

Yifei Zhou

yifeizhou02.github.io | yifei_zhou@berkeley.edu | 6073791318 | Berkeley, CA

EDUCATION

University of California, Berkeley

Ph.D. in Electrical Engineering and Computer Science

Aug.2023 -

Berkeley, CA

- Advisor: Sergey Levine
- Research Interest: Reinforcement Learning with LLMs

Cornell University

Bachelor of Arts in Computer Science and Mathematics

Aug.2020 - May.2023

Ithaca, NY

- GPA: 4.12/4.0
- Pauline and Irvine Tanner Dean's Scholar Scholarship (40/2000+)
- 2023 CRA Outstanding Undergraduate Researcher Finalist

Selected Papers

Yifei Zhou*, Hao Bai*, Mert Cemri, Jiayi Pan, Alane Suhr, Sergey Levine, Aviral Kumar, DigiRL: Training In-The-Wild Device-Control Agents with Autonomous Reinforcement Learning. NeurIPS 2024. Oral presentation at FM-WILD ICML 2024 (top 3%).

Yifei Zhou, Andrea Zanette, Jiayi Pan, Sergey Levine, Aviral Kumar, ArCHer: Training Language Model Agents via Hierarchical Multi-Turn RL. ICML 2024.

Yifei Zhou*, Qianlan Yang, Kaixiang Lin, Min Bai, Xiong Zhou, Yu-Xiong Wang, Sergey Levine, Erran Li Proposer-Agent-Evaluator (PAE): Autonomous Skill Discovery For Foundation Model Internet Agents, Pre-print.

Yifei Zhou*, Yuda Song*, Ayush Sekhari, Drew Bagnell, Akshay Krishnamurthy, Wen Sun, Hybrid RL: Using both offline and online data can make RL efficient, ICLR 2023.

Jiayi Pan, Yichi Zhang, Nicolas Tomlin, **Yifei Zhou**, Sergey Levine, Alane Suhr, Autonomous Evaluation and Refinement of Digital Agents. COLM 2024. Best paper award at CVPR MAR 2024.

Yifei Zhou*, Ayush Sekhari*, Yuda Song, Wen Sun, Offline Data Enhanced On-Policy Policy Gradient with Provable Guarantees. ICLR 2024.

Yifei Zhou*, Juntao Ren*, Fengyu Li*, Ramin Zabih, Ser-Nam Lim, Test-Time Distribution Normalization for Contrastively Learned Vision-language Models. NeurIPS 2023.

Yifei Zhou, Zilu Li, Abhinav Shrivastava, Hengshuang Zhao, Antonio Torralba, Taipeng Tian, Ser-Nam Lim, BT²: Backward Compatible Training with Basis Transformation. ICCV 2023.

Employment EXPERIENCE

Fundamental AI Research (FAIR), Meta

Visiting Researcher

Sep.2024 -

Menlo Park, CA

- Work under the Alignment pillar of FAIR.
- Carry out independent research projects on improving decision-making capabilities of Large Language Models.

Amazon

Applied Scientist Intern

May.2024 - Sep.2024

Santa Clara, CA

- Collaborate with AWS AI and AGI team.
- Work towards autonomous digital agent through Reinforcement Learning.

Professional Services

Reviewer for NeurIPS (2023-2024), ICML (2024), ICLR (2023-2025), CVPR (2023).