Alexa Pan

Washington, District of Columbia | (203) 285-9817 | alexapan1783@gmail.com | linkedin.com/in/alexa-pan-5678ab163

EDUCATION

Yale University, New Haven CT, United States   Sep 2020 – May 2024
BA Physics and Philosophy, Sigma Pi Sigma   GPA: 3.82
Relevant Coursework: Intermediate Machine Learning, Introductory Machine Learning, Probability Theory, Linear Algebra, Information Theory, Philosophy of Mind, Liability and Regulation for Frontier AI development

Oxford University, Oxford, United Kingdom   Jan 2023 – Jun 2023
BA Mathematics and Philosophy, Visiting Student
Relevant Coursework: Logic, Graph Theory, Practical Ethics, Philosophy of Cognitive Science

Raffles Institution, Singapore   Jan 2016 – Dec 2019
Awards: Silver, Singapore Physics Olympiad (2019). Gold, Singapore Mathematical Olympiad (2017). Best Written Response, Best Speaker, Hwa Chong Invitational Philosophy Olympiad (2018). SM1 Scholarship.

WORK EXPERIENCE

Redwood Research, *Member of Technical Staff (Work Trial),* Berkeley, United States   Sep 2025 – present

Forecasting Research Institute, *Research Analyst,* Washington DC, United States   Nov 2024 – Sep 2025
design and run large-scale experiments to elicit judgment-based forecasts on key indicators of AI progress and impact.

Center for AI Safety, *Technical Writer*, San Francisco CA, United States   Jun 2024 – Dec 2024
I contribute to the AI Safety Newsletter, which covers AI and AI safety developments for tens of thousands of subscribers.

ChinaTalk, *AI Analyst,* New York NY, United States   Jun 2024 – Oct 2024
Analyzed China's frontier AI research and industry developments to shed light on "how far are Chinese AI capabilities from the frontier", considering different domains and applications of AI and global AI governance implications.

Center on Long-term Risk, *Summer Research Fellow*, London, United Kingdom   Jun 2023 – Aug 2023
Conducted conceptual research on LLM capabilities and propensities and empirical research on ChatGPT's game theoretic reasoning and other cooperative capabilities. Supervised by Daniel Kokotajlo (OpenAI)

Concordia AI, *AI Safety Affiliate*, Beijing, China   Jun 2023 – Jan 2024
Translated, summarized, and edited academic papers, conference talks, and web articles related to technical AI safety for Chinese language audiences.

Owain Evans, *Research Assistant*, London, United Kingdom   Apr 2023 – Jun 2023
Created adversarial prompts and automated red-teaming experiments to evaluate the frequency and robustness of lie detection in LLMs. Co-authored "How to Catch an AI Liar". Supervised by Owain Evans.

QuLab @Yale Quantum Institute, *Research Intern,* New Haven CT, United States   May 2022 – Aug 2022
Assisted with experimental set-up for upcoming quantum error correction experiments, supervised by Michel Devoret.

PROJECTS AND ACTIVITIES

Yale Effective Altruism, *President*, Yale University   Jan 2022 – Dec 2022
Lead the executive team in strategy, planning and execution in community building on campus.

Machine Learning Safety Scholarship, *Participant*, remote   June 2022 – Aug 2022
Audited ten-week course in machine learning and research directions in ML safety, sponsored by the Center for AI Safety.

AI@RI Conference, *Founder*, Raffles Institution   2019
Organized the inaugural student conference on the ethics and philosophy of artificial intelligence.

SKILLS AND INTERESTS

Programming Skills: proficiency in Python; familiarity with C, SQL, HTML.
Language Skills: English & Mandarin (native fluency), French (elementary proficiency).