

Darren Y. Key

dyk34@cornell.edu | (404) 849-3621 | darrenkey.github.io | github.com/DarrenKey/ | linkedin.com/in/darren-key/

EDUCATION

Cornell University, College of Engineering

Ithaca, NY

Bachelor of Science in Computer Science

Expected May 2025

- Cumulative GPA: **4.122**
- Relevant and Current Coursework: Analysis of Algorithms • Machine Learning • Natural Language Processing • Functional Programming • OOP and Data Structures • Probability Theory • Discrete Structures • Linear Algebra • Applied Combinatorics
- Honors and Awards: Rawlings Presidential Scholars Program (~1%) • 4x AIME Qualifier • Dean's List

EXPERIENCE

Cornell Bowers College of Computing and Information Science

Ithaca, NY

Machine Learning Researcher, Advised by Professor Kevin Ellis

June 2022 – Present

- Researched improving the precision and recall of large language models trained on code, such as OpenAI's Codex
- Discovered pathways for transformer-based large language models utilized in code synthesis to create nontrivial test cases, and utilized these discoveries to develop a novel way of matching generated test cases to generated programs for program ranking
- Built comprehensive testing and evaluation pipelines for generated programs and test cases using python's multiprocessing and concurrent.futures libraries to parallelize testing, speeding the process 9-fold; adopted numpy + pandas for efficient data analysis
- Used Pytorch to create models, including a novel comparison model, to generate scores for ranking generated programs
- Achieved large empirical boost from previous state-of-the-art results of 65.8% to 78.1% on HumanEval, a standard program synthesis benchmark, and raised max F1 score from previous SOTA 0.75 to 0.82; paper submitted to ICLR [1]

Cornell College of Agriculture and Life Sciences

Ithaca, NY

Computational Biology Researcher, Hay Lab

July 2019 – May 2022

- Wrote python programs utilizing matplotlib, numpy, pandas, pybioviz, and biopython to visualize and analyze genomes
- Leveraged various UNIX command line tools such as bedtools and genomic analysis software to perform sequencing analysis
- Conducted PCR of *Sphingomonas* RD1 to confirm astaxanthin production, a high-valued health supplement; paper in-progress

PUBLICATIONS

[1] Key, D., Li, W., & Ellis, K. (2022). I Speak, You Verify: Toward Trustworthy Neural Program Synthesis. arXiv. <https://doi.org/10.48550/arXiv.2210.00848>

PERSONAL PROJECTS

24 Battle Royale – OCaml + ReactJS

February 2022

- Led team of 3 to develop an online multiplayer game involving real-time simultaneous gameplay and a working chat system
- Structured the project with Agile methodology, and used OCaml for backend and Next.js + ReactJS for frontend
- Utilized OCaml's built-in library for promises, Lwt, for concurrency, and created an equation parser using abstract syntax trees
- Employed Google Firebase to implement a NoSQL database of users and OAuth for users to sign-in with Google
- Adopted websockets for 2-way communication to a central server for online multiplayer and deployed using Oracle VMs

iOS Application: YEP Chat – Enhanced Emotes – Swift + UIKit

June 2021

- Designed and programmed an app on the iOS app store to enhance viewer experience on the livestreaming website Twitch
- Structured app with MVC design pattern and Object-Oriented Programming (OOP) fundamentals
- Received over 3000 installs and discussed app with Twitch's Director of Developer Relations

Minimum Spanning Tree Visualization - Python

June 2021

- Created a GUI application with tkinter to visualize Prim's and Kruskal's algorithms used to create minimum spanning trees
- Implemented the ability to add and remove nodes to visualize MST algorithms on custom graphs

LEADERSHIP AND COMMUNITY INVOLVEMENT

Cornell Data Science Project Team

Ithaca, NY

Insights Subteam

February 2022 – Present

- Created website with team of 4 to detect programming language of a given block of code using logistic regression + naïve bayes
- Utilized Scikit-learn for logistic regression and naïve bayes; developed a REST API using Flask and scraped data from Github
- Building computer vision models employing formulae detection to find user-specified LaTeX formulas in scientific papers

Cornell Quant Fund

Ithaca, NY

Quantitative Researcher

January 2022 – Present

- Worked with team of 4 using Markowitz model to create strategy from S&P 500 companies; achieved Sharpe ratio of ~1.57
- Collaborating with team of 9 to create a case involving maximizing P&L on SPY options for trading competition at Cornell Tech

TECHNICAL SKILLS AND ADDITIONAL INFORMATION

Programming Languages: Python • Java • OCaml • Swift • SQL • JavaScript • TypeScript • HTML • CSS

Software Development: Pytorch • Sklearn • Pandas • Numpy • Matplotlib • React • Next.js • Flask • Git • Docker • Jupyter Notebook