VOLUME II · ESSAY 01

When AI Starts Doing Mathematics

Proof may become abundant. Understanding may not.

Christophe Pennetier, André-Benoît De Jaegere

Generative Engineering · Volume II · written from October 2026, two years after Volume I · https://generative-engineering.aquia.studio/volume-ii/mathematics/

CONTENTS

The machine that searches, and the machine that says no 5
The caption and the film 7
Could the machine make the film? 10
The right to be useless 12
A fox at the table 15
The next reader 16
Sources 18

In 1957, a young topologist named Stephen Smale proved that you can turn a sphere inside out.

Picture a beach ball, painted blue on the outside and left plain on the inside. To evert it is to bring it, in one continuous motion, to plain outside and blue inside. The rules of the game are strange but strict: the surface may pass through itself, as if it were a ghost of rubber, but it may never be torn, punctured or creased into a sharp fold. Do not try this with a real beach ball. Real beach balls cannot pass through themselves, which is one of several reasons topologists prefer imaginary ones.

Under those rules, the obvious attempt fails. Push one half of the ball through the other and a tight ring forms where the surface has to double back on itself: a crease, which the rules forbid. Play the same game with a circle drawn on a page and it fails for good; a closed curve, it turns out, generally cannot be turned inside out at all.[1] Everyone expected the sphere to share the circle's fate. Smale proved that it does not. A sphere can evert.

His proof was correct. His former adviser, Raoul Bott, one of the founders of the field, flatly told him he was wrong, and explained why. Bott came round, as correct proofs have a way of making people do. But he remained frustrated, like many others, by what the historian of this episode calls the inscrutability of the proof. It said that an eversion existed. It did not say what one looked like. A later commentator compared it to describing a

soufflé in minute detail, down to the molecular chemistry, and expecting someone who had never seen a soufflé to cook one.[2]

The theorem was true, and nobody could see it.

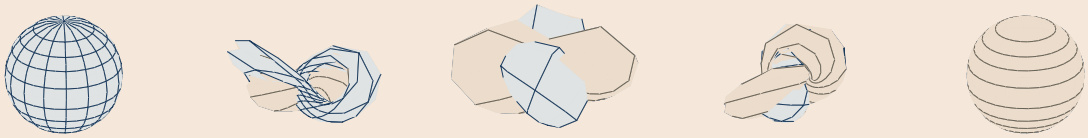
In 1961, Arnold Shapiro found a first explicit way to perform the eversion. He did not publish it. He explained it to a French mathematician, Bernard Morin, who passed it on, and who was among the first people anywhere to understand how a sphere turns inside out.

Morin was blind.

The history puts it plainly: that a blind man was one of the first to understand the eversion is "both a tribute to his ability and a convincing proof that 'visualization' goes far beyond the physical sense of sight." Morin went on to devise a simpler eversion of his own in 1967. Charles Pugh built wire-mesh models of its stages; Nelson Max turned them into a computer film in 1977. The models were later stolen from the Berkeley campus and never recovered. Their digital shadows survive in the film.[2] Somewhere, presumably, a burglar owns a very confusing sculpture.

Seeing, for Morin, had nothing to do with seeing.

FIG. 1 **A sphere can be turned inside out with no tear and no crease: what starts outside ends up inside.**



The sphere **A twisted ring** **Halfway** **Twisted back** **Inside out**

On the page, drag to turn it, play it, or step through its moments; in print, five of them. Blue faces with a grid are the side that starts outside; plain faces with lines one way are the inside. The surface passes through itself, as the rules allow, and never tears or creases. Drawn from the explicit formulas of Adam and Witold Bednorz, a family of ruled surfaces, with their suggested schedule; Smale's proof gave no such picture.[3]

Built

FIG. 2 **A correct proof that a sphere can be turned inside out took about two decades to become something anyone could watch.**

1957 · Smale's proof
Stephen Smale proves that a sphere can be turned inside out. His former adviser, Raoul Bott, flatly tells him he is wrong.

1961 · Shapiro's eversion
Arnold Shapiro finds a first explicit way to do it, and explains it to Bernard Morin, who is blind.

MAY 1966 · The drawings
Anthony Phillips publishes an eversion in Scientific American: most mathematicians see one for the first time.

1967 · Morin's eversion
Morin devises a simpler eversion of his own. Charles Pugh builds wire-mesh models of it.

1977 · The film
Nelson Max turns Pugh's models into a computer film. The models are later stolen; their digital shadows survive.

Dated events from Silvio Levy's history of sphere eversions, written for the Geometry Center's film *Outside In*, with Anthony Phillips's article in *Scientific American* (May 1966).[2][4]

Dated events

Hold on to that gap: between a proof that is true and an understanding that someone else can carry. Mathematics has always needed people to cross it, by explaining, reformulating, finding examples and teaching. Morin's work makes that labour visible. What is changing now is the scale at which new results can arrive, and the attention available to make sense of them.

The change is already visible. When we researched the original *Generative Engineering* study in November 2024, two years after ChatGPT appeared, Epoch AI launched FrontierMath, a benchmark of research-level problems written by mathematicians. The best AI systems solved less than 2 percent of it.[5] By March 2026, the record on that original test was 50 percent of its first three tiers.[6] In June, Epoch republished the benchmark after an audit found errors in more than two problems out of five, so scores before and after are not strictly comparable; on the corrected test's hardest tier, the top score shown in October is 100 percent.[7][8] In September, OpenAI announced a proposed proof concerning the Navier–Stokes equations, one of the Millennium Prize problems. In October, it released 719 mathematical manuscripts, organized in 372 families, about 42 percent of their top-line results formalized.[9][10]

FIG. 3 **Two years after the original study, machines went from under 2 percent of a research-level benchmark to a claimed Millennium problem and 719 manuscripts.**

NOVEMBER 2024 · FrontierMath
The month our original study is researched, Epoch AI launches FrontierMath: AI systems solve less than 2 percent of it.

JULY 2025 · Olympiad gold
An advanced Gemini model reaches an officially assessed gold-medal standard at the International Mathematical Olympiad.

MARCH 2026 · Half of FrontierMath
A model solves 50 percent of FrontierMath's first three tiers, in Epoch AI's own runs of the original test.

JUNE 2026 · A corrected test
Epoch AI republishes FrontierMath after an audit finds errors in 42 percent of its problems; earlier scores are not strictly comparable.

JUNE 2026 · First Proof
In a refereed test on ten research problems, seven receive at least one solution judged acceptable.

SEPTEMBER 2026 · Navier–Stokes
OpenAI announces a proposed proof of finite-time blow-up for a forced formulation, with a Lean formalisation.

OCTOBER 2026 · 719 manuscripts
OpenAI releases 719 manuscripts in 372 families, about 42 percent formalised. On the corrected test, FrontierMath's hardest tier reaches 100 percent.

Dated events from the sources cited in this essay: Epoch AI (2024, 2026), Google DeepMind (2025), First Proof (2026), OpenAI and its repository (2026). Benchmark scores are not research results, and the Navier–Stokes claim is not an accepted Millennium Prize resolution. [5][11][6][12][9][10][7][8]

Dated events

Proofs are becoming abundant. Understanding is not.

That is the subject of this essay, and the reason mathematics comes first in this volume. It is the field where the referee is cheapest, where a proof can be checked by a program, and therefore the field where the gap between a true result and a seen one is about to be tested at industrial scale. If Smale's inscrutable proof took twenty years to become a film, what happens when the inscrutable proofs arrive by the hundred, every month, from a machine?

The machine that searches, and the machine that says no

In the original study, mathematics already looked like the place where symbolic rigour, learned patterns and search would meet. The picture was a partnership: human intuition steering computational power.[13] The picture survives. Its comfortable division of roles does not.

The new systems are not the brute-force engines of popular imagination. In January 2024, AlphaGeometry solved olympiad geometry problems without human demonstrations, as its authors reported in *Nature*.[14] AlphaProof combined learned guidance, reinforcement learning and formal search, and with AlphaGeometry 2 reached a silver-medal standard at the 2024 International Mathematical Olympiad; the work was later published in *Nature*. A year later, an advanced Gemini model reached an officially assessed gold-medal standard writing its solutions in natural language.[15][16][11] Six months later I wrote that I did not feel replaced, not even imitated.[17] Olympiads are not research, which is why the First Proof project matters more: four systems, ten problems drawn from active research, thirty-nine submissions refereed by experts. Seven problems received at least one solution judged acceptable, and the referees found some genuinely interesting routes alongside the errors.[12] Long before any of this, in 2021, machine learning had already helped mathematicians spot connections in knot theory that they then proved.[18]

Two machines are now at work, and they are easy to confuse. The first proposes: it decides what to try, which intermediate statement might help, when a promising path has become an expensive misunderstanding. The second refuses: given an argument, it says whether each step follows. Lean is the second kind. It does not become a discoverer because a discoverer uses it, any more than a referee scores goals. I have called the two, elsewhere, imagination and proof: imagination makes the guess; proof earns the right to trust it.[19] In December 2024, the month after our study's research, a group of researchers argued in a position paper that formal reasoning with proof assistants had become "indispensable" for progress in AI for mathematics.[20] The referee, in other words, was hired first.

Think of the formal argument, for a moment, as a caption for a mathematical event. We will come back to the film. The checker's immediate concern is whether the caption follows the rules, and Lean, the proof assistant in which a formal proof is checked line by line, cares about

nothing else. Lean is the perfect caption: every step explicit, nothing left to goodwill. It is the first reader in the history of mathematics who cannot be bluffed by the word “clearly”. It is also the only reader in the room who never asks why. A Lean proof can be valid and as inscrutable as Smale’s, only longer. Checking a caption does not project the film.

Dragons in an empty room

Here is a small mathematical joke that deserves to be framed beside every triumphant screenshot of a proof checker. I announce that I have proved every dragon in this room is blue. You check the argument. It is impeccable. Then you notice there are no dragons in the room.

The statement was true. The announcement was doing rather more work than the theorem.

A formal proof certifies that a stated proposition follows from its assumptions. That is an enormous amount, and it is not everything. Lean’s own documentation explains that its guarantees depend on the axioms used; a placeholder can stand in for a missing proof; a command exists to list what a theorem actually rests on.[21] A proof of a weakened statement can be entirely valid. A hypothesis can quietly exclude the case that made the original problem hard. A definition can encode a slightly different object from the one a reader has in mind. None of these requires the checker to make a single mistake.

The fluid-dynamics announcements of September are the perfect classroom, because the same words cover different problems. OpenAI’s claim concerns finite-time blow-up for the three-dimensional Navier–Stokes equations with a smooth forcing term, with formal artefacts attached.[9] The official problem statement does include forced alternatives, so “forced” is not a disqualification; but it is an essential part of the claim, and it must not vanish into a headline about all turbulent fluids.[22] The Clay Mathematics Institute welcomed the announcement that the problem had “apparently been settled”, and reminded everyone that its own process is “deliberately unhurried”.[23] It means it. The last Millennium problem it confirmed, the Poincaré conjecture, was solved in preprints posted in 2002 and 2003, and the prize was awarded in March 2010; that July, Grigori Perelman turned down the million dollars.[24][25] Seven years from preprint to prize, then a polite no: in mathematics, that counts as a process working smoothly.

A neighbouring line of work, by Adarsh Ganeshram, Valentin Duruisseaux and Anima Anandkumar on the unforced Euler equations, used physics-

informed networks to find a candidate singular solution, then numerical certification and stability analysis; Terence Tao's assessment at the time stressed the substantial work still needed to turn a numerical candidate into a theorem.[26][27] A network can search a space of approximate functions. Numerical analysis can bound the errors. Analysis can say when an approximate object implies an exact one. A checker can certify an encoded argument. At every handoff, the question is what has actually been carried across, and the confidence earned at one layer must not be spent twice at the next.

And even a perfect proof does not explain itself. The kernel is not offended by an opaque construction. It has no seminar to teach next Tuesday. That is not a reason to distrust formal methods; informal mathematics makes the same mistakes with fewer tools for finding them. It is a reason to remember what the soufflé recipe was for. Smale's proof was correct in 1957. The film came twenty years later, and it came from people.

Suppose all those checks succeed. The theorem is the one we meant, the assumptions are the ones we intended, and the proof is valid. The question that troubled Bott remains: what have we understood? To answer that, we have to return from the certificate to the experience of doing mathematics.

The caption and the film

I am not a Fields Medallist. I have never been a research mathematician. But when I was young, at the Jet Propulsion Laboratory, and later during my PhD and my years in academia, the remark that came back most often about my mathematics was neither "right" nor "wrong". It was "elegant". Allow me this one immodesty, since it is my essay: I was proud of it then, and, rereading this sentence, I notice that I still am. It was never a matter of rank. It came from a passion, from using mathematics every day the way an engineer uses a favourite tool, and from intuitions that went looking for the door rather than hacking a path through the brambles. Change the space, and the problem shrinks.

Can one always be elegant? No, and it would be dishonest to pretend so. Some problems ask only whether you reached the goal, and for them brute force is perfectly respectable, provided it succeeds: nobody asks a bridge whether its calculations were beautiful. The real question is what kind of success we want once the force, brute or otherwise, belongs to a machine, and that question is older than the machines.

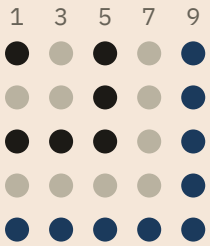
William Thurston, who spent his life seeing shapes other people could only prove, asked the question in its sharpest form. The issue, he wrote, is not how mathematicians prove theorems, nor even how they make progress. It is this: “How do mathematicians advance human understanding of mathematics?” When Appel and Haken proved the four-colour theorem with a massive computer calculation, the controversy that followed was not, in his reading, about whether the theorem was true. It reflected “a continuing desire for human understanding of a proof, in addition to knowledge that the theorem is true.”[28]

Where does that understanding live? Not, apparently, in the words. Asked by Jacques Hadamard how he thought, Einstein described elements that were visual and partly muscular; words came later, laboriously, when he had to communicate.[29] Henri Poincaré described the moment a connection in his work on Fuchsian functions arrived as he stepped onto an omnibus in Coutances, in the middle of a conversation about something else. He did not check it on the step. He checked it later, at home in Caen.[30] The flash and the verification belong to the same story, and only one of them fits on a page.

I think of the written proof as a caption, and of understanding as the film. The mathematician sees something: a motion, a symmetry, two distant objects with the same shape. Then comes the long work of turning that film into rules, and the rules into a sequence of statements another person can follow. The caption is precise, and it is incomplete. It leaves out the wrong turns, the half-analogies, the reason anyone thought of trying this at all. The reader has to rebuild a film from it, and every reader rebuilds a slightly different one.

Take the smallest example I know. The first n odd numbers always add up to n^2. One proof is an induction: true for one; and if it is true for n, adding the next odd number gives $n^2 + 2n + 1 = (n + 1)^2$. The recurrence is there in the algebra. A growing square makes the same recurrence visible, and on the page you can build it yourself.

FIG. 4 **Add the odd numbers in order and the dots always make a square.**



1 + 3 + 5 + 7 + 9 = 25 = 5 × 5

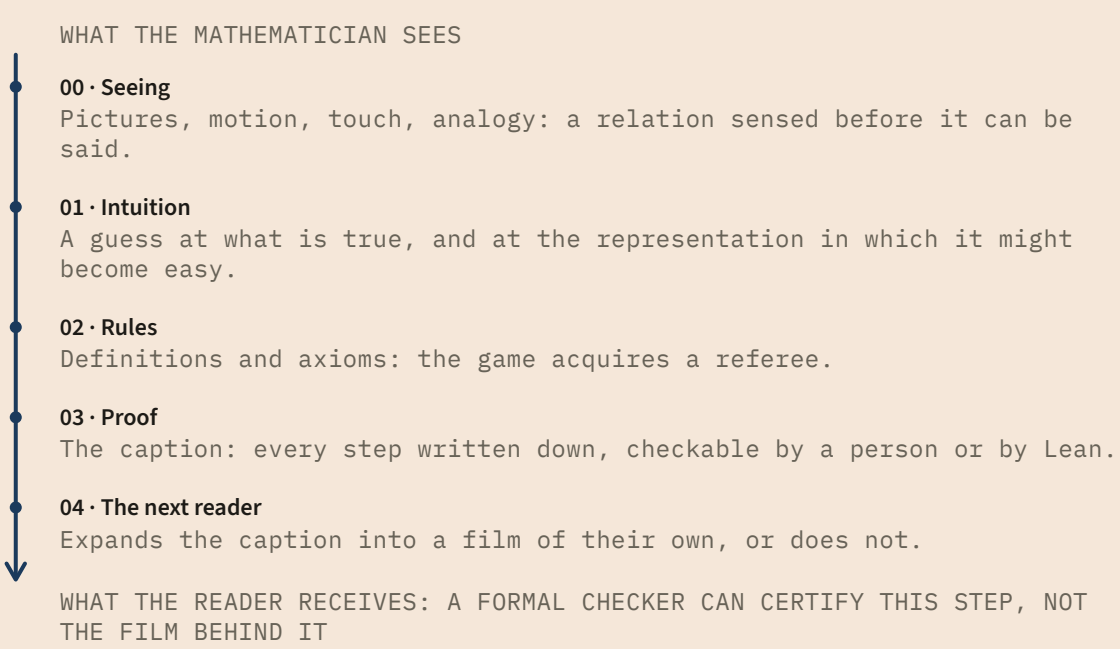
Each odd number is an L that wraps the square before it; the newest, 9, is drawn in blue.

Each number above the square counts the L that starts at the top of its column. The nth odd number, $2n - 1$, is n dots across and $n - 1$ down: it wraps the square of side $n - 1$ into the square of side n. The picture and the induction express the same argument in different forms.

Built

The induction carries the mechanism algebraically; the square makes it spatial. Both can prove, and both can explain, although one may make the relation easier for a particular reader to grasp. A good mathematician wants both, and an elegant proof is often the rare caption from which the film can be rebuilt.

FIG. 5 **Mathematics travels as a caption; the reader has to rebuild the film.**



Read from top to bottom. A conceptual figure, not a measurement: language is an instrument of thought as well as its delivery, and the loop also runs upward.

Conceptual

The metaphor needs one correction before it hardens into a theory. We also think with words. A definition can create an object we had never imagined, and a symbol can hold a distinction that memory would drop.

Cognitive scientists have argued that language is chiefly a tool for communicating rather than for thinking, and the evidence for separate systems is strong;[31] but nobody who has watched a good notation make a problem collapse believes words are mere packaging. The caption shapes the film it describes.

This is also why large language models surprised us. They were trained on captions, billions of them, and it turned out that a library of captions is not empty of films. Researchers found representations of geographical space and historical time inside models trained only on text.[32] The surprise was never that words contain words. It was how much of the world had already been pressed between them.

Which raises an awkward question about sight itself. If understanding is a film, must the mathematician have eyes? History answers rudely. Euler did a large part of his later work almost entirely blind, dictating, among other things, an algebra to a young tailor whose ignorance of mathematics forced the master to be clear.[33][34] Lev Pontryagin lost his sight at fourteen and became one of the great mathematicians of his century; his mother read him the texts aloud.[35] A study of three professional mathematicians blind from birth or childhood found their mathematics running on broadly the same brain network as sighted colleagues, with some extra help from the visual cortex.[36] The film, it seems, is not made of light. It is made of relations, transformations and structures, and they can arrive through the fingers, through motion, through a voice in the room.

Could the machine make the film?

The caption and the film leave us with a question we have not yet answered. We have seen machines produce arguments, and we have seen what it takes for a person to make an argument intelligible. Could the machine also learn to make the film: to build a representation, explore it, and recognize a relation that survives a change of scene?

The machine already searches, and the search is getting good. What it searches through is the real question. A system can become excellent at choosing the next step inside a proof and remain poor at noticing that the whole problem should be translated into another language: algebra into geometry, an object into a space of objects, an expression into an invariant. That translation is where elegance usually lives. It does not shorten the walk. It reveals a door.

This is where the world-model researchers come in, and it is worth being precise about what they are building, because the money has stopped waiting. The projects differ, but each gives us a concrete way to examine what a machine might carry behind its description of the world. Fei-Fei Li calls today's language models "wordsmiths in the dark": eloquent, learned, and ungrounded in the physical world.[37] Her company, World Labs, released Marble in November 2025, a model that turns text, images or video into editable three-dimensional worlds; in September, AMD agreed to buy the company for about $8.2 billion in stock.[38][39]

Yann LeCun's AMI Labs raised $1.03 billion in March to build world models on his JEPA architecture, while saying plainly that it has no plans to earn revenue for now; Meta's V-JEPA 2, from the same line of work, learned from video and was adapted to plan simple robot actions.[40][41][42] General Intuition, which learns from video-game footage, closed $220 million at a $6.2 billion valuation in late September.[43] Google DeepMind's Genie 3 generates worlds that one can walk through in real time.[44] The money has made its own prediction about the future, which is, after all, what a world model is for.

These are ambitions as much as results; AMI's chief executive expects the road from theory to product to take years. Beneath that commercial race lies a question of representation: what internal structure makes a relation easier to discover? There the mathematics is getting cleaner, which is usually a good sign. LeJEPA, by Randall Balestriero and LeCun, replaced a pile of training tricks with a single geometric requirement: the model's internal representations should spread evenly in every direction, like an isotropic Gaussian cloud, instead of collapsing into a corner where everything looks alike. It checks the requirement by looking at many random one-dimensional slices, relying on an old result of Cramér and Wold: if every slice looks right, the whole cloud is right.[45] One constraint in place of a dozen recipes: that is what elegance looks like in a machine-learning paper, the door rather than the path through the brambles.

Not every model in this landscape generates scenes. At Caltech, Anima Anandkumar developed neural operators that learn whole families of solutions to the equations of physics, and has started a company, Accelerated Understanding, to push physical simulation further.[46][47] All of these projects ask a different question from the language models: not which symbol comes next, but what situation is being represented, and what would happen if we pushed it.

That is what we do when we think well. We do not fall down every staircase to learn that stairs have consequences. We carry, compressed, the laws we have absorbed through a lifetime of watching things fall. Newton's leap, as Feynman told it, was to see that the Moon is falling too, continuously, around the Earth: the apple and the Moon are one story in two costumes. [48] The achievement was not a sharper image. It was recognizing which relation survives a change of scene.

The relevance to mathematics is not that a generated world is already a mathematical intuition. It is the possibility of carrying a structure that can be manipulated, tested and recognized in another setting. A mathematical world model, if anyone builds one, would do the same thing with mathematical objects. It would represent examples, counterexamples and dependencies; its moves would include changing coordinates, strengthening a hypothesis, translating a question into another field; and it would hand its best guesses to the machine that says no. I have sketched such an architecture before, as several kinds of intelligence coordinated around discovery, verification and explanation.[49][19] It remains a research direction, not a result. Change the space, and insight follows; but somebody, or something, has to choose the space.

The right to be useless

Suppose a machine learns to find a representation that makes a problem easy, and even helps us understand why it works. We would still have to ask which problems deserve that effort. The question of taste survives the improvement of the machine.

Or suppose it produces a thousand new theorems before breakfast, every one correct, every one never seen before. We would still know remarkably little… about the breakfast. Novelty is a relation to what has been recorded. Significance is a judgment about what changes.

G. H. Hardy put it with the confidence of a man who never had to fill in a grant application: "Beauty is the first test: there is no permanent place in the world for ugly mathematics."[50] Elegance, in that sense, is not decoration. An elegant argument exposes a mechanism, drops an unnecessary assumption, brings distant cases under one roof, travels to the next problem. Beauty is how we recognize those functions before we can price them.

There is a difficulty, though: the significance of an idea is often clearer after the exploration than before it. Taste has to leave room for work whose value cannot yet be demonstrated. This is where the argument reaches a word that sounds, at first, like an objection to the whole enterprise: useless.

The word for this moment was found by a friend, and it would be neither honest nor fair to pretend otherwise. In October, Tesla unveiled a Roadster that climbs a ramp by suction, flips through the air on jets of compressed air and lands on its wheels; a curator at the Smithsonian's air and space museum called it hopping.[51] Carlos Diaz, who hosts the Silicon Carne podcast, called it something better. This, he said, speaking from emotion rather than from the spec sheet, is the moment of *l'inutile*, and he carried the thought all the way to haute couture: the gown nobody needs and everybody looks at. French has one word for all of it. English, characteristically, needs a dozen.

Useless? Ah no, as Cyrano told the bore who found his nose merely big, that is a little short.[52] One could say so many things, varying the tone. The accountant: *unprofitable*, it returns nothing this quarter. The engineer: *redundant*, we already have one that works. The manager: *superfluous*, it was not in the specification. The puritan: *frivolous*, and therefore suspected of being fun.

So much for the prosecution. Now let the words testify. The preacher: *vain*, from the Latin for empty; ask any designer how much work an empty space takes. The Latinist: *futile*, from a word for a vessel that leaks; every laboratory leaks ideas, which is the whole point of a laboratory. The farmer: *idle*, like the field left fallow so that next year's harvest has somewhere to grow. The classicist: *otiose*, from *otium*, leisure; and the Greeks' word for leisure, *scholē*, is the one that gave us *school*. The lawyer: *gratuitous*, given for nothing, out of grace.[53] And the topologist, last and proudest: *superficial*, concerned only with surfaces, which in topology is the whole job. Smale's most famous theorem is superficial from beginning to end.

Carlos stays, gladly, with the futile and the superficial, where a flying car and a couture gown both live, and he is right to enjoy the view. I go further, and I arrive from the other end of the building: from research.

Research is mostly throwing things away. Ninety-nine ideas out of a hundred end in the bin, and the hundredth only exists because the ninety-nine were tried. You do not find the extreme by polishing the average; you find it by stripping the average from the field. Trash, and trash again, until

something refuses to be trash. Measured as output, the day was wasted. Measured as what you now know not to pursue, it was the most productive day of the month.

That is how I reach haute couture: through the back door, where the discarded toiles are kept. Luxury displays purchasing power. Couture, at its best, displays something else: a mastery of material, cut and proportion so rare that it cannot be justified by how often the garment is worn. The first kind of uselessness is vanity. The second is savoir-faire made visible, and the hundred trials that preceded it are part of the garment even though nobody sees them.

Design knows this too. A search page with one box and nothing else is harder to make than a portal crowded with weather, stock prices and horoscopes, because every addition has a sponsor and the empty space never attends the meeting. Elegance is the discipline of the empty space. In a proof, it is the invariant that makes nine of the ten auxiliary constructions disappear. The final simplicity hides the labour of deciding what not to keep. It looks effortless only to those who never had to throw anything away. Taste is what decides what to throw.

Abraham Flexner made the institutional case in 1939, in an essay whose title reads like a contradiction and works like a theorem, *The Usefulness of Useless Knowledge*: much of what changed the world grew from curiosity that answered to no immediate application.[54] Computer science has its own echo. Joel Lehman and Kenneth Stanley showed that, in deceptive problems, searching for novelty alone can outperform searching for the objective, because the objective rewards steps that lead nowhere.[55] A measure can be perfectly accurate about the wrong horizon. In a western I once wrote about the brain’s chemistry of surprise, stability and attention, the outlaw who plays surprise puts it more rudely: take away the surprise, and thought becomes bureaucracy.[56]

Here the machine could be a gift, or a flood. If testing a conjecture becomes cheap, a bad analogy need no longer cost a semester; one can trash faster, and better. Or one can produce a hundred variations of the same mistaken idea and call it exploration. A system trained to imitate the existing literature will deepen its grooves. The next useful idea may lie exactly where the literature is thinnest, which is also where imitation is weakest. Taste, in the end, is the ability to tell a new direction from a new decoration.

When the cost of trying an idea changes, the people able to carry it forward can change too. A connection that once remained in somebody’s

head may now become a conjecture, an example or a failed proof that teaches something. This is where the question of taste becomes a question about the people around the table.

A fox at the table

The line is older than any university. Archilochus, a Greek poet of the seventh century BC, left a fragment: the fox knows many things, but the hedgehog knows one big thing. As zoology, it is fair. The fox has a hundred tricks and runs through them all; the hedgehog has exactly one, it rolls into a ball of spines, and the fox, for all its repertoire, goes home hungry. In 1953 Isaiah Berlin turned the fragment into a way of sorting minds, in an essay on Tolstoy. Hedgehogs relate everything to a single central vision, one system in which everything makes sense: Plato, Dante, Dostoevsky. Foxes pursue many ends, often unrelated, and take the world in its variety rather than fitting it into one design: Aristotle, Shakespeare, Goethe. Tolstoy, Berlin argued, was a fox who believed in being a hedgehog.[57]

Mathematics keeps its own bestiary. Freeman Dyson's pair is birds, who survey the landscape from above, and frogs, who work in the mud below; he insisted that both are needed, and counted himself a frog.[58] Timothy Gowers drew the line between mathematicians who regard their central aim as solving problems, and those more concerned with building and understanding theories.[59] In the same essay he quotes Michael Atiyah, asked in 1984 how he chose his problems: "I just move around in the mathematical waters, thinking about things, being curious, interested, talking to people, stirring up ideas; things emerge and I follow them up."

That is a fox, swimming.

I know the fox's life from the inside. You see a connection between two fields. It is vivid enough to be irritating and incomplete enough to be indefensible. Making it real would take techniques you have not mastered, collaborators you have not convinced, and months your actual job will not release. So most of what you see goes in the bin, unproved. Meanwhile, at the table where the meals are served, the hedgehog has solved a well-posed problem. There is an answer; it can be shown; it is rewarded. Organizations are built around such answers, and perhaps by selection, perhaps by training, the world is mostly built by hedgehogs.

The machine changes this, and not in the direction people expect. It is, first of all, a magnificent hedgehog: one big trick, performed tirelessly and,

more and more often, better than we perform it. The short, visible, immediately rewarded work of executing a well-posed step is precisely what it does best. That is what makes the moment hard for the hedgehog. It can lift more than you; it can now prove more than you; and the reward that structured a career, the solved problem, is losing its scarcity.

For the fox, the same machine is the collaborator that never arrived. The fox can ask for an example, then a counterexample, then an attempted proof of the narrow claim on which the whole intuition hangs, before lunch, without assembling a team or begging for a semester. The fox does not need help proving everything in the fox's head. The fox needs help discovering how much of it should leave. A machine that patiently explains why the conjecture fails is a better ally than one that enthusiastically writes the abstract.

None of this means the future belongs to generalists. An expert with good tools gains more than a novice with the same tools, because the expert knows which result is surprising and where plausible arguments break. What the machine dissolves is the old bundle we called a mathematical job: choosing a question, finding an approach, producing the proof, checking it, explaining it, judging its significance. Those can now be separated, and the human does not automatically keep the first and the last. The division of labour should be earned by performance, not assigned by species.

A division of labour is also a division of opportunity. The fox can only test the conjecture if the instrument is available; the specialist can only interrogate the result if there is time, access and recognition for doing so. The table is no longer just a metaphor for a career. It is becoming an institution we have to design.

The next reader

A stack of manuscripts can grow faster than a community can read it. Seven hundred and nineteen of them, in 372 families, would keep a diligent mathematician busy for three lifetimes, and the repository itself warns that unformalized results may contain errors.[10][60] Mathematicians used to write "left as an exercise for the reader" as a small, affectionate cruelty. At 719 manuscripts, it becomes a career. Even if every one were correct, someone would still have to establish what was already known, which results overlap, and where a good idea is hiding under a bad

presentation. Correctness removes one uncertainty. It does not do the work of a literature. The download is free. The attention is not.

The September and October quarrels were about who sits at that table, and about the work that arrives with the seat. On 11 September, a statement signed by twenty-five Fields Medalists warned of “a severe misalignment” between AI development and mathematics, citing understanding and the formation of future researchers.[61] On 29 September, the advisory group AGMAI asked companies to stop testing advanced problems on proprietary models that mathematicians cannot access. On 7 October, the Association for Human Mathematics called on mathematicians to stop working with OpenAI, after its release of mathematical documents.[62][63]

China chose another posture. On 11 October, an alliance led by Shing-Tung Yau was launched at Tsinghua, joining eight mathematical research institutions around specialised AI models for mathematics, with shared computing resources and formal infrastructure.[64][65] Openly released provers such as DeepSeek-Prover-V2 and Goedel-Prover-V2 belong to the same landscape, though open weights are not the same as access to the strongest private systems.[66][67] In the United States, the Simons Foundation announced a $15 million effort, with Anthropic and OpenAI, to help mathematicians work in this new environment.[68]

Across these responses, two reflexes emerge: one protects the table, the other sets more places at it. I argued in *The Theorem Does Not Care* that a result does not become less true because it unsettles our professional identity, and that the anger belongs elsewhere: on who gets the instrument.[69] A proof, access to the model that found it, and control of the computation behind it are three different things. A PDF gives you the first. A monthly subscription does not give you the third. The fox needs a seat at the table, and so does the hedgehog; neither gains much if the table has become a product available by invitation.

Access is only the first part. Once the proof is on the table, somebody still has to make it usable by someone else.

The best responses so far did not build a wall. On 24 October, four mathematicians launched oai-math-revised, a community project to improve the exposition, references and organization of those manuscripts.[70] At Caltech, after a controversy, organizers and critics redesigned a mathathon around understanding rather than racing, under the theme “Old Problems, New Proofs”.[71] Both treat a machine-made proof the way

Morin treated Shapiro's: not as the end of the story, but as the beginning of a film somebody has to make.

A community cannot give every result the same attention. It has to decide which ones deserve to be unpacked, connected and taught, and how to recognize them before all that work is done. Taste now shapes the reading programme as well as the research programme.

I have called this reverse mathematical analysis: starting from the result, and rebuilding the territory around it. What changes when an assumption weakens? Is there a simpler core? Is the strange auxiliary object an accident, or the real discovery? Does the technique travel? That work deserves credit, visibly. If a model proposes a construction, one researcher finds the missing hypothesis, another formalizes the argument and a fourth discovers the general principle, "AI-assisted" tells us almost nothing about who did what. Otherwise we will have built an answer factory with an unpaid reading department.[72]

The same choice reaches the classroom. A student can ask the machine for a proof, read it until it feels familiar, and hand it in. Or a student can propose a conjecture, predict which hypothesis matters, ask the machine for counterexamples, and defend a revised statement alone. Both use AI. They are not the same education. The first produces a polished page. The second produces the next Morin: someone able to see what a correct proof is actually about.

Smale's theorem did not need Morin to be true. It needed Morin to be understood. The machines are about to give us a great many theorems that are true, and some that are beautiful in ways we have not yet learned to see. Refusing to read them would be a strange way to defend curiosity. Reading them is where the work begins.

A proof tells you that. Understanding tells you why. Taste tells you so what.

> *If proofs become cheap, does mathematical taste become the scarce resource?*

Sources

Each source says what it can carry: a peer-reviewed study, a preprint not yet reviewed, a book, a research organisation, an institution, a company's own announcement or report, code, or a press report. Facts quoted here were read at the source in October 2026; each entry gives its own date.

[1] Silvio Levy, Delle Maxwell and Tamara Munzner. Outside In: introduction to the film. The Geometry Center, 1995. Research centre film and guide.
[2] Silvio Levy. A brief history of sphere eversions. Making Waves: A Guide to the Ideas Behind Outside In, The Geometry Center, 1995. Book chapter, history of mathematics.
[3] Adam Bednorz and Witold Bednorz. Analytic sphere eversion using ruled surfaces. Differential Geometry and its Applications 64, 59 (2019); arXiv 1711.10466, 2019. Peer-reviewed study.
[4] Anthony Phillips. Turning a surface inside out. Scientific American 214 (5), 112–120, May 1966. Magazine article.
[5] Epoch AI. FrontierMath: Evaluating advanced mathematical reasoning in AI. Epoch AI, November 2024. Research organisation.
[6] Epoch AI. GPT-5.4 set a new record on FrontierMath. Epoch AI newsletter, March 2026. Research organisation.
[7] Epoch AI. FrontierMath Tiers 1–3 (v2). Epoch AI benchmark page, on the update of 12 June 2026, June 2026. Research organisation, open data.
[8] Epoch AI. FrontierMath Tier 4 (v2). Epoch AI benchmark page, updated 19 October 2026, October 2026. Research organisation, open data.
[9] OpenAI. On the Navier–Stokes Millennium Prize Problem. OpenAI, 8 September 2026. Company announcement, with a Lean formalisation.
[10] OpenAI. openai/math: mathematical manuscripts and supporting proof artifacts. GitHub repository, checked 25 October 2026, October 2026. Company repository.
[11] Thang Luong and Edward Lockhart, Google DeepMind. Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad. Google DeepMind, 21 July 2025. Company announcement of an officially assessed result.
[12] Mohammed Abouzaid, Nikhil Srivastava, Rachel Ward and Lauren Williams (First Proof). First Proof: Second Batch. arXiv 2606.18119, 16 June 2026. Preprint.
[13] Christophe Pennetier and André-Benoît De Jaegere. Mathematics (field chapter). Generative Engineering, the original study, researched November 2024. The original study.
[14] Trieu H. Trinh, Yuhuai Wu, Quoc V. Le, He He and Thang Luong. Solving olympiad geometry without human demonstrations. Nature 625, 476–482, 17 January 2024. Peer-reviewed study.
[15] Google DeepMind. AI achieves silver-medal standard solving International Mathematical Olympiad problems. Google DeepMind, 25 July 2024. Company announcement.
[16] Thomas Hubert et al.. Olympiad-level formal mathematical reasoning with reinforcement learning. Nature 651, 607–613 (published online 12 November 2025), 2026. Peer-reviewed study.
[17] Christophe Pennetier. Bridging Mathematics and Logistics – The World Model Analogy. Aquia, Collected, 5 January 2026. Author's essay.
[18] Alex Davies et al.. Advancing mathematics by guiding human intuition with AI. Nature 600, 70–74, 2021. Peer-reviewed study.
[19] Christophe Pennetier. The Three Minds of the Machine. Aquia, Collected, 9 January 2026. Author's essay.
[20] Kaiyu Yang, Gabriel Poesia, Jingxuan He et al.. Formal Mathematical Reasoning: A New Frontier in AI. arXiv 2412.16075, 20 December 2024. Preprint, position paper.
[21] Lean. Axioms. Lean Language Reference, accessed 25 October 2026. Technical documentation.
[22] Charles L. Fefferman. Existence and Smoothness of the Navier–Stokes Equation. Clay Mathematics Institute, official problem description, 2000. Institution.
[23] Clay Mathematics Institute. Navier–Stokes Announcement. Clay Mathematics Institute, 11 September 2026. Institution.
[24] Clay Mathematics Institute. Poincaré Conjecture. Clay Mathematics Institute, Millennium Prize Problems, prize awarded 18 March 2010. Institution.

[25] The Associated Press. Russian mathematician rejects $1 million prize. NBC News, 1 July 2010. Press report.
[26] Anima Anandkumar. Stable singularity of the Euler equations on R^3. guest post on What's New, 10 September 2026. Researcher's account.
[27] Terence Tao. Finite time blowup with smooth forcing term for the incompressible porous medium, Boussinesq, and incompressible Euler equations. What's New, 7 September 2026. Expert assessment.
[28] William P. Thurston. On proof and progress in mathematics. Bulletin of the American Mathematical Society 30 (2), 161–177, 1994. Peer-reviewed essay.
[29] Albert Einstein, letter to Jacques Hadamard, in John D. Norton. How Einstein Thought. University of Pittsburgh (letter published in Hadamard, The Psychology of Invention in the Mathematical Field, 1945), 1945. Primary testimony with scholarly commentary.
[30] Henri Poincaré. Mathematical Creation. The Monist (1910), from Science et méthode (1908); Henri Poincaré Papers, Université de Lorraine, 1910. Primary account.
[31] Evelina Fedorenko, Steven T. Piantadosi and Edward A. F. Gibson. Language is primarily a tool for communication rather than thought. Nature 630, 575–586, 2024. Peer-reviewed review.
[32] Wes Gurnee and Max Tegmark. Language Models Represent Space and Time. arXiv 2310.02207; ICLR 2024, 2023. Peer-reviewed conference paper.
[33] Nicolas Fuss. Eulogy of Leonhard Euler. Imperial Academy of Sciences, Saint Petersburg; translation hosted by MacTutor, 1783. Historical testimony.
[34] J. J. O'Connor and E. F. Robertson. Leonhard Euler. MacTutor History of Mathematics, University of St Andrews, accessed 25 October 2026. Scholarly history.
[35] J. J. O'Connor and E. F. Robertson. Lev Semenovich Pontryagin. MacTutor History of Mathematics, University of St Andrews, accessed 25 October 2026. Scholarly history.
[36] Marie Amalric, Isabelle Denghien and Stanislas Dehaene. On the role of visual experience in mathematical development: Evidence from blind mathematicians. Developmental Cognitive Neuroscience 30, 314–323, 2018. Peer-reviewed study.
[37] Fei-Fei Li. From Words to Worlds: Spatial Intelligence is AI's Next Frontier. Dr. Fei-Fei Li, on Substack, 10 November 2025. Researcher's essay.
[38] World Labs. Marble: A Multimodal World Model. World Labs, 12 November 2025. Company announcement.
[39] AMD. AMD to Acquire World Labs to Advance the Future of AI Compute. AMD Newsroom, 28 September 2026. Company announcement.
[40] Anna Heim. Yann LeCun's AMI Labs raises $1.03B to build world models. TechCrunch, 9 March 2026. Press report.
[41] AMI Labs. Research and company description. AMI Labs, accessed 25 October 2026. Company statement.
[42] Mahmoud Assran et al.. V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. arXiv 2506.09985, 2025. Preprint.
[43] Maria Deutscher. World model startup General Intuition closes $220M investment. SiliconANGLE, 29 September 2026. Press report.
[44] Google DeepMind. Genie 3: A new frontier for world models. Google DeepMind, 5 August 2025. Company announcement.
[45] Randall Balestriero and Yann LeCun. LeJEPA: Provable and Scalable Self-Supervised Learning Without the Heuristics. arXiv 2511.08544, November 2025. Preprint.
[46] Zongyi Li et al.. Fourier Neural Operator for Parametric Partial Differential Equations. arXiv 2010.08895; ICLR 2021, 2020. Peer-reviewed conference paper.
[47] Kimm Fesenmaier, Caltech Magazine. Caltech Professor and Alum Among TIME 100 Most Influential People in AI. Caltech Magazine, 31 August 2026. Institution.
[48] Richard P. Feynman, Robert B. Leighton and Matthew Sands. The Theory of Gravitation. The Feynman Lectures on Physics, Volume I, Chapter 7, 1963. Book.
[49] Christophe Pennetier. How to Build an AI Mathematician. Aquia, Collected, 8 December 2025. Author's essay.

[50] G. H. Hardy. A Mathematician’s Apology. Cambridge University Press, 1940. Book.
[51] Chris Isidore, CNN. Elon Musk unveils the Roadster, a new Tesla he says is a flying car. CNN, via ABC11, 16 October 2026. Press report.
[52] Edmond Rostand. Cyrano de Bergerac, Act I, scene 4. Wikisource, 1897. Play.
[53] Wiktionary. fūtilis; with the English entries vain, otiose, gratuitous and school. Wiktionary, read 25 October 2026. Collaborative dictionary.
[54] Abraham Flexner. The Usefulness of Useless Knowledge. Harper’s Magazine; text hosted by the Institute for Advanced Study, 1939. Essay.
[55] Joel Lehman and Kenneth O. Stanley. Abandoning Objectives: Evolution through the Search for Novelty Alone. Evolutionary Computation 19 (2), 189–223, 2011. Peer-reviewed study.
[56] Christophe Pennetier. Neural Banditos: A Multi-Armed Quest for the Shape of Thought. Aquia, Collected, 5 December 2025. Author’s essay.
[57] Isaiah Berlin. The Hedgehog and the Fox: An Essay on Tolstoy’s View of History. Isaiah Berlin Online, University of Oxford, 1953. Book.
[58] Freeman Dyson. Birds and Frogs. Notices of the American Mathematical Society 56 (2), 212–223, February 2009. Essay.
[59] W. T. Gowers. The Two Cultures of Mathematics. in Mathematics: Frontiers and Perspectives, American Mathematical Society, 2000. Essay.
[60] OpenAI. Sharing AI progress in mathematics. OpenAI, 6 October 2026, updated 15 October. Company research release.
[61] Twenty-five Fields Medalists, including Terence Tao. A Severe Misalignment of AI in Mathematics. What’s New (Terence Tao’s blog), 11 September 2026. Collective statement.
[62] AGMAI. Responsible Release of AI-Generated Mathematics. Advisory Group on Mathematics and Artificial Intelligence, 29 September 2026. Advisory recommendations.
[63] Association for Human Mathematics. AHM Statement on OpenAI’s October 6 Release of Mathematical Documents. guest post on What’s New, 7 October 2026. Advocacy statement.
[64] Tsinghua University News. Yau Mathematics and Artificial Intelligence Laboratory Alliance established at Tsinghua (□□□□□□□□□□□□□□□□□□□□□□□□□□□□). Tsinghua University, 13 October 2026. Institution.
[65] Long Yun, Science and Technology Daily, republished by Tsinghua University. Yau-led alliance focuses on AI and mathematics (□□□□□□□□□□□□□ AI □□□□). Tsinghua University, 14 October 2026. Press report.
[66] Z. Z. Ren et al.. DeepSeek-Prover-V2: Advancing Formal Mathematical Reasoning via Reinforcement Learning for Subgoal Decomposition. arXiv 2504.21801, 2025. Preprint, open model.
[67] Yong Lin et al.. Goedel-Prover-V2: Scaling Formal Theorem Proving with Scaffolded Data Synthesis and Self-Correction. arXiv 2508.03613, 2025. Preprint, open model.
[68] Simons Foundation. Simons Foundation Launches $15M Effort To Help Mathematicians Navigate AI Era. Simons Foundation, 15 October 2026. Foundation announcement.
[69] Christophe Pennetier. The Theorem Does Not Care. Aquia, Collected, 18 October 2026. Author’s essay.
[70] Abhinav Nellore, Johannes Schmitt, Soledad Villar and Rachel Ward. oai-math-revised: a community hub for revising OpenAI’s math write-ups. guest post on What’s New, 24 October 2026. Community project announcement.
[71] Caltech Mathathon organizers, critics and community members. Joint Statement about Mathathon. guest post on What’s New, 28 September 2026. Joint statement.
[72] Christophe Pennetier. The Answer Factory. Aquia, Collected, 9 September 2026. Author’s essay.