Richard So

347-281-3815 | richardso2021@gmail.com | github.com/richardso21 | in/richardso21 | sorichard.com

Education

Georgia Institute of Technology 08/2021 – 05/2025

B.S./M.S. Computer Science, Interactive Intelligence — GPA: 4.0 *Atlanta, GA*

- Coursework: Data Structures, Discrete Math, Algorithms Honors, Machine Learning, Computer Vision

Work Experience

Amazon Web Services | *Software Engineering Intern (ML)* 05/2024 — 08/2024

- Part of an internal investigative team aiming to quantitatively measure and analyze user/developer experience across AWS.
- Overhauled a pipeline to ingest unstructured user sessions into a queryable data lake leveraging AWS Lambda, S3, and Glue.
- Optimized a screenshot labeling process used to categorize user activity through AWS Rekognition and Augmented AI.

GT Financial Services Innovation Lab | *Research Assistant* 05/2024 — Present

- Explored benchmarking strategies, datasets, and metrics to evaluate against state-of-the-art financial LLMs.
- Compiled and curated large textual datasets for LLM fine-tuning with Scrapy, BeautifulSoup4 and NLTK.

Tanium | *Software Engineering Intern* 06/2023 — 08/2023

- Integrated a CRUD logger into an internal PostgreSQL database and RESTful API interface to elevate console visibility.
- Rapidly tackled 50+ Jira tickets within a 10-week internship period maintaining a Knex.js and React TypeScript codebase.
- Exercised test-driven development and data validation using Jest, Jasmine, and Joi to ensure UI and API reliability.

GT College of Computing | *Senior Teaching Assistant* 01/2023 — 05/2024

- Led biweekly lectures on the foundations of computer architecture, the C language, and memory allocation principles.
- Developed unit testing suites, docker images for auto-grading, and course software, servicing 1000+ students per semester.

Union Pacific | *Technology Intern* 05/2022 — 08/2022

- Designed explainable ML regression models to estimate rail shipment prices for customers using XGBoost and SHAP.
- Performed rigorous feature engineering to achieve a 31% RMSE decrease versus UP's existing pricing analytics solution.

GT EPIC Lab | *Undergraduate Research Assistant* 01/2022 — 08/2023

- Analyzed data across 400+ experimental trials to discover optimal human exoskeleton torque assistance profiles.
- Automated a time series data pipeline producing MATLAB structures for efficient access, analysis, and distribution.

Projects

LLM + 10-K | *Streamlit, Plotly, Google Gemini* (GitHub) 05/2024

- Constructed a web interface to query and plot financial metrics extracted from the SEC EDGAR 10-K filings database.
- Leveraged prompt engineering and Google Gemini 1.5 Flash to extract information reliably across different tickers.

Generative Data Augmentation for Image Classification | *PyTorch, Stable Diffusion, ControlNet* (GitHub) 04/2024

- Experimented with multiple image generative models to enhance image classification accuracy when data is scarce.
- Observed a 10% F1 increase for Resnet-50 on a compact dataset when augmented with ControlNet-generated images.

LC3Tools | *C++, Vue, Electron, LC-3 Assembly* (GitHub) 10/2023 — 05/2024

- Lead developer of the educational tooling suite to code, assemble, and simulate assembly programs for the LC-3.
- Added countless quality-of-life improvements through student and instructor feedback as a fork from the original project.

Alaskan Wildlife Image Segmentation | *Python, PyTorch, Pillow* (GitHub) 09/2021

- Utilized and refined the FgSegNet segmentation model to predict and automatically annotate animal presence in image data.
- 1st Award Winner of 2021 Terra NYC STEM Fair and Milton Fisher Scholarship for Innovation and Creativity.

Skills

Programming Languages | Python, TypeScript/JavaScript, C/C++, Go, MATLAB, Java

Frameworks & Libraries | React, Jest, Express, NumPy, Pandas, Scikit Learn, PyTorch

Databases & Misc. | Firebase, PostgreSQL, SQlite, Git, Github Actions, Docker, LaTeX, Vim