Stanford



Tianyi Zhou

Graduate Visiting Researcher Student, Computer Science

Bio

BIO

My research aims to build safe, capable, and responsible AI systems. My earlier work studied how algorithmic design in online information systems, including social networks, news platforms, and recommender systems, can amplify harm and how these harms can be mitigated. I now focus on the evaluation, post-training, and alignment of frontier models. I study how to identify and mitigate the risks that emerge as language models become more capable and agentic. In particular, I investigate how uncertainty-aware generation and verification can support safe and controllable model behavior while preserving reasoning and problem-solving capabilities. My work has been published in NeurIPS, ICLR, AAAI, WWW and CIKM.

PROFESSIONAL AFFILIATIONS AND ACTIVITIES

- Visiting student researcher, Stanford University Center for AI Safety (2026 - present)
- Affiliated Researcher, Wallenberg AI, Autonomous Systems and Software Program(WASP) (2022 - present)
- PhD student, KTH Royal Institute of Technology (2021 - present)
- Research Intern, Qatar Computing Research Institute (2025 - 2025)