# JIAYI WU

Phone: (+1) 401-743-3391 ⋄ Email: jiayi_wu4@brown.edu
Homepage: https://jiayiw005.github.io/
Google Scholar ⋄ Github ⋄ LinkedIn

## EDUCATION

**Brown University**                                     Sept 2024 - May 2028 (expected)
Sc.B. Mathematics-Computer Science and A.B. Applied Mathematics
Related courses: CSCI 2470 Deep Learning, CSCI 2952W Critical Data and Machine Learning Studies, CSCI 1010 Theory of Computation, CSCI 1715 Formal Proof and Verifications, CSCI 1951Y Using an Interactive Proof Assistant to Do Mathematics, APMA 1655 Introduction to Probability and Statistics with Theory, MATH 1530 Abstract Algebra, POLS 1460 International Political Economy

**Shanghai Qibao Dwight High School**                          Sept 2021 - June 2024
International Baccalaureate Diploma Programme (IBDP)
Related courses: Mathematics Analysis and Approaches HL, Computer Science HL, History HL, Theory of Knowledge, Global Politics-History Extended Essay

## RESEARCH INTERESTS

I am interested in the *theoretical underpinnings* of **alignment**, **reliability**, and **interpretability** in machine learning, with a particular focus on connections to complexity theory, game theory, and formal proof systems, situated within broader questions of **technical AI governance** and **complex systems**, where I aim to bridge rigorous mathematical reasoning with the design and evaluation of accountable, trustworthy AI systems.

## RESEARCH EXPERIENCE

**Benchmarking Overton Pluralistic Alignment for LLM Evals [1]**       May 2025 - Present
*Supervisors: Prof. Michiel Bakker, Elinor Poole-Dayan*       Massachusetts Institute of Technology

*Affiliation: Visiting Summer Researcher at Sloan School of Management and Center for Constructive Communication, Media Lab*

**AI Governance Glossary: Policy, Debate, and Etymology**          Apirl 2025 - Present
*Supervisor: Prof. Sarah Cen*                                        Stanford University

**Agent City Hall: Capturing and Replicating Human Reasoning [2] [3]**   Mar 2025 - Present
*Supervisor:Chance Jiajie Li*                         Massachusetts Institute of Technology

*Affiliation: Undergraduate Research Collaborator at City Science Group, Media Lab*

**AI Fairness Research, Socially Responsible Computing Handbook**      Dec 2024 - April 2025
*Supervisor: Prof. Suresh Venkatasubramanian and Prof. Julia Netter*          Brown University

*Affiliation: Undergraduate Research Assistant at Brown University Center for Technological Responsibility (CNTR), Department of Computer Science, and Data Science Institute (DSI)*

**Critical Auditing of Food-Delivery Algorithmic Management**        Nov 2024 - Present
*Supervisor: Prof. Harini Suresh*                                    Brown University

*Affiliation: Undergraduate Researcher at the Data in Society Collective (DISCO Lab), Brown University Department of Computer Science*

## PROJECT EXPERIENCE

**Multimodal Vision–Language Modeling for Autonomous Driving**     Aug 2025 - Present
*Affiliation: Latitude AI*     Break Through Tech AI/ML Fellowship

**ConnectRI Broadband Navigator Accessibility Audit**     Feb 2025 - May 2025
*Affiliation: Digital Equity Initiative*     Brown Initiative for Policy

**AsyLex Refugee Claim Entity Extraction & Judgement Prediction**     Sept 2024 - Mar 2025
*Affiliation: Ignite Fellowship*     AI4ALL

**Open-Sourced Android Mobile App for Education Resource Sharing** Nov 2021 - June 2023
*Affiliation: Founder, Technology Board*     Converter Education Initiative

**Modeling Honeybee Population Dynamics**     Feb 2022 - Feb 2023
*Affiliation: Team 12736*     High School Mathematical Contest in Modeling (HiMCM)

## PUBLICATIONS

[1] E. Poole-Dayan, **Jiayi Wu**, J. Pei, and M. A. Bakker, "Benchmarking Overton Pluralism in LLMs," submitted to the Thirty-Ninth Conference on Annual Neural Information Processing Systems (NeurIPS) Workshop on Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling.

[2] C. J. Li, **Jiayi Wu**, Z. Mo, A. Qu, Y. Tang, K. I. Zhao, Y. Gan, J. Fan, J. Yu, J. Zhao, P. Liang, L. Alonso, and K. Larson, "Position: Simulating Society Requires Simulating Thought," submitted to the Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS), Position Paper Track.

[3] C. J. Li, Z. Mo, Y. Tang, A. Qu, **Jiayi Wu**, K. I. Zhao, Y. Gan, J. Fan, J. Yu, J. Zhao, P. P. Liang, L. A. A. Pastor, and K. Larson, "HugAgent: Human-Grounded Benchmarking of Agent Beliefs in Social Contexts," submitted to the Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS) Workshop on LAW 2025: Bridging Language, Agent, and World Models for Reasoning and Planning.

## AWARDS

| | |
|---|---|
| Undergraduate Teaching and Research Awards (UTRA), Brown University | Fall 2025 |
| Chairman Scholarship (5%), Shanghai Qibao Dwight High School | Spring 2024 |
| Selected for Pioneer Journal Publication (1.8%), Pioneer Academics | Fall 2023 |
| Academic Excellence Scholarship First Place (1%), Shanghai Qibao Dwight High School | Spring 2023 |
| Meritourious Award (15%), COMAP's High School Mathematical Contest in Modeling | Spring 2023 |
| Global First Place Award (5%), China Thinks Big Competition(CTB) | Fall 2022 |

## SKILLS/MISCELLANEOUS

| | |
|---|---|
| **Programming Languages** | Java, Python, JavaScript, HTML/CSS, SQL/MySQL, MATLAB, Lean, Isabelle/HOL, ReasonML, Racket, C/C++, Processing |
| **Programming Tools** | PyTorch, Tensorflow, Sklearn, Pandas, Numpy, Vite, Vue.js, D3.js, Plotly, Android Studio, Microsoft Excel, Git/GitHub, Linux |