9th ACM SIGPLAN Int. Conf. on Certified Programs and Proofs (CPP 2020), 367–381. <https://arxiv.org/abs/1910.09336>
- [METI19] Ministry of Economy, Trade and Industry and Ministry of Education, Culture, Sports, Science and Technology (Japan), *The Coming Era of Mathematical Capitalism – How the Power of Mathematics Changes Our Future* (Summary), Industry–Academia Round-Table Discussion, March 26, 2019.
- [MO26] MathOverflow, “What are the key new ideas in the proof of nonsolvability of groups in OpenAI’s construction?” Question 513866, 2026. <https://mathoverflow.net/questions/513866/what-are-the-key-new-ideas-in-the-proof-of-nonsolvability-of-groups-in-openai-s-con> (Accessed: 5 August 2026).
- [Nel59] R. R. Nelson, *The Simple Economics of Basic Scientific Research*, Journal of Political Economy **67** (1959), 297–306.
- [OLS26] Parmy Olson, “Math’s AI Crisis Has a Lesson For the Rest of Us,” *Bloomberg Opinion*, forthcoming, Aug. 13, 2026. Available at: <https://www.bloomberg.com/opinion/articles/2026-08-13/math-faces-an-ai-spiritual-crisis-it-has-a-lesson-for-the-rest-of-us>.
- [OAI26] OpenAI, *Ten advances in mathematics and theoretical computer science*: announcement of solutions to ten long-open mathematical problems, 2026, <https://openai.com/index/ten-advances-in-mathematics/>
- [RHB26] A. Freitas Ramos, D. Barros Hulak, and R. J. Guerra Barretto de Queiroz. Formal Verification of an Explicit Counterexample to the Jacobian Conjecture. *Archive of Formal Proofs*, July 2026. https://isa-afp.org/entries/Jacobian_Counterexample.html, Formal proof development.
- [Rom90] P. M. Romer, *Endogenous Technological Change*, Journal of Political Economy **98** (1990), S71–S102.
- [SCH26] Eric Schmidt, “The AI Revolution,” Griswold Center for Economic Policy Studies Reunions Talk, Princeton University, moderated by Alan S. Blinder, May 22, 2026. Video: <https://www.youtube.com/watch?v=yS3rw9lPIPs>.
- [SM01] A. J. Salter and B. R. Martin, *The economic benefits of publicly funded basic research: a critical review*, Research Policy **30** (2001), 509–532.
- [Thu94] W. P. Thurston, *On Proof and Progress in Mathematics*, Bulletin of the American Mathematical Society **30** (1994), 161–177.
- [JCor75] C. G. J. Jacobi and A.-M. Legendre, *Correspondance mathématique entre Legendre et Jacobi*, Bulletin des sciences mathématiques et astronomiques, **9** (1875), 126–142. Available at: https://www.numdam.org/item/BSMA_1875__9__126_1/.

Version history.

- 2026.08.13 - Bloomberg published; ArXiv submitted;
- 2026.08.05 - Additions: the investigation on non-solv groups, Figure 2;
- 2026.08.01 - Version 1.