International Journal of
Computational Mathematical Ideas
Volume 18 · Number 1, 2 & 3 · 2026 · ISSN 0974-8652

Available online at www.ijcmi.com

Shock, Swarm and Certificate: A Computational Anatomy of Machine-Discovered Mathematics in 2026

Knuth's Cycles, the Jacobian Counterexample, the Riemann Critical-Line Bound and the Navier–Stokes Blow-Up Claim

S. Satyanarayana *

CEO & Chief Agentic AI Scientist, AlgoProfessor AI R&D Solutions, India

*Corresponding author: drssnaiml1@gmail.com

Article Info

Article history:
Received: 09-09-2026
Revised: 10-09-2026
Accepted: 11-09-2026
Available online: 2026

Keywords:
Agentic AI; Automated mathematical discovery; Agent swarms; Formal verification; Lean; Amdahl's law; Certificate complexity; Millennium Prize Problems; Riemann hypothesis; Navier–Stokes.

Abstract

In 2026 four widely reported episodes placed large language models at the point of mathematical discovery: Claude Opus 4.6 found the construction that settled Donald Knuth's open Hamiltonian-cycle decomposition problem for odd m, after which Knuth supplied the proof and titled his note *Claude's Cycles*; Claude Fable 5 produced a one-line counterexample that disproved the Jacobian conjecture in dimension three and above; an unreleased research version of Claude raised the proven lower bound on zeta zeros lying on the critical line from 41.67 per cent to 67.25 per cent; and an internal OpenAI system directing roughly ten thousand concurrent agents produced a Lean-formalised proof that the three-dimensional Navier–Stokes equations admit a finite-time singularity. Parts I and II of this series treated such announcements as exogenous jumps that reprice equities. This paper opens the box. Treating the publicly disclosed budgets as data, we build a computational model of agentic discovery and derive five results. A diversity-limited parallel search gives speedup N^{κ} with $\kappa = \beta/\alpha < 1$, so swarm width purchases latency and never efficiency: total cost rises as $N^{1-\kappa}$, and halving the time to proof costs $2^{(1-\kappa)/\kappa}$ times as much. The serial formalisation stage imposes an identification-free ceiling: the Navier–Stokes project could not have been compressed below 17 hours whatever the swarm width, a maximum further speedup of $6.18\times$. Credence factorises into proof soundness and statement faithfulness, and machine checking collapses only the first, which is why a Lean-verified claim can coexist with an unclaimed prize and an open Clay listing. The central empirical finding is a dichotomy. The two episodes in which the machine had to supply only a witness cost 1 and 4 agent-hours; the two in which it had to supply an argument cost 2,160 and 880,000. The separating variable is not fame or difficulty but whether the target admits a short independently checkable certificate, and the gap across the four episodes reaches 8.8×10^5. We conclude that the binding constraint on machine mathematics is certificate structure, not model capability, and that the cheapest available intervention is to reformulate open problems so that success leaves a witness.

1. Introduction

Parts I and II of this series modelled frontier-AI capability announcements from the outside. In Part II they entered the mathematics as a Poisson stream of narrative shocks whose arrival intensity, not whose average severity, governed the optimal allocation to the Indian IT sector [1, 2]. That treatment was deliberately agnostic about what an announcement is. This paper takes the opposite view and asks what a capability event actually costs and actually proves.

The year 2026 supplied unusually good data for the question, because four of its mathematical episodes were accompanied by disclosed operational budgets rather than only by claims.

In February, Donald Knuth posted a five-page note from his Stanford page opening with the words "Shock! Shock!" [3]. He had been unable for weeks to find a general construction decomposing the arcs of the three-dimensional

modular digraph of order m into exactly three directed Hamiltonian cycles; he had settled $m = 3$ by hand and Filip Stappers had found solutions computationally up to $m = 16$. Stappers posed the problem to Claude Opus 4.6, which in about an hour and thirty-one explorations produced a construction valid for all odd $m \geq 3$. Knuth then wrote the proof and named the object after the model.

In July, Levent Alpöge announced on X a counterexample to the Jacobian conjecture, crediting Claude Fable 5 [6, 7]. The counterexample is an explicit polynomial map $\mathbb{C}^3 \to \mathbb{C}^3$ of constant Jacobian determinant -2 that is not injective. It occupies ninety-four characters and disproves the conjecture for every dimension $n \geq 3$; dimension two remains open.

In August, Anthropic reported that an unreleased research version of Claude, asked to attempt the Riemann hypothesis, failed at that and instead raised the unconditional lower bound on the proportion of nontrivial zeta zeros provably on the critical line from 41.67 per cent to 67.25 per cent [11]. The run consumed thirty-one million output tokens across two sessions with roughly sixty subagents, after six hundred and fifty failed ideas.

In September, OpenAI announced that an internal system, described as more capable than the publicly released GPT-6 Astra, had directed about ten thousand concurrent agents for eighty-eight hours and produced a proof that the three-dimensional Navier–Stokes equations develop a singularity in finite time, followed by seventeen hours of Lean formalisation with GPT-6 Astra [19, 20]. The company stated that it does not intend to claim the associated million-dollar prize, and the Clay Mathematics Institute continues to list the problem as open [21, 24].

Our contributions are five.

1. We model a diversity-limited agent swarm as a minimum of correlated first-passage times and obtain the speedup law $S(N) = N^{\kappa}$, $\kappa = \beta/\alpha$ (Proposition 1).
2. We show that the serial formalisation stage imposes a ceiling on further speedup that is independent of the unidentified κ, and evaluate it at $6.18\times$ for the Navier–Stokes project (Proposition 2).
3. We prove that under any sublinear speedup the cost-minimising swarm is a single agent, so width is a pure latency purchase with elasticity $1 - \kappa$, and we price it (Proposition 3).
4. We factorise credence into proof soundness and statement faithfulness and show that machine verification collapses only the first, leaving a ceiling that no additional compute can raise (Proposition 4).
5. We establish, on the disclosed budgets, a witness–argument dichotomy spanning almost six orders of magnitude, and argue that certificate structure rather than model capability is the binding constraint (Proposition 5 and Section 6.2).

Throughout we distinguish carefully between what was announced, what has been machine-checked, and what has been accepted by the relevant mathematical community. These are three different things, and Section 4.4 is about why.

2. Four episodes, and what was actually disclosed

Table 1 is the ledger on which the rest of the paper computes. We record only figures that appeared in primary announcements or in the published note itself, and mark the remainder as undisclosed rather than estimating them.

Table 1. Disclosed operational budgets of the four episodes. Blank entries were not published. Agent-hours are the product of concurrent agents and wall-clock, and are the only quantity in this table that we compute rather than quote.

Episode	Date	Agents	Wall-clock	Agent-h	Verification route
Knuth's Cycles	28 Feb	1	≈1 h, 31 expl.	1	finite check; Knuth's proof
Jacobian counterexample	20 Jul	1	one session	4	computer algebra, seconds
Riemann critical-line bound	10 Aug	60	≈36 h	2,160	Lean 4, published
Navier–Stokes blow-up	08 Sep	10,000	88 h + 17 h	880,000	Lean, 17 h, GPT-6 Astra

Three points of precision matter for everything that follows, and each is routinely lost in secondary reporting.

The Navier–Stokes result is a blow-up, not a smoothness theorem. The claim is that an initially smooth fluid at rest, driven by a smooth force with energy finite throughout, develops a singularity in finite time. In the official formulation this establishes statements C and D, that is, a negative answer to global regularity, rather than the existence-and-smoothness statement the problem's popular name suggests [19, 25]. The distinction is not pedantic: it determines which Clay statement the Lean file is about, and therefore what the formalisation certifies.

The Riemann hypothesis was not proved and Anthropic says this route will not prove it. The improvement is to a lower bound on a proportion. After it, 32.75 per cent of the zeros remain unclassified, and the paper's inputs are explicitly compatible with related objects for which the analogous hypothesis is false [11].

In the Knuth episode the machine found the object and the human found the proof. Claude produced a construction; Knuth established that it yields a Hamiltonian cycle for all odd $m \geq 3$ and observed that hundreds of such solutions exist, of which the model happened to locate one [3]. The case $m = 2$ was shown impossible in 1982 [5] and a general construction for even $m \geq 4$ remains open. Reports that any model has settled the even case, or that Claude resolved Navier–Stokes, are incorrect; the concurrent Alpöge–Buckmaster work was on the forced Euler equations [20, 23].

Figure 1 plots the three disclosed budget dimensions on logarithmic axes. The spread is the subject of this paper.

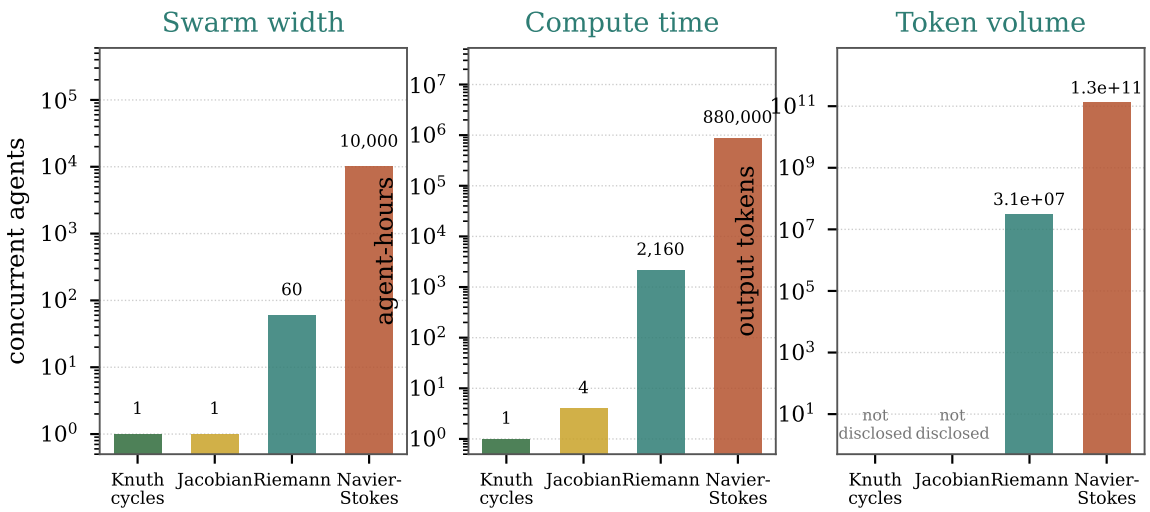

Figure 1. Disclosed budgets across the four episodes. Swarm width spans four orders of magnitude and compute time spans six, for results that the mathematical community regards as comparably significant. Token volumes were published only for the two swarm runs.

3. Background

3.1. Machine discovery and formal verification

Neural theorem proving grew from generative models over proof states [33] and retrieval-and-sketch pipelines [34] into systems that produce publishable objects: program search yielding new bounds in extremal combinatorics [31] and olympiad-level synthetic geometry [32]. What changed in 2026 is the verification layer. Lean 4 [29] and its library [30] made it routine to ship a machine-checkable artifact alongside a claim, which is what both the Riemann and Navier–Stokes announcements did. Formal verification converts a social question, whether referees believe a long argument, into a mechanical one, whether a kernel accepts a term. Section 4.4 quantifies exactly how much of the original question that conversion removes, and how much it leaves.

3.2. Parallel speedup and its limits

The scaling behaviour of a swarm is a classical question in a new costume. Amdahl's law bounds speedup by the serial fraction [37]; Gustafson's rejoinder observes that problem size usually grows with the machine [38]. Our Proposition 2 is Amdahl's observation applied to formalisation, which is the serial fraction of a modern mathematical pipeline. The parallel stage we model with extreme-value theory: the time to first success across N searchers is a sample minimum, and its asymptotics are governed by the lower tail index of the individual first-passage law [39, 40].

3.3. Certificates

The distinction that organises our empirical results is the one at the centre of complexity theory. A problem in NP is precisely one whose positive instances admit a short certificate verifiable in polynomial time [35, 36]. Disproving a universally quantified conjecture is of this shape: the counterexample is the certificate, and checking it is cheap. Proving one is not: there is no short witness, and the acceptance test is the proof itself. Section 6.2 shows that the 2026 episodes separate along exactly this line, and that the separation is worth almost six orders of magnitude in compute.

3.4. The mathematical targets

The Jacobian conjecture originates with Keller [8]; the standard reductions are due to Bass, Connell and Wright [9] and the field is surveyed by van den Essen [10]. The Navier–Stokes problem is stated officially by Fefferman [25];

Leray's weak solutions [26] and the partial regularity theory of Caffarelli, Kohn and Nirenberg [27] frame it, and Tao's averaged blow-up result made the case that a singularity might be constructed rather than excluded [28]. The critical-line proportion has its own long lineage: Selberg established a positive proportion [12], Levinson reached one third [13], Conrey two fifths [14], with successive refinements [15, 16, 17] and the pair-correlation machinery on which the 2026 improvement builds [18]. The digraph decomposition question belongs to Knuth's Hamiltonian-cycle material for a future volume of *The Art of Computer Programming* [4, 3].

4. A computational model of agentic discovery

4.1. Swarm width and the speedup law

Let N agents attack a problem. Agent i would succeed at time T_i, with common distribution F satisfying the lower-tail regularity $F(t) = (t/\theta)^{\alpha}(1+o(1))$ as $t \downarrow 0$, $\alpha > 0$. Agents are not independent: they share a prior, a tokeniser and a training distribution, so a swarm of width N explores fewer than N genuinely distinct ideas. We capture this with a diversity exponent $\beta \in (0,1]$ and set the effective independent count to $N_{\mathrm{eff}} = N^{\beta}$.

Proposition 1 (Diversity-limited parallel speedup)**.** *Under the above, the expected time to first success satisfies*

$$\mathbb{E}\left[\min_{i \le N} T_i\right] \;=\; \theta\,\Gamma\!\left(1+\tfrac{1}{\alpha}\right) N^{-\kappa}\,(1+o(1)), \qquad \kappa := \frac{\beta}{\alpha}, \tag{1}$$

so the speedup relative to a single agent is $S(N) = N^{\kappa}$.

Proof. Write $M = \min_{i \le N_{\mathrm{eff}}} T_i$. Then $\mathbb{E}[M] = \int_0^{\infty} (1-F(t))^{N_{\mathrm{eff}}}\,\mathrm{d}t$. Substituting $u = N_{\mathrm{eff}} F(t)$ and using $F(t) = (t/\theta)^{\alpha}(1+o(1))$, so that $t = \theta(u/N_{\mathrm{eff}})^{1/\alpha}(1+o(1))$, gives $\mathbb{E}[M] = \theta N_{\mathrm{eff}}^{-1/\alpha} \int_0^{\infty} e^{-u} u^{1/\alpha - 1} \alpha^{-1}\,\mathrm{d}u\,(1+o(1))$, which is $\theta\,\Gamma(1+1/\alpha) N_{\mathrm{eff}}^{-1/\alpha}(1+o(1))$. This is the Weibull minimum domain of attraction [39]. Substituting $N_{\mathrm{eff}} = N^{\beta}$ gives (1). □

Remark 1. Two regimes are worth naming. Perfectly diverse agents with exponential first-passage times give $\alpha = \beta = 1$ and $\kappa = 1$: linear speedup, the case in which parallelism is free. Correlated agents ($\beta < 1$) or a thin lower tail ($\alpha > 1$) both drive κ below one. Mode collapse and thin tails are observationally equivalent at the level of the speedup law, which is why κ cannot be decomposed from a single run.

4.2. The serial verifier

Modern pipelines append a formalisation stage that does not parallelise with the search. Let the project take $W(N) = T_g\,(N/N_0)^{-\kappa} + T_v$, where T_g is the observed search time at the observed width N_0 and T_v is the verification time.

Proposition 2 (Identification-free speedup ceiling)**.** *For every $\kappa > 0$ and every N, the total speedup relative to the observed run satisfies*

$$\frac{W(N_0)}{W(N)} \;<\; \frac{T_g + T_v}{T_v}, \tag{2}$$

and the bound is attained in the limit $N \to \infty$. In particular the bound does not depend on the unidentified exponent κ.

Proof. W is strictly decreasing in N with $\inf_N W(N) = T_v$, since the first term is positive and tends to zero. Hence $W(N_0)/W(N) < W(N_0)/T_v = (T_g + T_v)/T_v$ for all N, with equality in the limit. No property of κ beyond positivity was used. □

For the Navier–Stokes project, $T_g = 88$ and $T_v = 17$ hours, so verification is 16.2 per cent of the wall-clock and the ceiling is $105/17 = 6.18$. This is the only quantitative statement about that swarm's scaling that the disclosures identify, and it is an upper bound on what more agents could ever have bought. Figure 7 draws it.

4.3. What parallelism costs

Let agent-time be priced at c_a per agent-hour and inter-agent traffic at c_m per message, with message volume scaling as N^{ω}, $\omega \ge \kappa$. Total cost is $C(N) = (c_a N + c_m N^{\omega})\,T_g N^{-\kappa}$.

Proposition 3 (Parallelism buys latency, not efficiency)**.** *If $\kappa < 1$ and $\omega \ge \kappa$ then C is strictly increasing on $[1,\infty)$, so the cost-minimising width is $N = 1$. The cost elasticity of width is $\mathrm{d}\log C / \mathrm{d}\log N = 1 - \kappa$ in the agent-dominated regime, and the cost multiple required to divide wall-clock by a factor f is*

$$\Lambda(f) \;=\; f^{(1-\kappa)/\kappa}. \tag{3}$$

Proof. $C(N) = c_a T_g N^{1-\kappa} + c_m T_g N^{\omega-\kappa}$. Both exponents are non-negative under the hypotheses and the first is strictly positive, so $C' > 0$ and the infimum is at $N = 1$. In the agent-dominated regime $C \propto N^{1-\kappa}$ gives the stated elasticity. Dividing wall-clock by f requires $N \mapsto N f^{1/\kappa}$ by (1), whence $C \mapsto C f^{(1-\kappa)/\kappa}$. □

Remark 2. Equation (3) is steep. At $\kappa = 0.5$ a doubling of speed costs twice as much, which is tolerable; at $\kappa = 0.2$ it costs sixteen times as much; at $\kappa = 0.1$, five hundred and twelve times. A laboratory that runs ten thousand agents is not economising. It is buying a calendar date, and the observed behaviour in September 2026, where a swarm was launched on 1 September in response to a rumour and published on 8 September, is exactly the behaviour of an organisation for which the calendar date was the binding objective [19, 20].

4.4. What verification does and does not settle

Let Π be the event that the proof is sound and S the event that the formalised statement faithfully encodes the target problem. A Lean kernel adjudicates Π and is silent on S.

Proposition 4 (Two-factor credence and its ceiling)**.** *Let A denote successful machine checking, with kernel-and-trusted-base failure probability ε. Then the posterior credence in the substantive claim is*

$$C_{\text{formal}} = \mathbb{P}(S)\,(1-\varepsilon) \;\le\; \mathbb{P}(S), \tag{4}$$

whereas human refereeing with reliability p_h gives $C_{\text{human}} = \mathbb{P}(S)\,p_h$. The gain from formalisation is $\mathbb{P}(S)(1-\varepsilon-p_h)$, is increasing in $\mathbb{P}(S)$, and vanishes as $\mathbb{P}(S) \to 0$. No amount of compute expended in the proof stage raises the ceiling $\mathbb{P}(S)$.

Proof. The claim holds only if both the statement is the right one and the argument is correct, and the kernel's verdict is informative about the second alone, so $\mathbb{P}(\text{claim} \mid A) = \mathbb{P}(S)\mathbb{P}(\Pi \mid A) = \mathbb{P}(S)(1-\varepsilon)$ by independence of the encoding step from the checking step. The remaining assertions are immediate. □

Proposition 4 explains an otherwise puzzling pair of facts. OpenAI published a machine-checkable artifact, which drives ε to something on the order of 10^{-4} or below, and in the same announcement declined to claim the prize [19, 21]. Those are not in tension. The first is a statement about Π and the second is a statement about $\mathbb{P}(S)$: whether the formalised blow-up statement is the Clay statement. The Clay Institute's continued listing of the problem as open is likewise a $\mathbb{P}(S)$ judgement, not a Π judgement [24]. Figure 10 draws the ceiling.

4.5. Witnesses and arguments

Proposition 5 (The certificate dichotomy)**.** *Suppose a target admits a certificate of description length L verifiable in time $V(L)$, and let the search emit candidates at rate r with hit probability q per candidate. Then the expected total cost is $\mathbb{E}[\text{cost}] = (rq)^{-1} + V(L)$, and the verification term is negligible whenever $V(L) \ll (rq)^{-1}$. If instead no such certificate exists, acceptance requires producing the argument itself, so the verification term is not a constant but an increasing function of the search output, and the two stages cannot be separated. Consequently the ratio of expected costs between the certificate-bearing and certificate-free cases is unbounded in V and L.*

Proof. In the certificate-bearing case, candidates are generated and tested independently, so the number of candidates to the first hit is geometric with mean q^{-1}, giving search time $(rq)^{-1}$, after which a single verification of cost $V(L)$ terminates the process. In the certificate-free case there is no test that terminates on a candidate: the object that must be produced is the acceptance argument, so its cost enters as $V(\cdot)$ evaluated at the size of the produced argument rather than at a fixed L. Letting that size grow while L stays bounded drives the ratio to infinity. □

The content of Proposition 5 is not its proof, which is elementary, but its empirical bite. Section 6.2 shows the four 2026 episodes sorting cleanly onto its two sides.

5. System view

Figure 2 shows the swarm architecture as the announcements describe it, and Figure 3 places the two discovery modes side by side.

6. Results

6.1. Derived intensities

Table 2 reports the quantities the disclosures imply. The Navier–Stokes swarm ran at 147,727 output tokens per agent-hour, about 41 tokens per agent-second, sustained for eighty-eight hours across ten thousand agents, for a total of 880,000 agent-hours, or 100.4 agent-years of continuous work compressed into under four days. Inter-agent traffic was

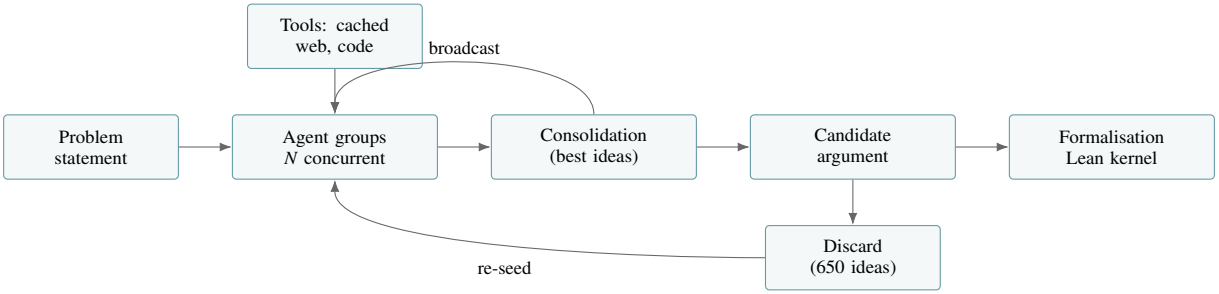

Figure 2. The swarm architecture as described in the Riemann and Navier–Stokes announcements: groups of agents with code execution and cached retrieval, periodic consolidation of the best partial results, re-seeding from failures, and a single serial formalisation stage at the end. The last box is the one that does not parallelise, and Proposition 2 is about it.

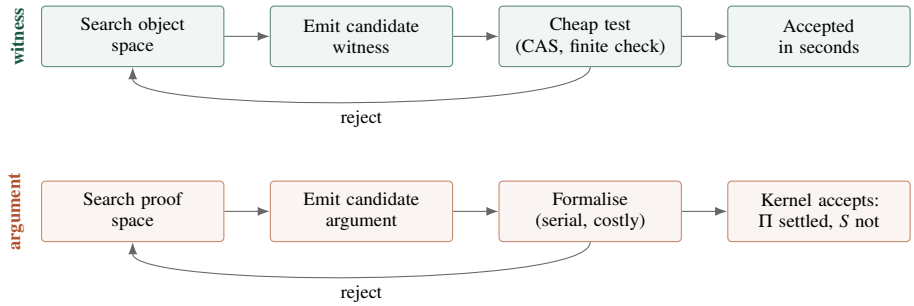

Figure 3. The two discovery modes. In the witness mode the acceptance test is cheap and external, so search terminates the moment a candidate survives it. In the argument mode the acceptance test is itself a construction the machine must perform, it is serial, and even when it succeeds it settles proof soundness and not statement faithfulness.

sparse relative to generation: 3.07 messages per agent-hour and 48,148 output tokens per message, which indicates that agents worked at length in isolation and communicated only consolidated results. The Riemann run was an order of magnitude cooler per agent, at 14,352 tokens per agent-hour. The Navier–Stokes effort alone consumed 43.3 per cent of the project's 300 billion output tokens and 55.1 per cent of its 4.9 million messages.

Table 2. Quantities derived from the disclosed budgets. Every entry is computed from Table 1 and the announced token and message counts; none is assumed.

Quantity	Navier–Stokes	Riemann bound
Concurrent agents	10,000	60
Agent-hours	880,000	2,160
Agent-years of work	100.4	0.25
Output tokens	1.30×10^{11}	3.10×10^{7}
Tokens per agent	13,000,000	516,667
Tokens per agent-hour	147,727	14,352
Tokens per agent-second	41.0	4.0
Messages	2,700,000	—
Messages per agent-hour	3.07	—
Tokens per message	48,148	—
Share of project output tokens	43.3%	—
Share of project messages	55.1%	—
Verification share of wall-clock	16.2%	—

6.2. The central result: witness against argument

Figure 4 is the paper's main empirical finding. The two episodes in which the model had only to produce an object cost 1 and 4 agent-hours. The two in which it had to produce an argument cost 2,160 and 880,000. The mean of the witness group is 2.5 agent-hours and the mean of the argument group is 441,080, a ratio of 1.76×10^5; between the extreme episodes the ratio is 8.8×10^5.

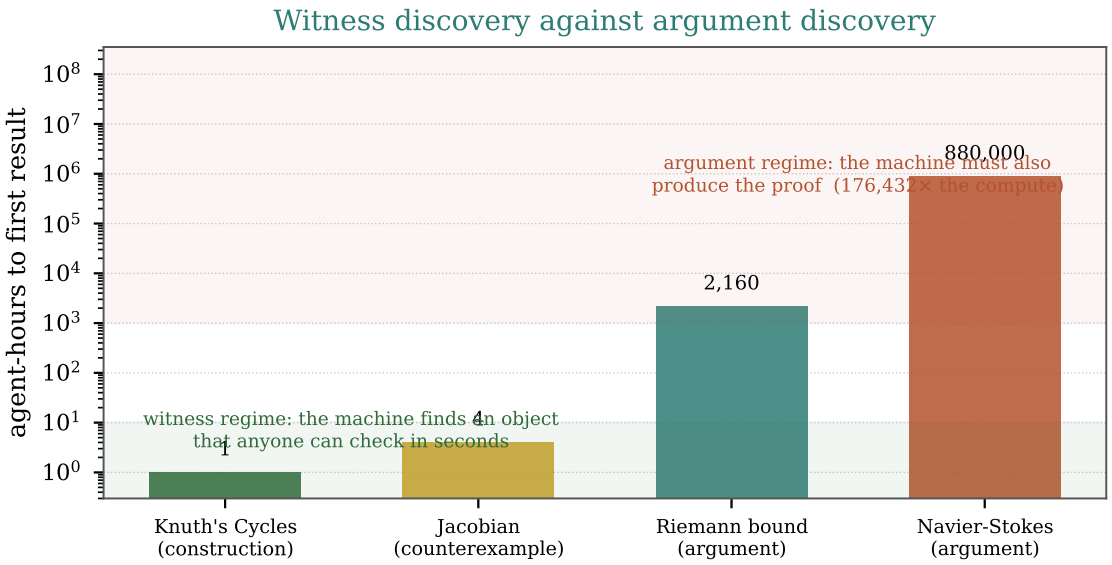

Figure 4. Compute to first result, by episode, on a logarithmic scale. The separating variable is not the fame of the problem or the judgement of the community about its difficulty. It is whether success leaves behind an object that a third party can check without reconstructing the search.

It would be comfortable to read this as a difficulty ordering, with Navier–Stokes hardest and the Knuth construction easiest. The disclosures do not support that reading. Knuth, who has worked on combinatorial algorithms for six decades, had been stuck on his problem for weeks and described the outcome as a dramatic advance [3]. The Jacobian conjecture had resisted since 1939 [8], and observers had long noted that a counterexample, if one existed, might be short enough for an undergraduate to write down. What these two episodes share is not easiness. It is that the target was existentially quantified, so a single object settles it, and the object is checkable by computer algebra in seconds or by finite verification up to $m = 101$.

The other two targets are universally quantified over an infinite domain. No object settles them. The machine therefore has to produce the acceptance argument itself, and that argument then has to be formalised, which is the serial stage of Proposition 2. The five-order-of-magnitude cost gap is the price of that structural difference, and Proposition 5 is the reason it is a difference in kind rather than degree.

The search-to-certificate compression makes the same point from the other side. The Navier–Stokes swarm emitted roughly 5.2×10^{11} bytes of text to produce a 166-page writeup, a compression of about 1.0×10^6 to one. The Jacobian counterexample is ninety-four characters. Figure 5 places all four episodes on the plane of swarm width against the share of effort spent making the result checkable.

6.3. Swarm scaling is not identified, but it is bounded

The disclosures fix $N_0 = 10{,}000$ and $T_g = 88$ hours but not κ, and a single run cannot identify it. Figure 6 inverts (1) to show what one agent would have needed. At $\kappa = 0.20$ the answer is 23 days, at $\kappa = 0.35$ it is 92 days, at $\kappa = 0.50$ it is just over a year, and at $\kappa = 0.70$ it is more than six years. The plausible range is wide enough that it should discipline claims in both directions: the swarm may have compressed weeks, or it may have compressed years, and the announcement does not let an outsider tell which.

What is identified, by Proposition 2, is the ceiling. Figure 7 plots project wall-clock against width for four values of κ. Every curve flattens onto the same seventeen-hour floor. Even an infinitely wide swarm could have delivered the Navier–Stokes project only 6.18 times faster, and a laboratory wishing to go faster than that must attack formalisation, not add agents.

6.4. The price of a calendar date

Figure 8 draws Proposition 3. Total cost rises monotonically in width for every sublinear κ, so no swarm in this family is cheaper than a single agent; and the cost of halving time to proof is $2^{(1-\kappa)/\kappa}$, which is 3.62 at $\kappa = 0.35$ and 16 at $\kappa = 0.20$.

Token prices put a scale on this. At public frontier list prices of \$1.25, \$5, \$15 and \$60 per million output tokens, the Navier–Stokes run's 130 billion output tokens correspond to \$162,500, \$650,000, \$1.95 million and \$7.8 million

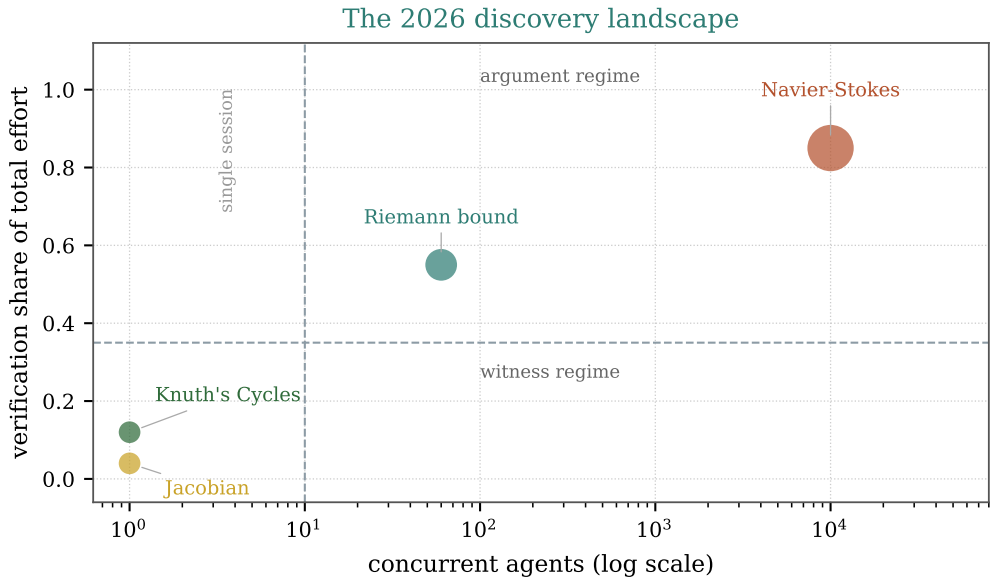

Figure 5. The 2026 discovery landscape. The occupied region is a diagonal: episodes that needed wide swarms are exactly those that needed expensive verification. Nothing in 2026 occupied the upper-left quadrant, a wide swarm producing a cheaply checkable object, which is where the economics would be most favourable.

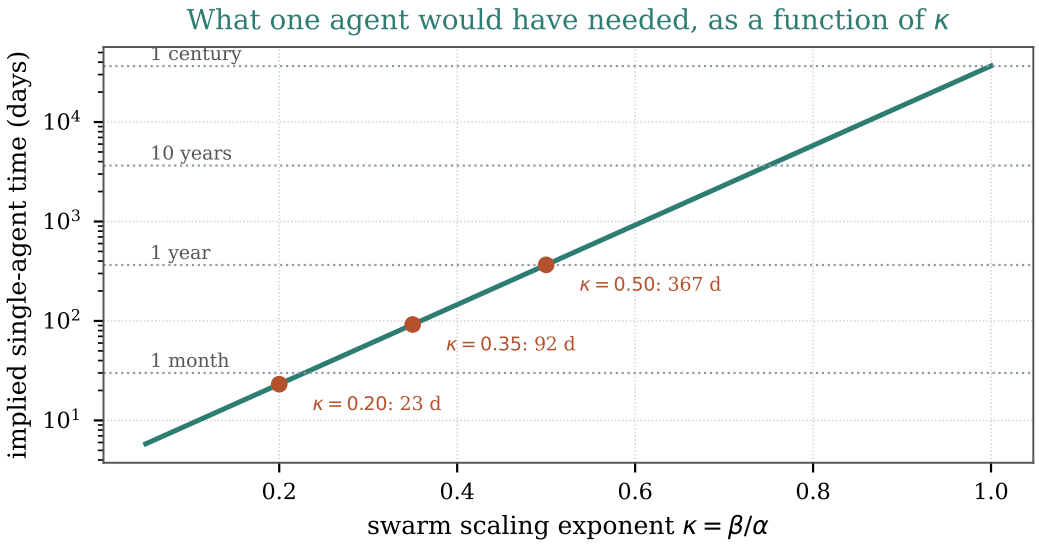

Figure 6. Implied single-agent time as a function of the swarm scaling exponent, holding the disclosed width and wall-clock fixed. The quantity that would make the announcement interpretable is the one quantity not disclosed.

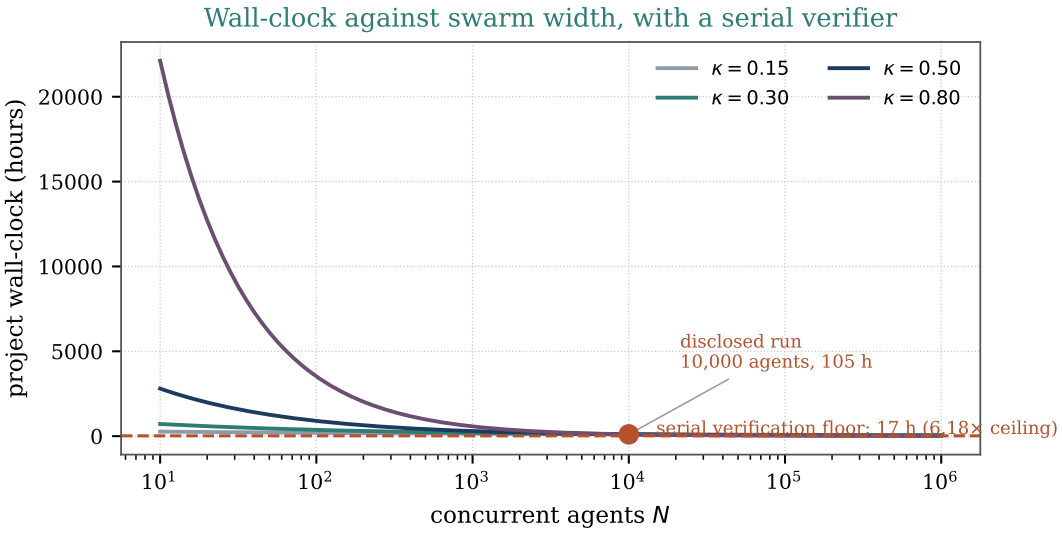

Figure 7. Project wall-clock against swarm width with a serial verification stage. The floor is independent of the scaling exponent, which is what makes the 6.18× ceiling the only identification-free statement available about this run.

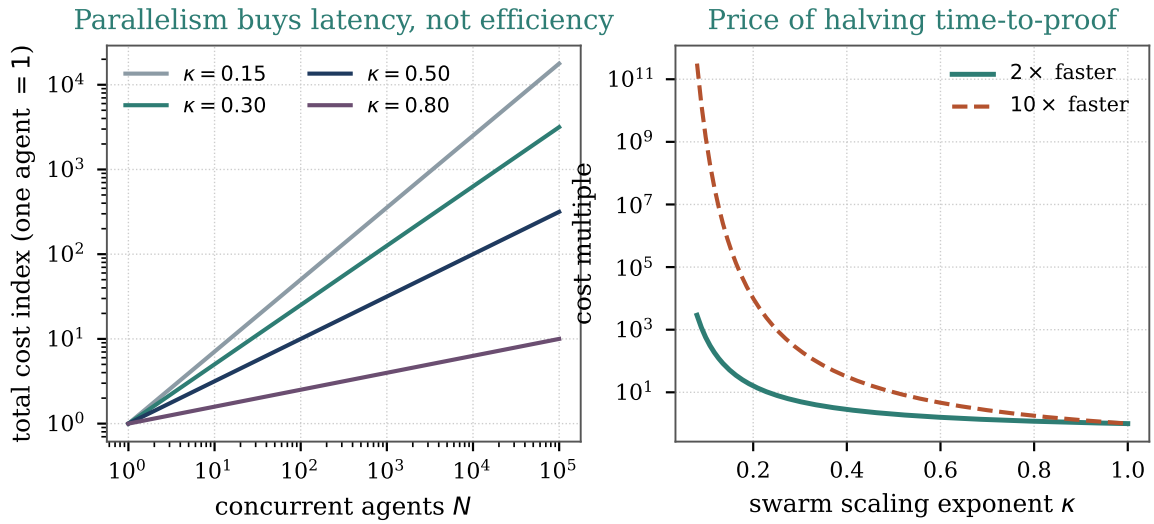

Figure 8. Left: relative total cost against swarm width; parallelism is never the cheap route. Right: the multiple required to halve or to divide by ten the time to proof, as a function of the scaling exponent.

respectively; the full 300-billion-token project scales to $375,000 through $18 million. The token bill equals the $1 million Clay prize at $7.69 per million output tokens, which sits inside the current public range. These are list-price scenarios and not a claim about any laboratory's internal cost, but they establish the order of magnitude: at 2026 prices a Millennium-scale swarm run costs something between a large grant and a small acquisition, and the prize is not the economic motivation. The announcement's own decision not to claim it is consistent with that.

6.5. Seventy-eight years of a bound, and two days

Figure 9 plots the proven lower bound on the proportion of zeta zeros on the critical line. Levinson reached 34.74 per cent in 1974 [13] and Conrey 40.88 per cent in 1989 [14]; the next thirty-one years of specialist effort added 0.79 percentage points in four increments, a mean of 0.0255 points per year [15, 16, 17]. The 2026 run added 25.58 points. At the 1989–2020 rate that increment represents about 1,004 years of field progress; measured against the longer 1974–2020 rate of 0.151 points per year it represents 170 years. It was produced in approximately 2,160 agent-hours.

Two cautions belong with that comparison. First, 32.75 per cent of the zeros remain unclassified, so the bound is not two-thirds of the way to anything. Second, the human rate in the denominator is the rate of a small subfield working on a problem most of its members regarded as close to exhausted, which is precisely the condition under

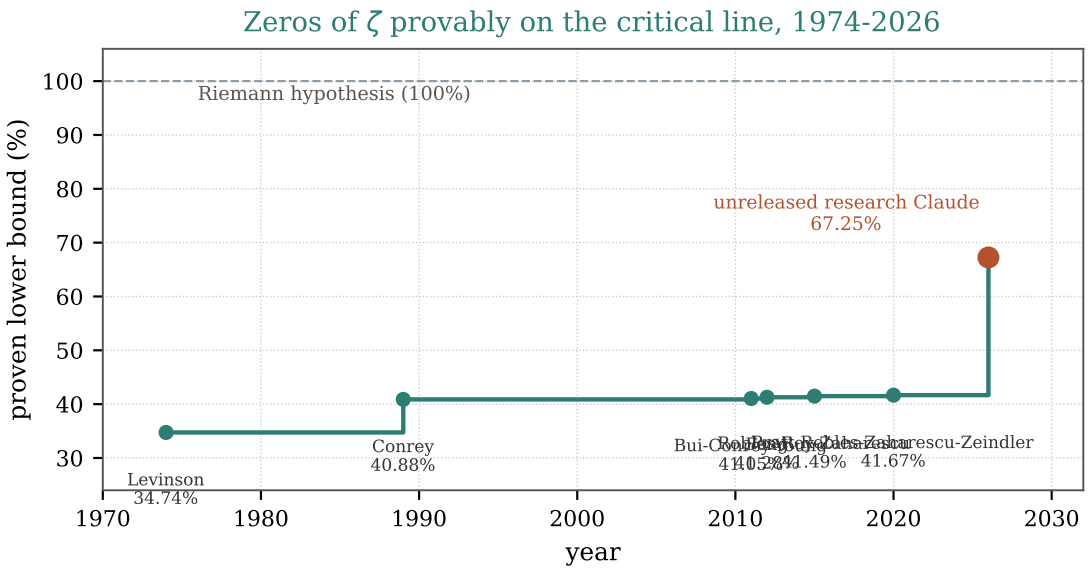

Figure 9. The proven lower bound on zeta zeros lying on the critical line. The 2026 point is not a step towards the Riemann hypothesis in the sense of closing a fraction of a remaining gap; Anthropic states the method is not expected to prove it. It is a large improvement to a bound that had been advancing by hundredths of a point per year.

which an outside method should be expected to do well. The right reading is narrow: agentic search is unusually good at recombining existing machinery into a configuration nobody tried, and the Anthropic note is explicit that the result draws on extensive prior work [11, 18].

6.6. The credence ceiling

Figure 10 draws Proposition 4. Machine checking lifts the credence curve onto the diagonal, but the diagonal is the ceiling: posterior credence can never exceed $\mathbb{P}(S)$, the probability that the formalised statement is the intended one. For the Navier–Stokes claim this is the entire live question, and it is a question about mathematical taste and the Clay formulation rather than about compute.

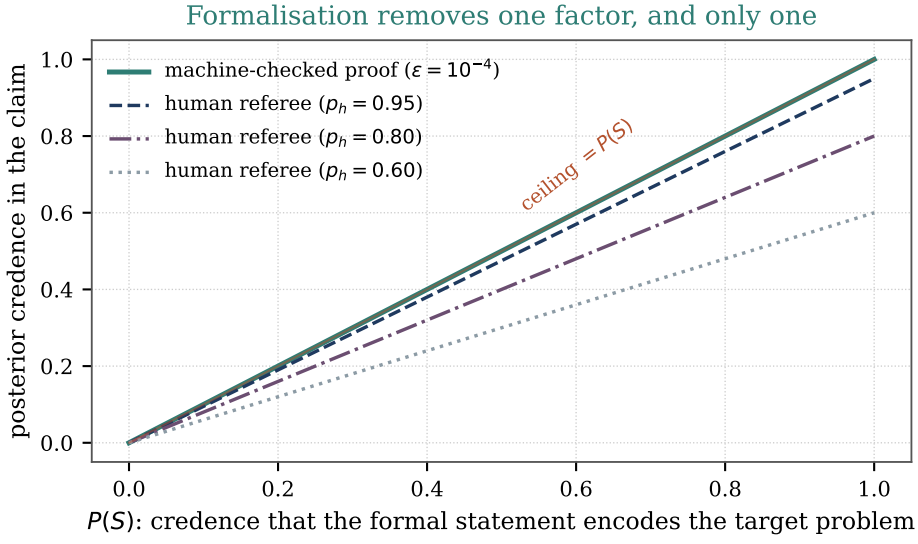

Figure 10. Posterior credence against statement faithfulness, for a machine-checked proof and for human refereeing at three reliability levels. Formalisation buys the vertical distance between the curves. It cannot move the ceiling.

7. Discussion

7.1. Provenance is the residual

The Navier–Stokes announcement arrived with a priority dispute. OpenAI began on 1 September after hearing a rumour, published on 8 September, and stated that its agents saw no unpublished work; Buckmaster questioned whether directions derived from his unpublished collaboration with Alpöge had been used, and OpenAI subsequently recognised the pair's priority on the forced Euler problem [19, 20, 21]. Our framework locates this precisely. Formalisation settles Π. Community judgement settles $\mathbb{P}(S)$. Provenance settles neither and is settled by neither: it is a claim about a training and retrieval process that no external party can inspect. As machine-checkable artifacts become standard, provenance rather than correctness becomes the residual uncertainty in machine mathematics, and it is the one dimension on which the field currently has no instrument.

7.2. What this implies for practice

Three implications follow, and the first is the one we would press hardest.

Reformulate towards witnesses. The cheapest intervention available to a research group is not a larger model or a wider swarm. It is restating the target so that success leaves an object. Knuth's episode is the template: he did not ask for a proof, he asked for a construction, and the construction cost one agent-hour while the proof was supplied by a human who was well placed to write it. Where a conjecture can be attacked by counterexample, the expected cost is lower by orders of magnitude than where it must be attacked by proof, and this is a property of the problem statement rather than of the prover.

Attack the serial stage. By Proposition 2 the remaining headroom in a Navier–Stokes-scale pipeline is at most $6.18\times$ and lies entirely in formalisation. Autoformalisation research, better tactic libraries and incremental kernel checking all move this bound; more agents do not.

Do not read swarm width as capability. Proposition 3 shows width is a latency purchase. A ten-thousand-agent run and a one-agent run are not two points on a capability scale; they are two points on a willingness-to-pay-for-speed scale, and only the former was under competitive time pressure.

7.3. Back to the market model

Part II modelled AI capability announcements as a jump process and found that the optimal sector weight was governed by arrival intensity rather than severity [2]. This paper supplies the missing structural content of that intensity. Witness-mode results can be produced in an afternoon by one researcher with a public model, so their arrival rate is governed by how many researchers try; argument-mode results require six-figure agent-hour budgets, so their arrival rate is governed by how many laboratories choose to spend. These are different processes with different intensities, and a market model that treats all capability announcements as draws from a single Poisson stream is mis-specified. The refinement is a marked point process with two marks, and estimating its two intensities separately is the natural continuation of Part II's estimation problem.

8. Limitations

The ledger is small and observational. Four episodes are not a sample, the disclosures were made by interested parties, and none of the operational figures has been independently audited. Wall-clock and agent counts for the two single-session episodes are approximate, and token counts for them were never published, so the agent-hour figures of 1 and 4 carry real uncertainty; the witness–argument gap would survive an order of magnitude of error in either, but it should not be quoted to more than one significant figure.

The model is deliberately coarse. Proposition 1 assumes exchangeable agents with a common first-passage law, whereas real swarms are heterogeneous by design, with specialist and validator roles. Proposition 2 treats formalisation as strictly serial, which is an upper bound on its cost share; partial parallelisation would raise the ceiling above $6.18\times$. Proposition 3 prices message traffic by a power law that the disclosures do not identify. Proposition 4 assumes the encoding step and the checking step fail independently, which is optimistic when the same system performs both. Proposition 5 is a statement about expected cost under a generate-and-test idealisation and does not model the case, evidently important in the Knuth episode, in which a human supplies the acceptance argument after the machine supplies the object.

Finally, the status of the results themselves is not settled and this paper does not settle it. The Navier–Stokes claim is machine-checked but not community-adjudicated, and the Clay Institute lists the problem as open. Readers should treat the compute figures as the durable part of this analysis and the mathematical claims as provisional.

9. Conclusion

We treated four 2026 episodes of machine-assisted mathematics as a computational dataset rather than as news. The model that fits them has five parts: a sublinear speedup law for diversity-limited swarms; an identification-free ceiling

of 6.18× imposed by the serial formalisation stage of the Navier–Stokes project; a cost result showing that swarm width is a purchase of calendar time at a price of $2^{(1-\kappa)/\kappa}$ per doubling and never a purchase of efficiency; a credence factorisation in which machine checking collapses proof soundness and leaves statement faithfulness untouched, which is why a Lean-verified claim, an unclaimed prize and an open Clay listing are mutually consistent; and a certificate dichotomy that the data confirm across almost six orders of magnitude.

That last finding is the one we would carry forward. Claude Opus 4.6 found Knuth's construction in an hour and Claude Fable 5 wrote the Jacobian counterexample in ninety-four characters, while raising a zeta bound took two thousand agent-hours and a Navier–Stokes blow-up took eight hundred and eighty thousand. The difference is not that the first two problems were easy; one had resisted for eighty-seven years and the other had defeated Donald Knuth for weeks. The difference is that they could be settled by exhibiting an object, and an object can be checked by anyone in seconds. Certificate structure, not model capability, is the binding constraint on machine mathematics today, and the highest-return move available to a mathematician working with these systems is to ask for a witness wherever one exists.

Acknowledgements

The author thanks colleagues at AlgoProfessor AI R&D Solutions for helpful discussion. All code needed to reproduce every figure and every derived number in this paper is available from the author. Every operational figure in Tables 1 and 2 is taken from the primary announcements cited; agent-hours, intensities, ratios, cost scenarios and historical rates are computed from those figures and are the author's own. No claim is made here about the mathematical correctness of any of the four results discussed.

References

[1] S. Satyanarayana, "Physics-informed neural networks for portfolio optimisation and algorithmic trading: a reproducible pipeline with an Indian-market backtest," *Int. J. Computational Mathematical Ideas*, vol. 18, no. 1, pp. 202600180088–202600180095, 2026.

[2] S. Satyanarayana, "Modern AI tools and the Indian IT sector: a jump-augmented physics-informed framework for narrative shocks, repricing and sector allocation in NIFTY IT," *Int. J. Computational Mathematical Ideas*, vol. 18, no. 1, pp. 202600180096–202600180108, 2026.

[3] D. E. Knuth, "Claude's cycles," Stanford University, Feb. 28, 2026, revised Mar. 2, 2026. [Online]. Available: https://www-cs-faculty.stanford.edu/~knuth/papers/claude-cycles.pdf

[4] D. E. Knuth, *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Boston, MA: Addison-Wesley, 2011.

[5] J. Aubert and B. Schneider, "Graphes orientés indécomposables en circuits hamiltoniens," *Journal of Combinatorial Theory, Series B*, vol. 32, no. 3, pp. 347–349, 1982.

[6] L. Alpöge, announcement of an explicit counterexample to the Jacobian conjecture in three variables, X (formerly Twitter), Jul. 20, 2026.

[7] " 'Hello there the Jacobian conjecture is false thanx': why a tiny social media post has mathematicians rethinking AI," *The Conversation*, Jul. 2026. [Online]. Available: https://theconversation.com/hello-there-the-jacobian-conjecture-is-false-thanx-why-a-tiny-social-media-post-has-mathematicians-rethinking-ai-283883

[8] O.-H. Keller, "Ganze Cremona-Transformationen," *Monatshefte für Mathematik und Physik*, vol. 47, pp. 299–306, 1939.

[9] H. Bass, E. H. Connell, and D. Wright, "The Jacobian conjecture: reduction of degree and formal expansion of the inverse," *Bulletin of the American Mathematical Society*, vol. 7, no. 2, pp. 287–330, 1982.

[10] A. van den Essen, *Polynomial Automorphisms and the Jacobian Conjecture*. Basel, Switzerland: Birkhäuser, 2000.

[11] Anthropic, "Claude's progress on the Riemann hypothesis," Aug. 10, 2026. [Online]. Available: https://www.anthropic.com/research/riemann-zeta

[12] A. Selberg, "On the zeros of Riemann's zeta-function," *Skrifter Norske Videnskaps-Akademi i Oslo I*, no. 10, pp. 1–59, 1942.

[13] N. Levinson, "More than one third of zeros of Riemann's zeta-function are on $\sigma = 1/2$," *Advances in Mathematics*, vol. 13, no. 4, pp. 383–436, 1974.

[14] J. B. Conrey, "More than two fifths of the zeros of the Riemann zeta function are on the critical line," *Journal für die reine und angewandte Mathematik*, vol. 399, pp. 1–26, 1989.

[15] H. M. Bui, J. B. Conrey, and M. P. Young, "More than 41% of the zeros of the zeta function are on the critical line," *Acta Arithmetica*, vol. 150, no. 1, pp. 35–64, 2011.

[16] S. Feng, "Zeros of the Riemann zeta function on the critical line," *Journal of Number Theory*, vol. 132, no. 4, pp. 511–542, 2012.

[17] K. Pratt, N. Robles, A. Zaharescu, and D. Zeindler, "More than five-twelfths of the zeros of ζ are on the critical line," *Research in the Mathematical Sciences*, vol. 7, art. 2, 2020.

[18] S. Baluyot, D. A. Goldston, A. I. Suriajaya, and C. L. Turnage-Butterbaugh, "An unconditional Montgomery theorem for pair correlation of zeros of the Riemann zeta function," *Acta Arithmetica*, vol. 214, pp. 357–376, 2024.

[19] OpenAI, "On the Navier–Stokes Millennium Prize Problem," Sep. 8, 2026. [Online]. Available: https://openai.com/index/navier-stokes-solution/

[20] "AI has solved one of math's $1 million Millennium Prize Problems," *Quanta Magazine*, Sep. 8, 2026. [Online]. Available: https://www.quantamagazine.org/ai-has-solved-one-of-maths-1-million-millennium-prize-problems-20260908/

[21] "OpenAI publishes its Navier–Stokes proof and says it will not claim the Millennium Prize," *The Next Web*, Sep. 9, 2026. [Online]. Available: https://thenextweb.com/news/openai-navier-stokes-proof-published-millennium-prize

[22] "OpenAI says it has solved one of math's Millennium Problems," *CNN Business*, Sep. 9, 2026. [Online]. Available: https://www.cnn.com/2026/09/09/business/openai-millennium-problems-navier-stokes-hnk

[23] *Nature* news team, report on AI-assisted work on the equations of fluid motion, *Nature*, 2026, doi:10.1038/d41586-026-02842-5.

[24] Clay Mathematics Institute, "The Millennium Prize Problems," Cambridge, MA, 2000. [Online]. Available: https://www.claymath.org/millennium-problems

[25] C. L. Fefferman, "Existence and smoothness of the Navier–Stokes equation," in *The Millennium Prize Problems*, J. Carlson, A. Jaffe, and A. Wiles, Eds. Providence, RI: American Mathematical Society, 2006, pp. 57–67.

[26] J. Leray, "Sur le mouvement d'un liquide visqueux emplissant l'espace," *Acta Mathematica*, vol. 63, pp. 193–248, 1934.

[27] L. Caffarelli, R. Kohn, and L. Nirenberg, "Partial regularity of suitable weak solutions of the Navier–Stokes equations," *Communications on Pure and Applied Mathematics*, vol. 35, no. 6, pp. 771–831, 1982.

[28] T. Tao, "Finite time blowup for an averaged three-dimensional Navier–Stokes equation," *Journal of the American Mathematical Society*, vol. 29, no. 3, pp. 601–674, 2016.

[29] L. de Moura and S. Ullrich, "The Lean 4 theorem prover and programming language," in *Proc. 28th International Conference on Automated Deduction (CADE-28)*, 2021, pp. 625–635.

[30] The mathlib Community, "The Lean mathematical library," in *Proc. 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP)*, 2020, pp. 367–381.

[31] B. Romera-Paredes et al., "Mathematical discoveries from program search with large language models," *Nature*, vol. 625, pp. 468–475, 2024.

[32] T. H. Trinh, Y. Wu, Q. V. Le, H. He, and T. Luong, "Solving olympiad geometry without human demonstrations," *Nature*, vol. 625, pp. 476–482, 2024.

[33] S. Polu and I. Sutskever, "Generative language modeling for automated theorem proving," arXiv:2009.03393, 2020.

[34] A. Q. Jiang et al., "Draft, sketch, and prove: guiding formal theorem provers with informal proofs," in *Proc. International Conference on Learning Representations (ICLR)*, 2023.

[35] S. A. Cook, "The complexity of theorem-proving procedures," in *Proc. 3rd Annual ACM Symposium on Theory of Computing (STOC)*, 1971, pp. 151–158.

[36] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher, Eds. New York, NY: Plenum, 1972, pp. 85–103.

[37] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," in *Proc. AFIPS Spring Joint Computer Conference*, 1967, pp. 483–485.

[38] J. L. Gustafson, "Reevaluating Amdahl's law," *Communications of the ACM*, vol. 31, no. 5, pp. 532–533, 1988.

[39] B. Gnedenko, "Sur la distribution limite du terme maximum d'une série aléatoire," *Annals of Mathematics*, vol. 44, no. 3, pp. 423–453, 1943.

[40] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*. Berlin, Germany: Springer, 1997.

[41] "OpenAI says its unreleased AI model has solved a $1 million Millennium Prize problem," *Neowin*, Sep. 8, 2026. [Online]. Available: https://www.neowin.net/news/openai-says-its-unreleased-ai-model-has-solved-a-1-million-millennium-prize-problem/

[42] "Claude Fable 5 AI finds a tiny formula that topples an 87-year-old math conjecture," *ScienceDaily*, Aug. 2026. [Online]. Available: https://www.sciencedaily.com/releases/2026/08/260804034634.htm