Lean Copilot: Large Language Models as Copilots for Theorem Proving in Lean

Peiyang Song[1], Kaiyu Yang[1], Anima Anandkumar[1]

[1]Caltech

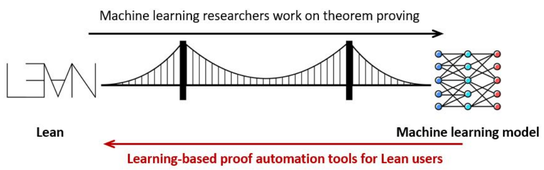
NeuS 2025

AI for (Formal) Mathematics Reasoning

- **Formal reasoning (theorem proving)**
  - Proof can be rigorously checked by computers
  - Formalization and theorem proving are labor-intensive
  - Applications in math and formal verification

Lean's Mathlib **95K** theorems, **1M** lines of code
Analysis, algebra, combinatorics, etc.

- **Theorem proving with machine learning**
  - Exciting recent progress on neural theorem proving
  - But not readily accessible for Lean users!
    - Challenging integration into Lean's workflow
    - Unrealistic requirement of compute power & runtime



- **LLM-aided theorem proving**
  - Fully autonomously vs. *as a copilot*

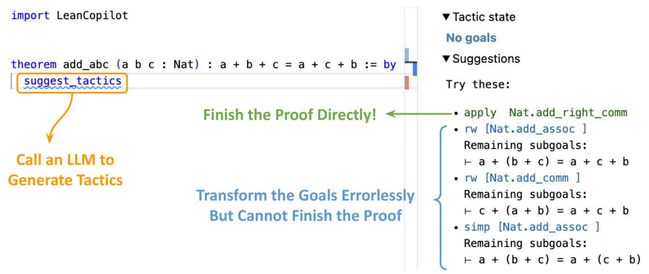
Method	Avg. # human-entered tactics (↓)	% theorems proved autonomously (↑)	Avg. % proof steps automated (↑)
AESOP	3.86	24.4%	40.1%
SUGGEST_TACTICS	3.10	45.2%	58.3%
SEARCH_PROOFS	**2.08**	**63.7%**	**74.2%**

  - Work out of the box w\o changes to Lean's workflow
  - Installed as a Lean package w\o additional setups

- **User experience**
  - Simple usage: everything wrapped up as tactics
  - Respond instantaneously on CPU-only laptops

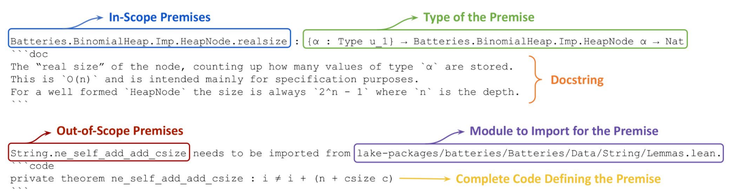
LLM-based Proof Automation

- **Tactic suggestion**
  - Allow to constrain output tactics with a prefix
  - Perform type check and mark with different colors



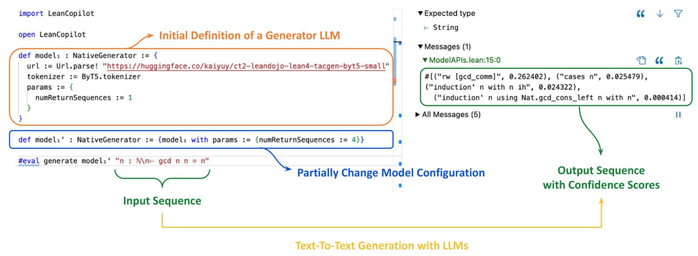
- **Proof search**
  - Combine LLM-generated steps with best-first search
  - Produce multi-step proof certified correct

- **Premise selection**
  - Retrieve useful lemmas from Mathlib
  - Provide type & docstring (if any) information
  - Offer import & code preview for out-of-scope ones

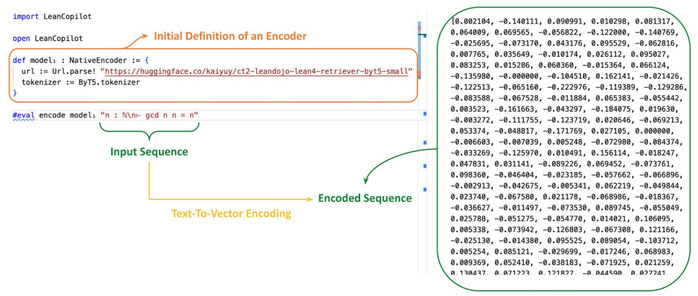


Neural Network Inference in Lean

- **Native neural network inference**
  - Use different models / hyperparameters
  - Support multiple platforms, w\ or w\o GPUs
  - Enable applications not limited to theorem proving
    - Text-to-text generation
      - Underlying tactic suggestion and proof search



- Text-to-vector encoding
  - Underlying premise selection



- **Server process for any models**
  - Allow users to bring their own models
  - Provide a Python API server to run models externally