Joey Yudelson

201-679-4465
Jyudel@gmail.com

EXPERIENCE

Redwood Research — *Research Resident + Intern*

December 2022 - April 2023

- Researched novel methods for using mechanistic interpretability to detect anomalous inputs and behavior in small, transformer-based ML models.
- Designed and engineered experiments to test a method of mechanistic anomaly detection, with the ultimate goal being the development of new AI safety evaluations.
- Started as a contract worker; was subsequently offered an internship to complete the research started.

Glass Lab, Rutgers University— *Researcher*

September 2021 - July 2022

- Developed and maintained novel data analytics tools for modeling subjects' behavior in economic games.
- Wrote usable, thorough, and well-commented code.

Juno Journey, Israel — *Data Scientist + Software Engineer*

January 2019 - June 2020

- Used cutting-edge machine learning models (BERT, GPT2) to analyze large quantities of data.
- Developed algorithms for a content recommendation system.
- Coordinated back-end and front-end developers.
- Started as an intern; was subsequently offered a part-time position.

EDUCATION

Rutgers University — *Bachelors, Summa Cum Laude, 3.923 GPA*

September 2019 - August 2022

Majored in Computer Science, Minored in Mathematics

Relevant Coursework: Machine Learning, Numerical and Complex Analysis, Statistics and Probability, Operating Systems Design

PROJECTS

RoboTraffic — *2nd Place*

March 2017

- Developed steering software for an autonomous car as part of the international RoboTraffic Competition at the Technion.
- Worked with sensor suites and microprocessors to optimize a miniature car for safe, legal driving.

SKILLS

- Data analysis & machine learning proficiency (PyTorch, DVC, etc.)
- Data analysis with Python (Pandas, Numpy, etc.)
- Fast, scalable software development in Rust
- Writing
- Fast Learner

AWARDS

Gold Medal at the Greater New York Math Fair, for work in Weierstrass Functions

2nd Place at TechNegev Hackathon for the West Negev

LANGUAGES

- Python: Proficient
- Rust: Intermediate
- English: Native Speaker
- Hebrew: Working Proficiency