

Weiming Sheng

✉ ws2720@columbia.edu ☎ 845-318-9002 in LinkedIn @ GitHub

Education

Columbia University — <i>PhD in Philosophy</i> (GPA 4.0/4.0)	New York, NY	08/2023 – 05/2029
Vassar College — <i>BA in Philosophy with Honors</i> (GPA 4.0/4.0)	Poughkeepsie, NY	09/2019 – 05/2023
Oxford University, St. Anne's College — <i>Visiting Student</i>	Oxford, UK	10/2022 – 06/2023

Publications

“Fictionalism About Personae: Folk Psychology as an Interpretability Strategy”

Philosophy Meets Machine Learning Workshop, ICML 2026 2026

I argue that LLMs generate fictional characters rather than instantiate mental states, and that the predictive success of folk psychology for LLM systems is explained by the model's capacity to generate folk-psychologically coherent characters.

“Building a Correct-by-Design Lakehouse: Data Contracts, Versioning, and Transactional Pipelines for Humans and Agents”

Submitted 2025

We show how typed table contracts, Git-like data versioning, and transactional pipeline execution can make illegal states unrepresentable and avoid partial effects.

Professional Experience

Cambridge AI Safety Hub — *Research Fellow, MARS V* Remote / Cambridge, UK 06/2026 – Present

- Selected for a competitive part-time AI safety research program; working in Elizaveta Tennant's (Google DeepMind / UCL) moral alignment stream on a team project on agentic moral alignment.

ValuesLab, Columbia University — *Research Mentor* New York, NY 2026 – Present

- One of three mentors for the 2026 Student Research Fellowship at Katja Vogt's lab for AI alignment; contributed to the cohort's project selection and mentoring a fellow on data augmentation for vision-language models (VLMs).

Bauplan — *Research Intern* New York, NY 05/2025 – 08/2025

- Developed rigorous formal semantics for a version-controlled data lake system (“Git for data”), establishing the logical foundations for branched data lineage and atomic commits.
- Designed and implemented a first-order logic specification in **Alloy**, using automated model finding to verify the consistency and correctness of the proposed versioning primitives.

Columbia University — *Graduate Teaching Assistant* New York, NY 09/2024 – Present

- Lead two weekly recitation sessions in Symbolic Logic for 100+ students, focused on analytical problem-solving.
- Received a teaching effectiveness score of 4.5/5 in student evaluations.

Floyd Philosophy Education — *Founder / Program Developer* Hangzhou, China 05/2018 – 08/2020

- Organized 3 in-person camps introducing philosophy to 100+ Chinese students.
- Led the marketing team, generating 300+ total applicants.
- Won National 3rd Prize at the UN Development Programme (UNDP) China Nonprofit Competition, 2019.

Graduate Coursework

PhD-level: Analysis & Probability I • Theoretical Statistics I • Advanced Reinforcement Learning

Masters-level: Machine Learning • Theory of Machine Learning and Interactive Decision Making

Skills & Achievements

Languages: English (Native), Mandarin (Native)

Technical Skills: Python, R, Scheme, LaTeX, Office Suite

Honors: Phi Beta Kappa (2023) • Philip Nochlin Prize for Best Senior Thesis (2023)