JIAYI WU

Phone: (+1) 401-743-3391 ⋄ Email: jiayi_wu4@brown.edu
Google Scholar ⋄ GitHub ⋄ LinkedIn ⋄ Personal Webpage

EDUCATION

Brown University Sept 2024 - May 2028 (expected)
A.B. Mathematics and Sc.B. Computer Science
Advisor: Isabel Vogt and Suresh Venkatasubramanian
Related Coursework:

- Machine Learning / Artificial Intelligence: CSCI 2952N Advanced Topics in Deep Learning, CSCI 2470 Deep Learning (Graduate Level), APMA 1941J Topics in High-Dimensional Probability, CSCI 2952W Critical Data and Machine Learning, APMA 1655 Introduction to Probability and Statistics with Theory, MATH 0540 Linear Algebra with Theory
- Theory / Formal Methods: CSCI 1010 Theory of Computation, CSCI 1715 Formal Proof and Verification, CSCI 1951Y Using an Interactive Proof Assistant to Do Mathematics, CSCI 1973 Individual Independent Study (on Automated Theorem Proving), MATH 1530 Abstract Algebra, MATH 1540 Topics in Abstract Algebra

RESEARCH INTERESTS

I hope to build the *formal* and *empirical* infrastructure for **trustworthy machine learning** across two complementary dimensions:

- **AI for mathematics and automated reasoning**: formal theorem proving and program synthesis are fully specified, unambiguous, machine-checkable substrates for studying models' reasoning processes. Concretely, I'm interested in designing model architectures that leverage various learning paradigms (reinforcement learning and neurosymbolic programming in particular) to enable more robust and interpretable reasoning.
- **Principled evaluation, auditing, and guarantees**: Evaluating trustworthiness requires pipelines that adapt to different systems at different stages of model design, development, and deployment. I'm interested in enabling these pipelines to reason about security, alignment, and accountability across these stages and throughout the full model lifecycle and supply chain.

PUBLICATIONS

* represents equal contribution.

[1] Elinor Poole-Dayan, **Jiayi Wu**, Taylor Sorensen, Jiaxin Pei, Michiel A. Bakker, "Benchmarking overton pluralism in LLMs," in *The Fourteenth International Conference on Learning Representations (ICLR 2026)*, also accepted to *NeurIPS 2025 Workshop on Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling*, 2025. [Online]. Available: https://arxiv.org/abs/2512.01351.

[2] Chance Jiajie Li*, **Jiayi Wu***, Zhenze Mo, Ao Qu, Yuhan Tang, Kaiya Ivy Zhao, Yulu Gan, Jie Fan, Jiangbo Yu, Jinhua Zhao, Paul Pu Liang, Luis Alberto Alonso Pastor, Kent Larson, "Simulating society requires simulating thought," in *The Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025) Position Paper Track*, 2025. [Online]. Available: https://arxiv.org/abs/2506.06958.

[3] Chance Jiajie Li*, Zhenze Mo*, Yuhan Tang, Ao Qu, **Jiayi Wu**, Kaiya Ivy Zhao, Yulu Gan, Jie Fan, Jiangbo Yu, Hang Jiang, Paul Pu Liang, Jinhua Zhao, Luis Alberto Alonso Pastor, Kent Larson, "HugAgent: Evaluating LLMs in simulating individual-level human reasoning on open-ended tasks," in *NeurIPS 2025 Workshop on Bridging Language, Agent, and World Models for Reasoning and Planning (Spotlight)*, also accepted to *NeurIPS 2025 Workshop on PersonaLLM: Workshop on LLM Persona Modeling (Oral)* and *NeurIPS 2025 Workshop on Socially Responsible and Trustworthy Foundation Models (ResponsibleFM)*, 2025. [Online]. Available: https://arxiv.org/abs/2510.15144.

RESEARCH EXPERIENCE

Interoperability for Formal Mathematics: Automating Proof Translation Across ITPs Feb 2026 - Present
California Institute of Technology Pasadena, CA

Summer Undergraduate Research Fellowships (SURF) project advised by Anima Anandkumar, in collaboration with Robert Joseph George.

Neurosymbolic Inference of Regular Stream Types Jan 2026 - Present
Brown University Providence, RI

Undergraduate Teaching and Research Awards (UTRA) project advised by Nikos Vasilakis, in collaboration with Lukas Lazarek, Zekai Li, and Evangelos Lamprou.

Pattern Mining and Automated Tactic Discovery in Theorem Proving Oct 2025 - Present
Brown University Providence, RI

Research project (CSCI 1973 Individual Independent Study) advised by Robert Lewis and Stephen Bach, in collaboration with Gavin Zhao, Zekai Li, and Ilija Ivanov.

Formalizing and Benchmarking Overton Pluralism in Large Language Models [1] May 2025 - Dec 2025
Massachusetts Institute of Technology Cambridge, MA

Visiting summer researcher project advised by Michiel Bakker, in collaboration with Elinor Poole-Dayan.

Motif of Thoughts (MoT): Reusable Abstractions for Neurosymbolic Reasoning [2] [3] Mar 2025 - Present
Massachusetts Institute of Technology Cambridge, MA

Research project in collaboration with Chance Jiajie Li.

Algorithmic Fairness, AI Governance, and AI Safety Primer, SRC Handbook Dec 2024 - Present
Brown University Providence, RI

Research project advised by Suresh Venkatasubramanian, Julia Netter, and Serena Booth at Center for Technological Responsibility, Reimagination and Redesign (CNTR).

PROJECT EXPERIENCE

The Cutoff Phenomenon for Markov Chains: Entropy, Curvature, and Mixing Times Feb 2026 - Present
Brown University Providence, RI

Class final project at APMA 1941J Topics in High-Dimensional Probability.

Probing Structural Signals in Lean 4 Proof Graphs with GNNs Sept 2025 - Dec 2025
Brown University Providence, RI

Class final project at CSCI 2470 Deep Learning (Graduate Level).

Formalizing Hartshorne's Foundations of Projective Geometry in Isabelle Sept 2025 - Dec 2025
Brown University Providence, RI

Class final project at CSCI 1951Y Using an Interactive Proof Assistant to Do Mathematics.

TECHNICAL SKILLS

Programming Languages	Java, Python, JavaScript, HTML/CSS, SQL, MATLAB, Lean, Isabelle, ReasonML, Racket, C/C++, Processing
Programming Tools	PyTorch, TensorFlow, scikit-learn, NumPy, pandas, Git/GitHub, Linux