MARK XU

m@rkxu.me ◇ markxu.com

EXPERIENCE

Research Intern Aug 2021 - Present
Alignment Research Center *San Francisco, CA*

- Designing algorithms that train agents that "tell you everything they know" as an alignment strategy.

Research Intern Jan 2021 - March 2021
Machine Intelligence Research Institute *Berkeley, CA (Remote)*

- Studying act-based agents, goals across Cartesian boundaries, and myopia to understand how to do relaxed adversarial training.

Board of Directors March 2020 - Present
Helpful; helpfulengineering.org *Sylvania, OH*

- Oversee Helpful, a non-profit entity that incubates open-source solutions for COVID-19 and other global challenges.

Contract Researcher Sept 2020 - Dec 2020
OpenAI *San Francisco, CA (Remote)*

- Identified limitations to AI safety via debate from both practical and theoretical angles.

Machine Learning Intern June 2020 - Aug 2020
Pure Storage *Mountain View, CA (Remote)*

- Prototyped latency anomaly detection methods, laying groundwork for development of a new fleet-management feature.
- Developed novel methods for time-series data analysis and decomposition.

Software Engineering Intern June 2019 - Sept 2019
Pure Storage *Mountain View, CA*

- Uncovered performance issues in VASA integration by developing VM boot storm simulation microbenchmark.
- Improved engineer productivity by halving expected runtime of multiple benchmarks and deploying test validation infrastructure.
- Stabilized internal log processing tools by resolving long-standing deployment issues.

Research Fellow June 2018 - Sept 2018
California Institute of Technology *Pasadena, CA*

- Determined plausible alternate paths for interhemisphere information transfer using causal inference techniques.

EDUCATION

California Institute of Technology Expected 2023

Bachelors of Mathematics GPA: 3.9

SKILLS

Technical Skills Python (numpy, scipy, scikit-learn, pandas, pytorch, etc.), C++, C, Java

Coursework Algebra, Algorithms, Analysis, Computability Theory, Information Theory, Machine Learning, Model Theory, Set Theory, Topology