Name: Shuo Li Phone: (445) 888 2015 Email: shuoli0128@gmail.com

Shuo Li

Seattle, WA | shuoli0128@gmail.com | +1 445-888-2015

Google Scholar | LinkedIn | Personal Website | GitHub

Education

- Ph.D. in Computer Science (GPA: 3.95/4.0) — August 2025 | University of Pennsylvania
- Master in Robotics (GPA: 3.97/4.0) — May 2020 | University of Pennsylvania

Research Interests

Agentic AI · AI Safety (Alignment & Uncertainty Quantification) · LLM Evaluation.

Representative Publications

[1].A Fast, Reliable, and Secure Programming Language for LLM Agents. *Under Review.*
[2].SeekerGym: Benchmarking Agentic Information Seeking under Uncertainty. *Under Review.*
[3].BrowerArena: Evaluating LLM Agents on Real-World Web Navigation Tasks. *Under Review.*
[4].Alignment of Large Language Models with Constrained Learning. *NeurIPS 2025.*
[5].MR. Guard: Multilingual Reasoning Guardrail using Curriculum Learning. *EMNLP, 2025.*
[6].One-Shot Safety Alignment for Large Language Models. *NeurIPS 2024*, Spotlight. [Code]
[7].TRAQ: Trustworthy Retrieval Augmented Question Answering. *NAACL 2024*. [Code]
[8].Uncertainty in language models: Assessment through rank-calibration. *EMNLP 2024*. [Code]

Work Experience

Safe Coding Agent via formal Verification — **April – Current, 2026**
Research Scientist @Google Deepmind

- Shipped three agentic safety datasets, including adversarial prompt injection, real user trajectory.
- Improved formal method based policy guardian.

Safe Coding Agent via Verification — **June, 2025 - March, 2026**
Applied Scientist @Amazon AWS AI Lab

- Shipped a production-grade agent verification module, now integrated into a flagship AWS product used by enterprise customers.
- Designed a formal-methods verification pipeline, achieving 46.4% verification accuracy and reducing agent execution failures.
- Built Trajectory Analyzer, a reusable agent debugging library adopted by two internal agent teams to hill-climb evaluation leaderboards.
- Led collaboration with the Uncertainty Quantification team, guiding project direction and ensuring seamless progress.
- Pioneered post-training to generate formal specifications from natural language, expected to improve verification quality while reducing cost and latency.

Name: Shuo Li Phone: (445) 888 2015 Email: shuoli0128@gmail.com

Research Directions

Agentic Safety

- **Developed a programming language for LLM agents**, reducing execution time by 56%, improving security by 53%, and increasing reliability through conformal prediction [1].
- Built an agentic retrieval gym with provable reward signals; proposed SeekerAgent, outperforming frontier models retrieval recall by 68% [2].
- Co-designed **BrowserArena**, a real-world web navigation benchmark enabling action-level logging & trace comparison [3].
- Founded & co-host UPenn LLM Agent Reading Group Penn-Agents.

Trustworthy and Safe LLMs (Alignment & Guardrails)

- Innovated a single-step safety alignment method to navigate LLM helpfulness and safety tradeoffs. Reduced computational complexity by 90%, spotlight at NeurIPS 2024 [6].
- Advanced one-step method via an iterative optimization process. Improved safety constraint satisfaction rate by 57%. Accepted to NeurIPS 2025 [4].
- Built reasoning-aware multilingual guardrail via GRPO. Obtained more than 15% higher safety against jailbreaking than SOTA approaches. Accepted to EMNLP 2025 [5].

Uncertainty Quantification for LLM

- Established the first provable guarantee for RAG via conformal prediction. Accepted at NAACL 2024; Best paper at 2023 TEACH ICML [7].
- Pioneered a rank-based metric for LLM uncertainty quantification, improved robustness and consistency over other metrics. Accepted at EMNLP 2024 [8].

Technical skills

- Agentic Safety & Post-Training: RLHF, GRPO, Uncertainty Quantification, Static Analysis
- Programming Languages: Python, MATLAB, C/C++.
- Libraries & Frameworks: PyTorch, Transformers, Scikit-Learn, LangChain, TRL, VeRL.
- Tools: AWS, Git, Docker, Parallel Computing.

Impact Highlights

- Invented agent safety mechanisms (agentic verification + formal verification).
- Shipped agent verification into a flagship AWS product used by enterprise customers.
- Built reusable trajectory analyzer adopted by two internal agent teams.