Yanming Wan

ymwan@cs.washington.edu | https://www.wanyanming.com

Profile

- **PhD student interested in Social Reinforcement Learning and Large Language Models.**
- Currently studying in **University of Wahington CSE**, advised by **Prof. Natasha Jaques**.

Education

University of Washington **Seattle, WA, USA**

PhD Student, Paul G. Allen School of Computer Science & Engineering, Advised by Natasha Jaques *Sept. 2023 – present*

- GPA: 3.95/4.00 (Up till Autumn Semester, 2025)
- Received **Amazon AI PhD Fellowship** *Oct. 2025*
- Received **Madrona Prize** at Allen School's annual Industry Affiliates Research Showcase *Oct. 2025*

Tsinghua University **Beijing, China**

Undergraduate Student, Yao Class, Institute for Interdisciplinary Information Sciences (IIIS) *Aug. 2019 – June 2023*

- **GPA: 3.97/4.00, Rank: 2/30**
- Received **Recognition Prize of Yao Award** *Sept. 2022*
- Received **Jiang-Nanxiang Scholarship** (Unique in Yao Class) *Dec. 2021*
- Won the **Gold Medal** in 2018 Chinese Mathematical Olympiad, and admission guaranteed *Nov. 2018*

Research Experiences

Seed Edge, ByteDance **Bellevue, WA, USA**

Student Researcher, Mentored by Yizhong Wang *Sept. 2025 – present*

- Exploring methods to enable reasoning models to handle out-of-distribution environments by actively interacting with the environments to acquire information and infer underlying structures, with a focus on task-free exploration.

Google DeepMind **Seattle, WA, USA**

Student Researcher, Hosted by Jiaxing Wu *Sept. 2024 – March 2025*

- Current LLM training methods like RLHF prioritize helpfulness and safety but fall short in personalization. Traditional methods to personalization often rely on extensive user history, limiting their effectiveness for context-limited users.
- We propose to incorporate an intrinsic motivation to improve the conversational agents' model of the user as an additional reward alongside multi-turn RLHF. This auxiliary reward encourages the agent to actively elicit user traits by optimizing conversations to increase the accuracy of its user model, and consequently the policy agent can deliver more personalized interactions through obtaining more information about the user.

Social RL Group, University of Washington **Seattle, WA, USA**

Research Assistant and Teacher Assistant, Advised by Natasha Jaques *Sept. 2023 – present*

- Variantional Preference Learning: To address the need for pluralistic alignment, we propose to infer a novel user-specific latent and learning reward models and policies conditioned on it without additional user-specific data.
- Follow Instructions with Social and Embodied Reasoning: To better natural language instruction in collaborative embodied tasks, we propose to make explicit inferences of human's goals and intentions as intermediate reasoning steps.

CoCoSci Lab, Massachusetts Institute of Technology **Cambridge, MA, USA**

Undergraduate Visiting Student, Advised by Josh B. Tenenbaum and Jiayuan Mao *Feb. 2022 – July 2022*

- To collaborate with human partners successfully in complex environments, robots should be able to interpret and follow natural language instructions in contexts.
- Introduced HandMeThat, a benchmark for a holistic evaluation of instruction understanding and following in physical and social environments, which highlights the additional challenge of understanding instructions with ambiguities based on physical states and human actions and goals.

Publications

Yanming Wan*, Jiaxing Wu*, Marwa Abdulhai, Lior Shani, Natasha Jaques. Enhancing Personalized Multi-Turn Dialogue with Curiosity Reward. In *NeurIPS*, 2025.

Hyunji Nam, **Yanming Wan**, Mickel Liu, Jianxun Lian, Natasha Jaques. Learning Pluralistic User Preferences through Reinforcement Learning Fine-tuned Summaries. In *submission*, 2025.

Yanming Wan, Yue Wu, Yiping Wang, Jiayuan Mao†, Natasha Jaques†. Infer Human's Intentions Before Following Natural Language Instructions. In *AAAI*, 2025.

Sriyash Poddar*, **Yanming Wan***, Hamish Ivison, Abhishek Gupta†, Natasha Jaques†. Personalizing Reinforcement Learning from Human Feedback with Variational Preference Learning. In *NeurIPS* (spotlight), 2024.

Guang Yang, Muru Zhang, Lin Qiu, **Yanming Wan**, Noah A. Smith. Toward a More Complete OMR Solution. In *ISMIR*, 2024.

Yanming Wan*, Jiayuan Mao*, Joshua B. Tenenbaum. HandMeThat: Human-Robot Communication in Physical and Social Environments. In *NeurIPS Datasets and Benchmarks Track*, 2022.