

AI4SLT: AI for Statistical Learning Theory (in Lean 4)

Fanghui Liu

fanghui.liu@sjtu.edu.cn

*Institute of Natural Sciences, School of Mathematical Sciences
Shanghai Jiao Tong University (SJTU)*

Joint work with Yuanhe Zhang (Warwick) and Jason D. Lee (UC Berkeley)



自然科学研究院
Institute of Natural Sciences

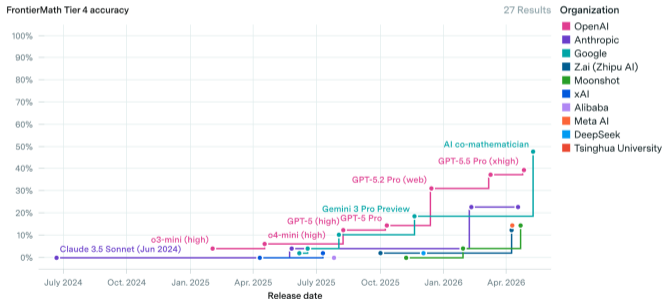


上海交通大学
科学工程计算教育部重点实验室

LLMs can solve IMO, expert-level, open problems...

- 2025: IMO gold medal, university level (PutnamBench)
- 2026: some Erdős problems, Algebra problems (Ju et al., 2026; Zheng et al., 2026)

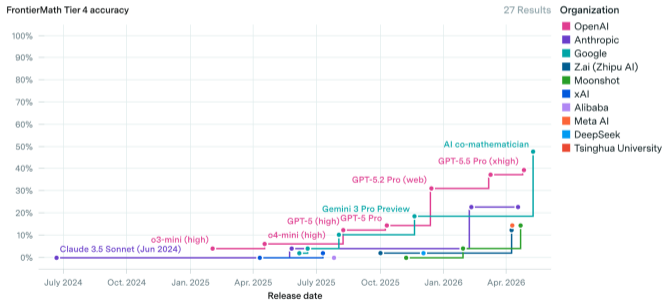
AI performance on the most challenging expert-level mathematics problems



LLMs can solve IMO, expert-level, open problems...

- 2025: IMO gold medal, university level (PutnamBench)
- 2026: some Erdős problems, Algebra problems (Ju et al., 2026; Zheng et al., 2026)
- 2027, 2028...

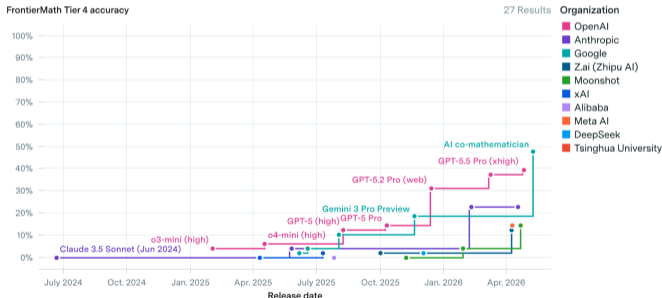
AI performance on the most challenging expert-level mathematics problems



LLMs can solve IMO, expert-level, open problems...

- 2025: IMO gold medal, university level (PutnamBench)
- 2026: some Erdős problems, Algebra problems (Ju et al., 2026; Zheng et al., 2026)
- 2027, 2028...

AI performance on the most challenging expert-level mathematics problems

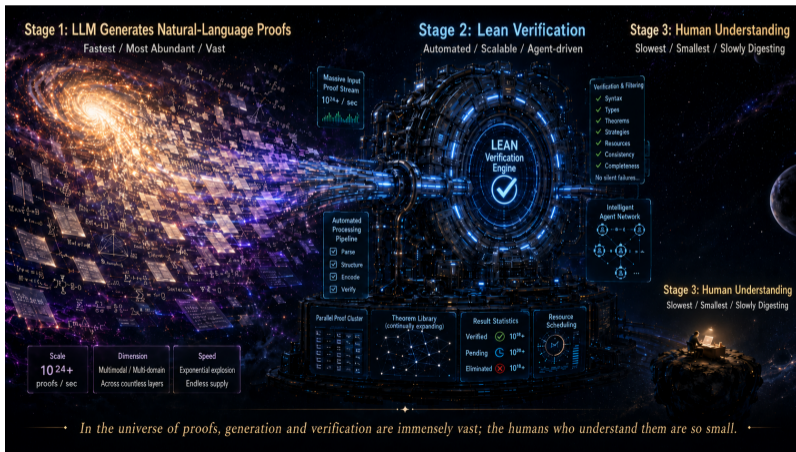
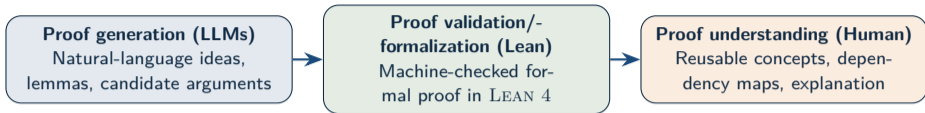


EPOCH AI | CC-BY

epoch.ai

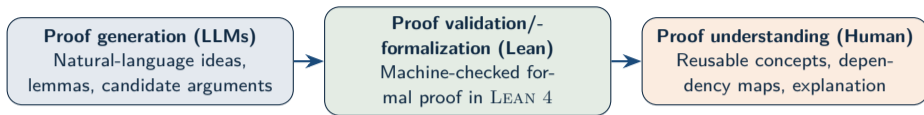
- Traditional coding \Rightarrow Vibe coding
- Traditional (math) proving \Rightarrow Vibe proving...

(Industrial) Revolution in Mathematics



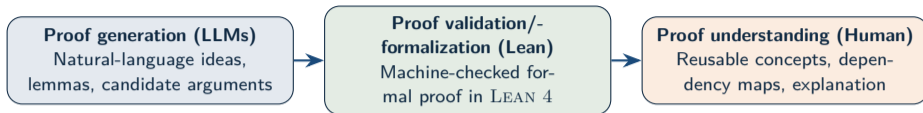
(Generated by GPT5.5)

(Industrial) Revolution in Mathematics



- **Validation is bottleneck**
 - plausibility is not correctness;
 - small gaps can invalidate a proof;
 - informal checking does not scale to libraries.
- **Understanding is the final goal**
 - Math aims to advance human understanding
 - not just a machine for generating formal proofs

(Industrial) Revolution in Mathematics

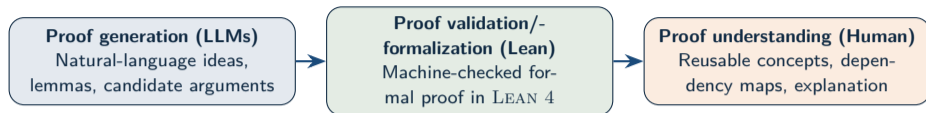


- **Validation is bottleneck**
 - plausibility is not correctness;
 - small gaps can invalidate a proof;
 - informal checking does not scale to libraries.
- **Understanding is the final goal**
 - Math aims to advance human understanding
 - not just a machine for generating formal proofs

Lean (verification) is not a silver bullet but helps **reduce** and **transfer** risk.

- **consistency** vs. **complete** [Gödel's incompleteness theorems]
- statement consistency

(Industrial) Revolution in Mathematics

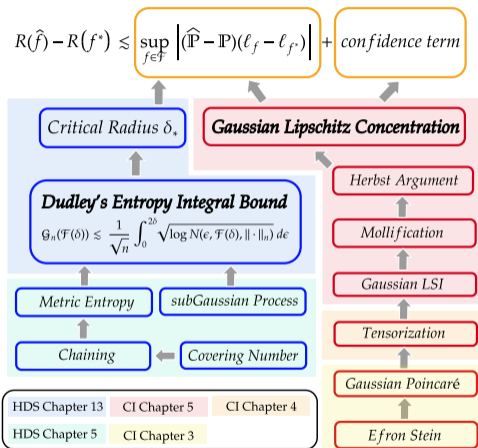


- **Validation is bottleneck**
 - plausibility is not correctness;
 - small gaps can invalidate a proof;
 - informal checking does not scale to libraries.
- **Understanding is the final goal**
 - Math aims to advance human understanding
 - not just a machine for generating formal proofs

Question shift

Can AI suggest a proof? \implies Can **WE** validate, reuse, and help understanding?

Target: **Localized** empirical process framework for SLT



Wainwright (2019) (High-dimensional Statistics, HDS)

Boucheron et al. (2013) (Concentration Inequality, CI)

Concentration

High-probability control of the empirical process fluctuation.

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq t\} \leq 2 \exp\left(-\frac{t^2}{2L^2}\right).$$

Capacity control

Localize around f^* :

$$\mathcal{F}(\delta) = \{f \in \mathcal{F} : d(f, f^*) \leq \delta\},$$

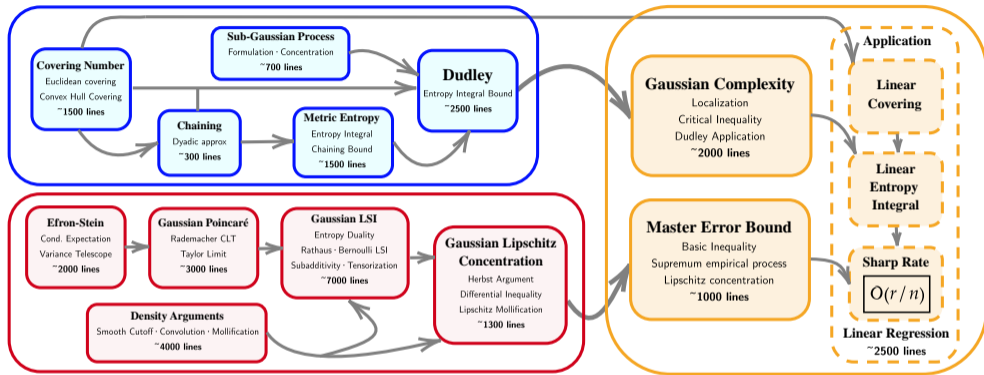
then bound complexity by metric entropy:

$$G_n(\mathcal{F}(\delta)) \lesssim \frac{1}{\sqrt{n}} \int_0^{2\delta} \sqrt{\log \mathcal{N}(\epsilon, \mathcal{F}(\delta), d)} d\epsilon.$$

Natural-language SLT vs. Lean 4 obligations

Mathematical phrase	What Lean asks for
“Take a supremum over a class”	Topology, measurability, boundedness or extended-real formulation
“Apply concentration”	Exact Lipschitz structure, distributional assumptions, integrability
“Use covering numbers”	Metric-space structure, finite nets, cardinal coercions, monotonicity
“Integrate entropy”	\mathbb{R} vs. $\mathbb{R}_{\geq 0} \cup \{\infty\}$ integral interface
“By density”	Approximation sequence, convergence mode, preservation of constants

The dependency graph behind the formalization



Infrastructure gap

For many modern research areas, Mathlib has the foundations but not the domain-specific route from textbook argument to executable proof.

Deliverables

Main implementation

- Gaussian analysis toolbox: Efron–Stein, Gaussian Poincaré, density, Gaussian LSI, Lipschitz concentration.
- Dudley’s entropy integral.
- Localized least-squares framework with linear and ℓ_1 -constrained regression rates.
- Human-AI workflow: humans specify proof architecture; AI agents implement tactical Lean proofs.

Scale

Lean 4 code	~ 30,000 lines
Theorems/lemmas	1000+
Supervised work	~ 500 hours
Task specs	~ 60 preserved
Trust boundary	no sorry or axiom

Deliverables

Main implementation

- Gaussian analysis toolbox: Efron–Stein, Gaussian Poincaré, density, Gaussian LSI, Lipschitz concentration.
- Dudley’s entropy integral.
- Localized least-squares framework with linear and ℓ_1 -constrained regression rates.
- Human-AI workflow: humans specify proof architecture; AI agents implement tactical Lean proofs.

Scale

Lean 4 code	~ 30,000 lines
Theorems/lemmas	1000+
Supervised work	~ 500 hours
Task specs	~ 60 preserved
Trust boundary	no sorry or axiom

Takeaway

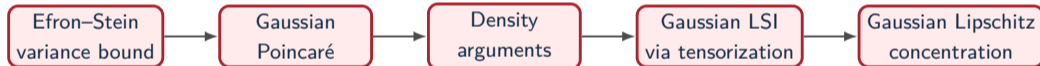
The paper is not only a translation of SLT into Lean: it builds reusable empirical-process infrastructure that was largely missing from the Lean ecosystem.

Gaussian concentration: target theorem

Gaussian Lipschitz concentration

Let $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be L -Lipschitz. Then, for all $t \geq 0$,

$$\mathbb{P}\{f(X) - \mathbb{E}f(X) \geq t\} \leq \exp\left(-\frac{t^2}{2L^2}\right).$$

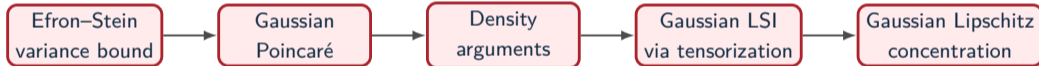


Gaussian concentration: target theorem

Gaussian Lipschitz concentration

Let $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be L -Lipschitz. Then, for all $t \geq 0$,

$$\mathbb{P}\{f(X) - \mathbb{E}f(X) \geq t\} \leq \exp\left(-\frac{t^2}{2L^2}\right).$$



Formalization burden

Each step pulls in different Lean infrastructure: conditional expectation, product measures, CLT limits, Sobolev density, entropy, differentiability, and integration.

Deriving Gaussian Lipschitz Concentration

- **Efron–Stein** gives coordinate-wise variance control:

$$\text{Var}(Z) \leq \sum_{i=1}^n \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right]$$

- **Rademacher CLT + Efron–Stein + Taylor expansion** yield **Gaussian Poincaré**:

$$\text{Var}(f(X)) \leq \mathbb{E} \|\nabla f(X)\|_2^2, \quad \forall f \in C^\infty, X \sim \mathcal{N}(0, I_n)$$

Deriving Gaussian Lipschitz Concentration

- **Efron–Stein** gives coordinate-wise variance control:

$$\frac{d^2}{d\lambda^2} \log \mathbb{E} e^{\lambda Z} |_{\lambda=0} = \text{Var}(Z) \leq \sum_{i=1}^n \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right]$$

- **Rademacher CLT + Efron–Stein + Taylor expansion** yield **Gaussian Poincaré**:

$$\text{Var}(f(X)) \leq \mathbb{E} \|\nabla f(X)\|_2^2, \quad \forall f \in C^\infty, X \sim \mathcal{N}(0, I_n)$$

- **Bernoulli LSI + Tensorization + Entropy duality** yield Gaussian LSI

$$\text{Ent}(g^2) \leq 2 \mathbb{E} \|\nabla g(X)\|_2^2, \quad \forall g \in C^\infty$$

Deriving Gaussian Lipschitz Concentration

- **Efron–Stein** gives coordinate-wise variance control:

$$\frac{d^2}{d\lambda^2} \log \mathbb{E} e^{\lambda Z} |_{\lambda=0} = \text{Var}(Z) \leq \sum_{i=1}^n \mathbb{E} \left[(Z - \mathbb{E}^{(i)} Z)^2 \right]$$

- **Rademacher CLT + Efron–Stein + Taylor expansion** yield **Gaussian Poincaré**:

$$\text{Var}(f(X)) \leq \mathbb{E} \|\nabla f(X)\|_2^2, \quad \forall f \in C^\infty, X \sim \mathcal{N}(0, I_n)$$

- **Bernoulli LSI + Tensorization + Entropy duality** yield Gaussian LSI

$$\text{Ent}(g^2) \leq 2 \mathbb{E} \|\nabla g(X)\|_2^2, \quad \forall g \in C^\infty$$

- **Apply LSI** to $g = e^{\lambda f/2}$. Since $\|\nabla f\|_2 \leq L$,

$$\text{Ent}(e^{\lambda f}) \leq \frac{\lambda^2 L^2}{2} \mathbb{E} e^{\lambda f} \quad \Rightarrow \quad \text{Herbst argument} \quad \log \mathbb{E} e^{\lambda(f - \mathbb{E}f)} \leq \frac{\lambda^2 L^2}{2}.$$

- **mgf bound + (one-side) Chernoff bound** gives the tail.

Example interface: Efron–Stein inequality

Mathematical statement

Let $X = (X_1, \dots, X_n)$ be a vector of n independent random variables and $Z = f(X)$ square-integrable. With $\mathbb{E}^{(i)}$ denoting conditional expectation over all coordinates except i ,

$$\text{Var}(Z) \leq \sum_{i=1}^n \mathbb{E} \left[\left(Z - \mathbb{E}^{(i)}[Z] \right)^2 \right].$$

Leave-one-coordinate conditional expectation

```
noncomputable def condExpExceptCoord
  (i : Fin n) (f : (Fin n -> Omega) -> R) :
  (Fin n -> Omega) -> R :=
  fun x => integral y, f (Function.update x i y) d(mus i)

theorem efronStein
  (f : (Fin n -> Omega) -> R) (hf : MemLp f 2 mus) :
  variance f mus <=
    sum i : Fin n, integral x, (f x - condExpExceptCoord i f x)^2 dmus := by
```

Why is Lean formalization difficult?

We want to prove

$$\text{Law}(X^{(i)}) = \mu_1 \otimes \cdots \otimes \mu_n$$

Key Point

- ❑ **Natural language:** *Resampling one independent coordinate preserves the joint distribution*
- ❑ **Lean:** *Prove that a pushforward of a product measure equals the original product measure*

Why is Lean formalization difficult?

We want to prove

$$\text{Law}(X^{(i)}) = \mu_1 \otimes \cdots \otimes \mu_n$$

Key Point

- ❑ **Natural language:** *Resampling one independent coordinate preserves the joint distribution*
- ❑ **Lean:** *Prove that a pushforward of a product measure equals the original product measure*

product measures measurable maps pushforwards Fubini's theorem measure rectangles.

Why is Lean formalization difficult?

We want to prove

$$\text{Law}(X^{(i)}) = \mu_1 \otimes \cdots \otimes \mu_n$$

Key Point

- ❑ **Natural language:** *Resampling one independent coordinate preserves the joint distribution*
- ❑ **Lean:** *Prove that a pushforward of a product measure equals the original product measure*

product measures measurable maps pushforwards Fubini's theorem measure rectangles.

Design principle for reuse

Formalize intermediate tools as library-quality lemmas, not as one-off proof hacks.

Density arguments (normally missing in textbooks)

Gaussian LSI: from $C_c^\infty(\mathbb{R}^n)$ to $W^{1,2}(\gamma^{\otimes n})$

For $f \in C_c^\infty(\mathbb{R}^n)$, we have

$$\text{Ent}(f^2) \leq 2\mathbb{E} \|\nabla f(X)\|_2^2, \quad X \sim \mathcal{N}(0, I_n).$$

Density arguments (normally missing in textbooks)

Gaussian LSI: from $C_c^\infty(\mathbb{R}^n)$ to $W^{1,2}(\gamma^{\otimes n})$

For $f \in C_c^\infty(\mathbb{R}^n)$, we have

$$\text{Ent}(f^2) \leq 2\mathbb{E} \|\nabla f(X)\|_2^2, \quad X \sim \mathcal{N}(0, I_n).$$

□ Formalization

- Define Gaussian Sobolev space $W^{1,2}(\gamma)$
- Define the Gaussian Sobolev norm
- Prove the density theorem
- Prove convergence of the entropy/gradient term
- Pass the Gaussian LSI to the limit

Density arguments (normally missing in textbooks)

Gaussian LSI: from $C_c^\infty(\mathbb{R}^n)$ to $W^{1,2}(\gamma^{\otimes n})$

For $f \in C_c^\infty(\mathbb{R}^n)$, we have

$$\text{Ent}(f^2) \leq 2\mathbb{E} \|\nabla f(X)\|_2^2, \quad X \sim \mathcal{N}(0, I_n).$$

□ Formalization

- Define Gaussian Sobolev space $W^{1,2}(\gamma)$
- Define the Gaussian Sobolev norm
- Prove the density theorem
- Prove convergence of the entropy/gradient term
- Pass the Gaussian LSI to the limit

Gaussian Lipschitz concentration: from $C_c^\infty(\mathbb{R}^n)$ to Lipschitz

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2 \exp\left(-\frac{t^2}{2L^2}\right)$$

Mollification for Lipschitz functions

For a Lipschitz f , define

$$\rho_\epsilon(x) = \epsilon^{-n} \rho(x/\epsilon), \quad f_\epsilon = f * \rho_\epsilon.$$

The approximation preserves Lipschitz constants and satisfies $f_\epsilon \rightarrow f$ uniformly.

Dudley's entropy integral: theorem target

Mathematical statement

Let (A, d) be a pseudo-metric space, $T \subseteq A$ totally bounded, $\text{diam}(s) \leq D$, and $\{X_t\}_{t \in T}$ a normalized sub-Gaussian process with parameter σ . Under integrability and path-continuity hypotheses,

$$\mathbb{E} \left[\sup_{t \in T} X_t \right] \leq 12\sqrt{2} \sigma \int_0^D \sqrt{\log \mathcal{N}(\epsilon, T, d)} \, d\epsilon.$$

Dudley's entropy integral: theorem target

Mathematical statement

Let (A, d) be a pseudo-metric space, $T \subseteq A$ totally bounded, $\text{diam}(s) \leq D$, and $\{X_t\}_{t \in T}$ a normalized sub-Gaussian process with parameter σ . Under integrability and path-continuity hypotheses,

$$\mathbb{E} \left[\sup_{t \in T} X_t \right] \leq 12\sqrt{2} \sigma \int_0^D \sqrt{\log \mathcal{N}(\epsilon, T, d)} \, d\epsilon.$$

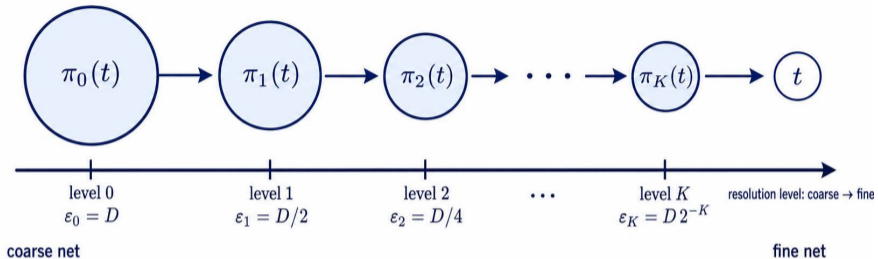
Why it is central

- Converts stochastic complexity into metric geometry.
- Enables localized Gaussian complexity bounds.

Why it is hard in Lean

The proof combines finite nets, sub-Gaussian maximum inequalities, dyadic chaining, Fatou-type limits, and real vs. nonnegative integration APIs.

Chaining

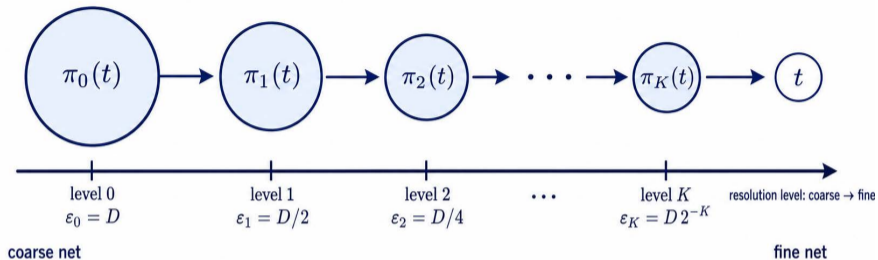


$$\pi_j(t) \in T_j, \quad d(t, \pi_j(t)) \leq \epsilon_j, \quad \epsilon_j = D 2^{-j}$$

$$X_t - X_{\pi_0(t)} = \sum_{j=1}^K (X_{\pi_j(t)} - X_{\pi_{j-1}(t)}) + (X_t - X_{\pi_K(t)})$$

(Generated by GPT5.5)

Chaining



$$\pi_j(t) \in T_j, \quad d(t, \pi_j(t)) \leq \varepsilon_j, \quad \varepsilon_j = D 2^{-j}$$

$$X_t - X_{\pi_0(t)} = \sum_{j=1}^K (X_{\pi_j(t)} - X_{\pi_{j-1}(t)}) + (X_t - X_{\pi_K(t)})$$

(Generated by GPT5.5)

Main proof-engineering issue

The convenient measure-theoretic integral lives in $\mathbb{R}_{\geq 0} \cup \{\infty\}$, while downstream SLT inequalities are stated over \mathbb{R} .

Application to least-squares: Master error bound and critical radius

Critical inequality

A radius δ is admissible if

$$\frac{G_n(\mathcal{F}(\delta))}{\delta} \leq \frac{\delta}{2\sigma}.$$

The smallest positive solution is the critical radius δ^* .

Master theorem

For $t \geq \delta^*$,

$$\text{Prob}(\|\hat{f} - f^*\|_n^2 \geq 16t\delta^*) \leq \exp\left(-\frac{nt\delta^*}{2\sigma^2}\right).$$

Capacity bound via Dudley

For star-shaped \mathcal{F}^* ,

$$G_n(\mathcal{F}(\delta)) \leq \frac{24\sqrt{2}}{\sqrt{n}} I(\delta),$$

$$I(\delta) = \int_0^{2\delta} \sqrt{\log \mathcal{N}(\epsilon, \mathcal{F}(\delta), \|\cdot\|_n)} d\epsilon.$$

Takeaway

After the Lean infrastructure is built, applications reduce to covering-number calculus.

Human-AI collaborative formalization

Structured TASK.md protocol

Each task specification contains:

1. also need NL target statement;
2. tactic-level proof plan (for LLMs);
3. relevant file paths and expected crossref;
4. hard boundaries for what the agent should not rewrite.

Collaborative level

Autonomous	~ 15%
Sketch-guided	~ 30%
Specified under NL proofs	~ 40%
Human-critical	~ 15%

Human-AI collaborative formalization

Structured TASK.md protocol

Each task specification contains:

1. also need NL target statement;
2. tactic-level proof plan (for LLMs);
3. relevant file paths and expected crossref;
4. hard boundaries for what the agent should not rewrite.

Collaborative level

Autonomous	~ 15%
Sketch-guided	~ 30%
Specified under NL proofs	~ 40%
Human-critical	~ 15%

Takeaway

The dominant bottleneck is not tactic writing; it is converting compressed mathematical prose into precise, decomposed proof architecture.

Practical lessons for formalization

- ❑ **Decompose aggressively.** Target small lemmas that fit within one agent context window.

Practical lessons for formalization

- ❑ **Decompose aggressively.** Target small lemmas that fit within one agent context window.
- ❑ **Specify more than the theorem.** Include local APIs, intended proof route.

Practical lessons for formalization

- ❑ **Decompose aggressively.** Target small lemmas that fit within one agent context window.
- ❑ **Specify more than the theorem.** Include local APIs, intended proof route.
- ❑ **Treat persistent failure as mathematical signal.** Recheck the statement and look for missing assumptions or counterexamples.

Practical lessons for formalization

- ❑ **Decompose aggressively.** Target small lemmas that fit within one agent context window.
- ❑ **Specify more than the theorem.** Include local APIs, intended proof route.
- ❑ **Treat persistent failure as mathematical signal.** Recheck the statement and look for missing assumptions or counterexamples.
- ❑ **Make hidden regularity reusable.** Integrability, continuity, nonemptiness, and boundedness hypotheses become reusable theorem contracts.

Practical lessons for formalization

- ❑ **Decompose aggressively.** Target small lemmas that fit within one agent context window.
- ❑ **Specify more than the theorem.** Include local APIs, intended proof route.
- ❑ **Treat persistent failure as mathematical signal.** Recheck the statement and look for missing assumptions or counterexamples.
- ❑ **Make hidden regularity reusable.** Integrability, continuity, nonemptiness, and boundedness hypotheses become reusable theorem contracts.
- ❑ **To agent: don't frequently change the proof!**

Practical lessons for formalization

- ❑ **Decompose aggressively.** Target small lemmas that fit within one agent context window.
- ❑ **Specify more than the theorem.** Include local APIs, intended proof route.
- ❑ **Treat persistent failure as mathematical signal.** Recheck the statement and look for missing assumptions or counterexamples.
- ❑ **Make hidden regularity reusable.** Integrability, continuity, nonemptiness, and boundedness hypotheses become reusable theorem contracts.
- ❑ **To agent: don't frequently change the proof!**
 - Know more details (e.g., implicit assumptions)
 - Know which key infrastructure is reused (structure view, dependence graph)
 - proof connections (e.g., Gaussian Poincaré by Efron-Stein, Hermite decomposition, OU semigroup)

If Mathlib is “complete”, can **we** autoformalize research-level problems even with proof gap?

If Mathlib is “complete”, can **we** autoformalize research-level problems even with proof gap?



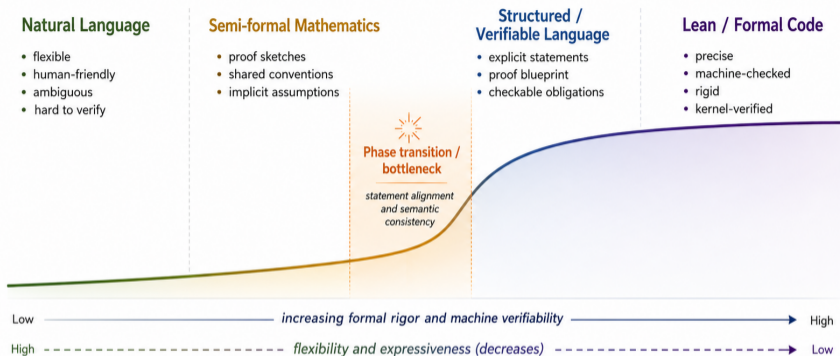
LeanMarathon: Toward Reliable AI Co-Mathematicians through
Long-Horizon Lean Autoformalization

Yuanhe Zhang* Yuekai Sun† Taiji Suzuki‡ Jason D. Lee§ Fanghui Liu¶



Erdős problems (#1051, #1196, #164, #1217) ([Barreto et al., 2026](#); [Alexeev et al., 2026](#))

From Natural Language to Lean: a Verification Phase Transition



Key challenge: the hardest step is often not proof checking, but aligning mathematical intent with formal statements.

(Generated by GPT5.5)

Yuanhe Zhang, Ilja Kuzborskij, Jason D. Lee, Chenlei Leng, Fanghui Liu. DAG-Math: Graph-of-Thought-Guided Mathematical Reasoning in LLMs. ICLR'26.

Conclusion

□ Yuanhe Zhang, Jason D. Lee, Fanghui Liu. *AI4SLT: Empirical Processes in Lean 4 for Formal Statistical Learning Theory*. ICML'26.

What the paper establishes

- End-to-end infrastructure for SLT;
- Gaussian Lipschitz concentration + Dudley entropy integral stack;
- Localized least-squares framework

Why it matters

- Verified reusable SLT toolbox;
- Template for human–AI large-scale formalization;
- **Training ground for students.**

Conclusion

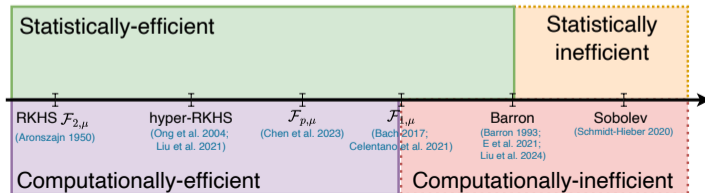
□ Yuanhe Zhang, Jason D. Lee, Fanghui Liu. *AI4SLT: Empirical Processes in Lean 4 for Formal Statistical Learning Theory*. ICML'26.

What the paper establishes

- End-to-end infrastructure for SLT;
- Gaussian Lipschitz concentration + Dudley entropy integral stack;
- Localized least-squares framework

Why it matters

- Verified reusable SLT toolbox;
- Template for human–AI large-scale formalization;
- **Training ground for students.**



Conclusion

□ Yuanhe Zhang, Jason D. Lee, Fanghui Liu. *AI4SLT: Empirical Processes in Lean 4 for Formal Statistical Learning Theory*. ICML'26.

What the paper establishes

- End-to-end infrastructure for SLT;
- Gaussian Lipschitz concentration + Dudley entropy integral stack;
- Localized least-squares framework

Why it matters

- Verified reusable SLT toolbox;
- Template for human–AI large-scale formalization;
- **Training ground for students.**

Q & A

fanghui.liu@sjtu.edu.cn

www.lfhsGRE.org

References

- Boris Alexeev, Kevin Barreto, Yanyang Li, Jared Duker Lichtman, Liam Price, Jibran Iqbal Shah, Quanyu Tang, and Terence Tao. Primitive sets and von Mangoldt chains: Erdős Problem #1196 and beyond, 2026. URL <https://arxiv.org/abs/2605.00301>.
- Kevin Barreto, Jiwon Kang, Sang-hyun Kim, Vjekoslav Kovač, and Shengtong Zhang. Irrationality of rapidly converging series: a problem of Erdős and Graham, 2026. URL <https://arxiv.org/abs/2601.21442>.
- Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. URL <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>.

References ii

- Haocheng Ju, Guoxiong Gao, Jiedong Jiang, Bin Wu, Zeming Sun, Leheng Chen, Yutong Wang, Yuefeng Wang, Zichen Wang, Wanyi He, et al. Automated conjecture resolution with formal verification. *arXiv preprint arXiv:2604.03789*, 2026.
- Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.
- Daniel Zheng, Ingrid von Glehn, Yori Zwols, Iuliya Beloshapka, Lars Buesing, Daniel M Roy, Martin Wattenberg, Bogdan Georgiev, Tatiana Schmidt, Andrew Cowie, et al. AI Co-Mathematician: Accelerating mathematicians with agentic AI. *arXiv preprint arXiv:2605.06651*, 2026.