

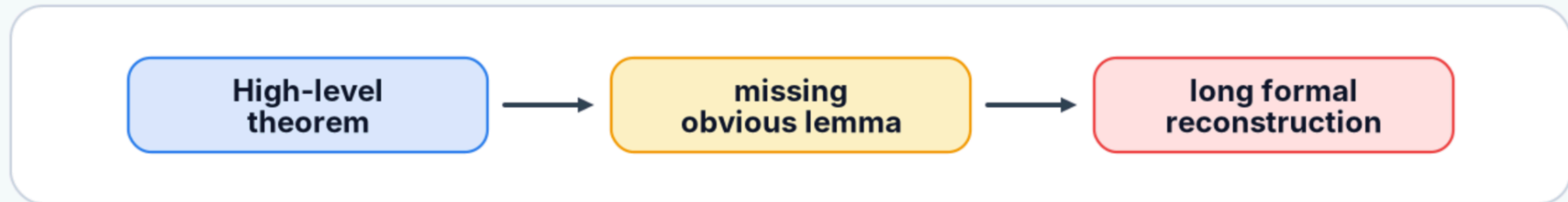
MathlibLemma: Folklore Lemma Generation and Benchmark For Formal Mathematics

Xinyu Liu, Zixuan Xie, Amir Moeini, Claire Chen, Shuze Daniel Liu, Yu Meng, Aidong Zhang, Shangtong Zhang

From using Mathlib to expanding Mathlib

Date: 05/17/2026


Formalization often gets stuck on tiny lemmas that humans treat as obvious



Folklore lemma = a reusable fact mathematicians use without comment, but that may be absent from Mathlib.

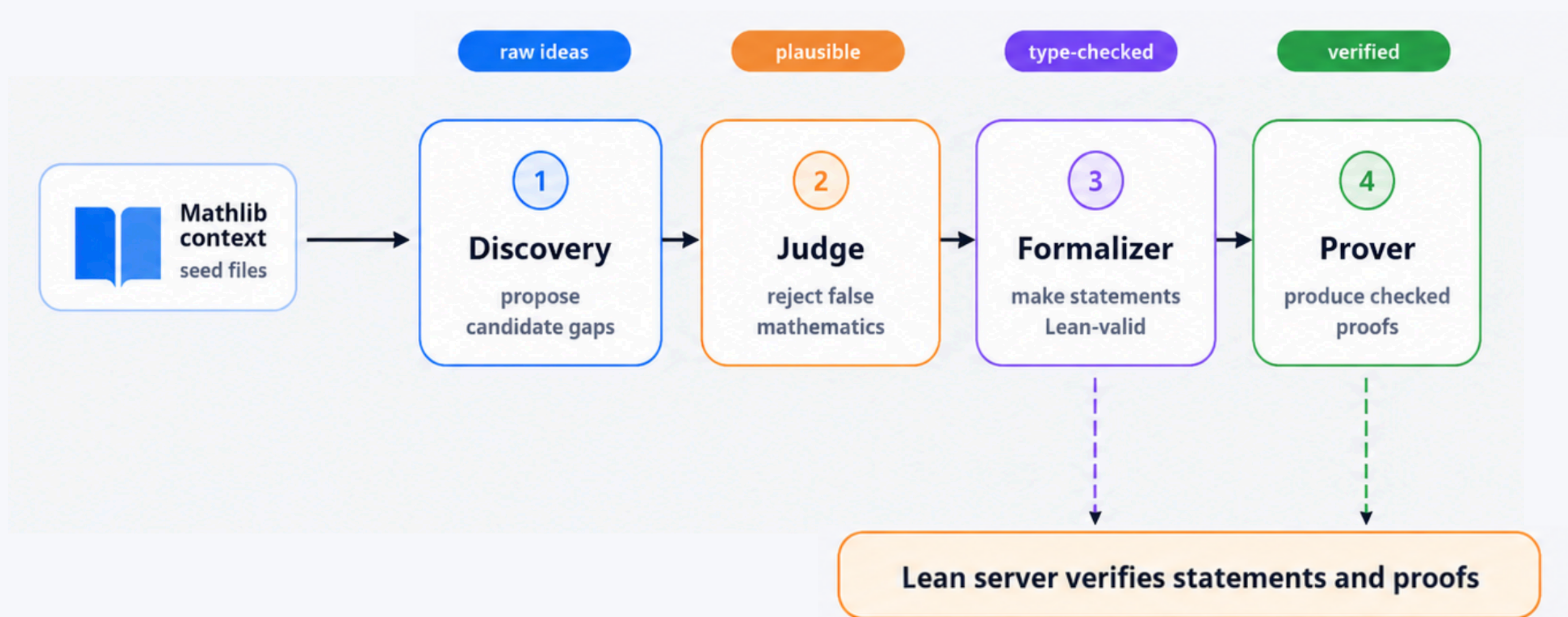
MathlibLemma turn missing folklore facts into a mining task



 **Goal**
Automate library expansion at scale.

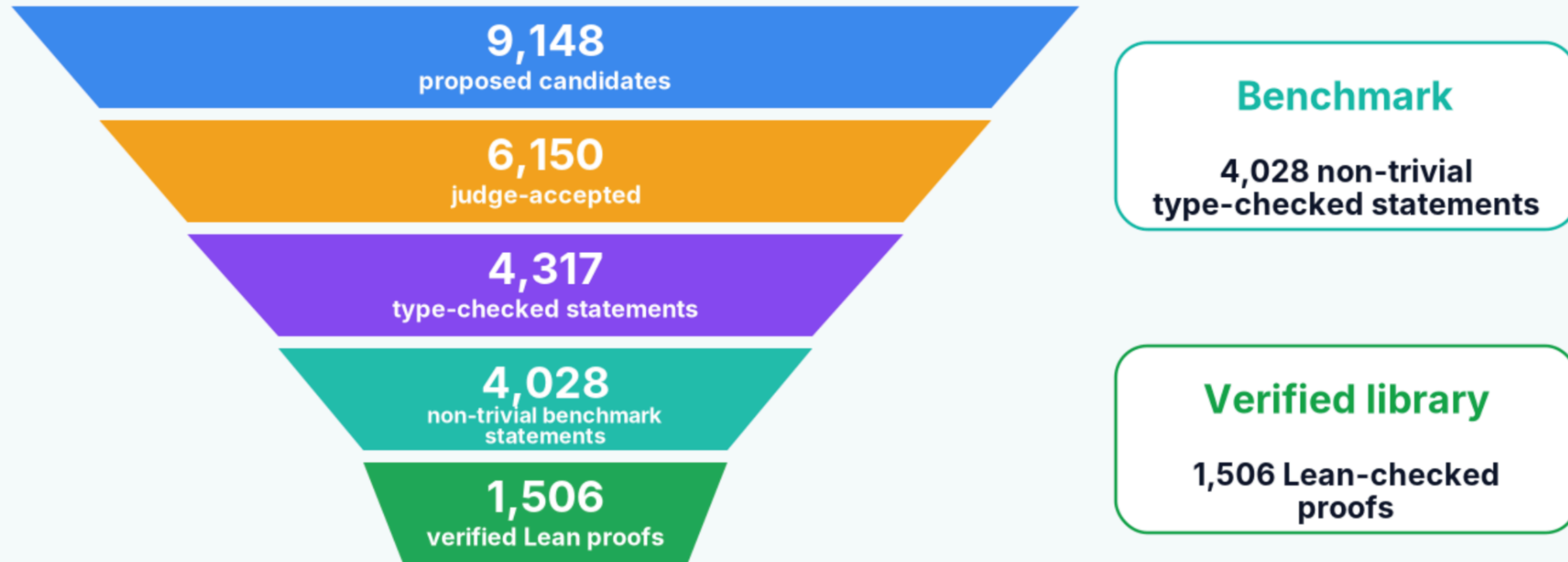
The pipeline separates discovery, filtering, formalization, and proving

Each stage targets a different failure mode: false math, invalid Lean, or failed proof search.



Modularity makes failure modes observable and separately addressable.

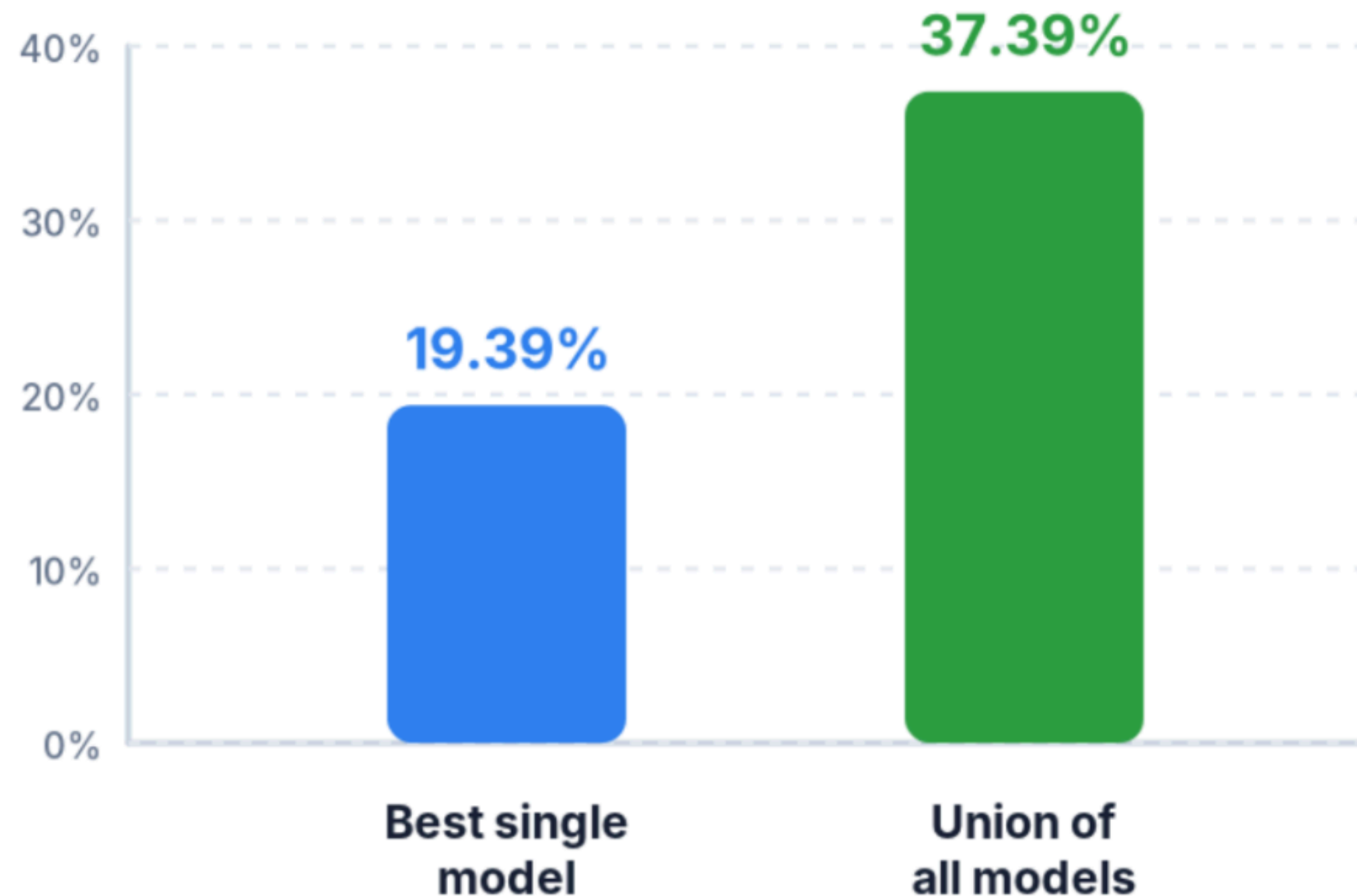
The pipeline turns raw ideas into thousands of type-checked and proved artifacts



Verified proofs pass Lean checking + proof-bypass screening

Current provers solve only part of the benchmark, while many unsolved lemmas are still valid

Closed-book prover success (Success@2)



Residual validity audit



of sampled model-unsolved residuals were human-proved

Benchmark is challenging, but not saturated.

Generated folklore lemmas can become real Mathlib API after human curation

A small upstreaming pilot shows that selected outputs can meet real library standards.



Analysis merged

gronwallBound_mono

Monotonicity for Gronwall bounds

Measure theory merged

Kernel.restrict_const

Constant kernels commute with restriction

Probability merged

centralMoment_congr_ae

Central moments respect a.e. equality

Thank you!

Find our paper: [arXiv:2602.02561v2](https://arxiv.org/abs/2602.02561v2)

Contact us: xinyuliu@virginia.edu

shangtong@virginia.edu