WENJIE LI

I'M YOUNG. I HAVE A BIG DREAM. I'M WORKING HARD FOR IT.

liwj2022@shanghaitech.edu.cn Shanghai, China

Profile

I am a first-year graduate student majoring in AI at ShanghaiTech University. My research interests include Data Science and ML Safety. My advisor is Pengcheng Zeng, and we are now developing statistical methods and computational tools for analyzing cutting-edge large scale genomic data and healthcare data. Besides my coursework, I am also admitted by Stanford Existential Risk Initiative (SERI, summer 2022 cohort) and Berkeley AI Safety Initiative for Students (BASIS), working as a student fellow. We aimed at contributing to mitigating X-risks from Advanced AI. From last winter, I started working under the supervision of Philip Thomas from UMASS on the ML Fairness problem and I have already written a blog based on their Science paper, introducing the Seldonian framework. See here to get more info about this project if interested. Currently I'm working on an research project supervised by Marius Hobbhahn from Max Planck Institute.

Education & Experience

Artificial Intelligence, MSc, ShanghaiTech, Shanghai

September 2022 — June 2025

New cohort of 2022 Fall

Computer Science, Bachelor, NCEPU, Beijing

September 2018 — June 2022

The first year in college, I ranked the top 4 in my class and successfully passed the examination of major transfer from Automation to Computer Science.

In the sophomore year, I planned to study abroad for a master's degree. Then I took my first IELTS and got a total score of 6.5 by self-studying. (reading 7.5, listening 7, speaking 6.0, writing 6.0).

The junior year, I successfully got the exchange qualification, but voluntarily gave it up due to the COVID, and then prepared for the postgraduate entrance examination of China.

In the senior year, I was awarded the honorable member both in the summer camp held by Information school and BME school of ShanghaiTech. And finally I decide to join the School of Information with an interview exemption.

Courses completed: Linear Algebra (98) Java Programming (92) Digital Logic (90) Virtual Reality (100) Circuit Theory (93) Web Development (93) Introduction to AI (91)

Links

Blog

Github

Linkedin

Honors

University-level Scholarship

Merit student of the department

Excellence Award, Beijing, Innovation and Entrepreneurship Program for Chinese Students

Third prize, National English Competition for University Students

Over 10k Scholarship from Stanford, Berkeley and Center for AI Safety

More in the future...

Skills

English

Programming

Guitar

Skateboarding

liwj2022@shanghaitech.edu.cn Shanghai, China

Miscellaneous

Student Fellow

June 2022 — Present

Center for International Security and Cooperation,Freeman Spogli Institute, Stanford

I was admitted by Stanford to join the first cohort of SERI fellows in China. I work with Prof.Philip on some Fairness example problems. Currently I've finished a write-up to introduce the "Seldonian Framework" we used to constrain model's undersired behaviour like racist and discrimination. The further goal is to submit a paper focusing on Fairness and Safety problem in Reinforcement Learning context.

SPAR Mentee

March 2023 — Present

Berkeley AI Safety Initiative for Students

From this Spring, I started working in BASIS on some interpretability problem under the supervision of Marius. Currently we are still exploring how to detect trojans/backdoors in a toy-like transformer model. We hope this work might give some hints when someone want to scale up the task in the future,

AGISF, Participant

July 2022 — September 2023

BlueDot Impact

I finished the 12-week, part-time courses to help people learn about AGI safety and make professional connections to the research field.

MLSS Scholar

February 2023 — May 2023

Center for AI Safety

I finished the 8-week course where we discuss how researchers can shape the process that will lead to strong AI systems and steer that process in a safer direction. Basically we cover various technical topics to reduce X-Risks from strong AI, like “Robustness”, “Monitoring”, “Alignment” and “Systemic Safety”.

Horizon Fellow, EA Hong Kong

June 2023 — August 2023

This summer, I will join the Horizon Fellowship Program. I hope to get more insights and have more thoughts on how to do good effectively, and how I can contribute to some of the world’s most pressing problems.