Yuxin (Shirley) Li

E-mail | Webpage | GitHub | Google Scholar

Computer Science, The Hong Kong University of Science and Technology

RESEARCH INTERESTS

I am interested in building foundations components for LLM reasoning by examining models from two complementary perspectives: formal logic as a normative framework for how models should reason, and human cognition as a descriptive benchmark for how models tend to reason in practice. **My current research focuses on integrating formal languages (Lean4) into LLM systems to enable structured, verifiable reasoning.** I investigate how formal structures can serve as scaffolds for scientific reasoning, including but not limited to mathematics and physics. I am also interested in multi-agent LLM social behavior, including how LLMs exhibit cognitive biases, and develop collective opinions.

EDUCATION

The Hong Kong University of Science and Technology *Sep. 2022 - Jun. 2026(expected)*
BEng. in Computer Science *Advisor: Long Chen*

SELECTED PUBLICATIONS

* indicates equal contribution

[1] **Yuxin Li**, Minghao Liu, Ruida Wang, Wenzhao Ji, Zhitao He, Rui Pan, Junming Huang, Tong Zhang, Yi R. Fung
Lean4Physics: Comprehensive Reasoning Framework for College-level Physics in Lean4
[ICLR 2026]

[2] Ruida Wang*, **Yuxin Li***, Yi. R. Fung, Tong Zhang
Let's Reason Formally: Natural-Formal Hybrid Reasoning Enhances LLM's Math Capability
[EMNLP 2025 (Main)]

[3] Ruida Wang*, Rui Pan*, **Yuxin Li***, Jipeng Zhang, Yizhen Jia, Shizhe Diao, Renjie Pi, Junjie Hu, Tong Zhang
MA-LoT: Model-Collaboration Lean-based Long Chain-of-Thought Reasoning Enhances Formal Theorem Proving [ICML 2025]

RESEARCH EXPERIENCE

Lean4Physics: Comprehensive Reasoning Framework for College-level Physics in Lean4 2025
Role: leading author | *Advisor: Yi R.(May) Fung and Tong Zhang* | *Preprint:* arXiv:2510.26094

- This work proposed *Lean4PHYS*, a reasoning framework for college-level physics problems in Lean4. It includes *LeanPhysBench*, the first benchmark in the field, and *PhysLib*, a community-driven repository that sets the foundation for the field.
- *Lean4PHYS* pioneered the extension of formal reasoning from mathematics to physics; leading large-scale evaluations across 8 frontier LLMs (GPT-4o, Claude 4, Gemini 2.5 Pro, etc.), which yielded an average **improvement of 11.89%** in formal reasoning accuracy with *PhysLib*.
- Designed structure of the reasoning framework and implemented the modular architecture of *LeanPhysBench*, handcrafted the formalization of college-level physics problems, implemented the open-source models inference pipeline, and conducted in-depth analysis to the proof.

Natural-Formal Hybrid Reasoning Enhances LLM's Math Capability 2025
Role: co-first author | *Advisor: Tong Zhang and Yi R.(May) Fung* | *Publication:* arXiv:2505.23703

- This work proposed an end-to-end framework that augments natural-language mathematical reasoning with formal-language verification. Designed NL-FL Problem Alignment to reformulate NL QA tasks as FL existence theorems, enabling direct interaction with a formal prover. Developed a Mixed Problem Input mechanism allowing the formal agent to solve QA and existence problems concurrently.
- This framework achieved **89.80%** on MATH-500 and **84.34%** on AMC, outperforming the NL baseline; solved several problems that were unreachable by the NL baseline, even with more trials.
- Contributed to the design of the `HybridReasoning` pipeline and implemented the answer extraction mechanism.

Model-Collaboration Lean-based Long Chain-of-Thought Reasoning enhances Formal Theorem Proving 2025
Role: co-first author | *Advisor: Tong Zhang, UIUC* | *Publication:* arXiv:2503.03205

- Proposed MA-LoT, a comprehensive model-collaboration framework to balance NL reasoning and FL verification under the Long CoT paradigm for Lean4 theorem proving, allowing the model to generate in-depth formal reasoning through NL planning and analysis.
- Decoupled cognitive subtasks: natural language whole-proof generation (prover) and iterative formal error correction (corrector), unified through structured role collaboration.
- Introduced a novel LoT-Transfer Learning pipeline that enables Long CoT reasoning in domain-specific settings without requiring task-specific annotations.
- Achieved **61.07%** accuracy on the Lean4 MiniF2F-Test benchmark, outperforming DeepSeek-V3 (33.61%), InternLM-Step-Prover (50.70%), and Godel-Prover (55.33%).
- Engaged in the initial idea discussion on the model-collaboration logic between the prover and corrector, contributed to the prompting scheme and the structure of Long CoT content, and implemented the evaluation code for closed-source models.

ACADEMIC EXPERIENCE

Peer Mentor *Sep. 2024 - Present*
Affiliation: Department of Computer Science and Engineering, HKUST

- Mentored thirteen freshman and sophomore in Computer Science major through monthly sessions (10 hours/month) on course and research planning, study skills, and time management.
- Provided coursework guidance and Q&A support for courses such as Discrete Mathematics, Machine Learning, and Information Technology.

Research Assistant *Jun. 2024 - Present*
Advisor: Yongren Shi, The University of Arizona
Research Topic: Explore Selective Disclosure Bias with Networks of LLM-based Agents

- Developed a simulation to study how selective disclosure bias affects opinion segregation.
- Built simulation environments to model how LLM-based agents form and update opinions under selective information sharing.
- Designed experiments to examine the effects of disclosure patterns on opinion clustering and polarization in agent networks.

SERVICES

Conference Reviewer: KSS (2025)

SKILLS & CERTIFICATIONS

Standard Tests: TOEFL 5.5/6 (110/120)
Technical Skills: Python, C++, Java, Scala, RISC-V, LaTeX.