

Anant Shyam

914-406-6790 | ais64@cornell.edu | [Website](#) | [LinkedIn](#) | [GitHub](#)

EDUCATION

Cornell University

Ithaca, NY

Bachelor of Science in Computer Science (Honors), Minor in Mathematics

Aug '21 – May '25

- Awards: Cornell Computer Science Outstanding Course Staff Award Nominaton
- GPA: 3.709 (Dean's Merit List Scholar for 5 semesters)
- Graduate Level Coursework: Advanced Machine Learning Systems, Mathematics of Deep Learning
- Teaching Experiences: Systems Programming, Introduction to Analysis of Algorithms, Functional Programming, Object-Oriented Programming and Data Structures, Python Programming

PROFESSIONAL EXPERIENCES

MathWorks

Natick, MA

Engineering Development Group Intern

May '24 – Aug '24

- Developing a pricing assistant using MATLAB which solves classical computational finance workflows for customers such as option pricing, translating econometric models, and formulating portfolio optimization problems.
- Exploring various prompt engineering techniques in order to effectively get pre-trained Large Language Models to properly process customer queries. Plan to explore fine-tuning and model distillation with pre-trained large language models in order to improve accuracy and reduce cost for inference with large models.

Cornell University

Ithaca, NY

Machine Learning Researcher

Jan '24 – Present

- Project: Accelerating Large Language Model Inference on Associative Processing Units (APU) (Advisor: Professor Zhiru Zhang). Optimizing low precision (4-bit) matrix multiplication on APU in order to allow for these computations to be parallelized across multiple cores. Implemented an efficient ReLU activation layer on the APU, whose total time for data movement and computation was just about 88 ms, for high dimensional (4096 x 4096) dimensional inputs. Using Activation-aware Weight Quantization (AWQ) to quantize a pretrained LLaMA-7B model from full precision (16 bits) to a more lightweight 4 bit model for faster inference.
- Project: Distilling GRAFT (Advisors: Dean Kavita Bala, Professor Bharath Hariharan). Developing a more lightweight model to perform similarly to GRAFT, a vision-language model aligning text with satellite imagery.

Cleo

Rockford, IL

Software Engineer Intern

Jun '23 – Aug '23

- Developed a prototype model using Tensorflow which attempts to write code in a hypothetical object-oriented language similar to Cleo's language with accuracy of approximately 60%. Pitched my model to the Director of Product Development Team and explained how it can be expanded to learn how to program in Cleo's programming language.
- Trained and fine-tuned multiple existing models to write code that handles complex data transformations, similar to those that Cleo supports. Pitched these trained models to the Product Development Team, and explained their potential to reduce the room for error and time needed for Cleo to handle data transformations.
- Implemented a feature on the Cleo Integration Cloud, the UI where all of Cleo's customers' jobs are displayed, which allows for direct navigation from a customer's jobs to another customer's jobs using TypeScript, Angular, and RxJS.

Interactive Brokers

Greenwich, CT

Software Engineer Intern

Jun '22 – Aug '22

- Enhanced the tool that the Risk Team used to approve or reject margin changes for equities to reduce room for error and time required to review these margin changes.
- Retrieved margin-rates related data (specifically the margin rule input) from numerous relational databases using MySQL. Stored the margin rule input using Perl, and properly displayed the data for the customer on the UI using HTML, allowing the customer to make more informed decisions about the stocks that they want to purchase.
- Developed a Python script which interprets margin-rates data including the stock type, asset classes, and registration date from a CSV file, and updates the appropriate relational database with this data. Incorporated an argument parser, allowing the user to pass in additional parameters to be put into the database from the command line.

TECHNICAL SKILLS

Languages: Python, Java, OCaml, C, C++, MATLAB

Libraries/Frameworks: PyTorch, Tensorflow, NumPy, Hugging Face