

Aayush Agnihotri

aa2328@cornell.edu | [linkedin.com/in/aayush-agnihotri](https://www.linkedin.com/in/aayush-agnihotri) | +1 (908) 922-9055 | aayush-agnihotri.me | github.com/aayush-agnihotri

EDUCATION

Cornell University

Expected May 2026

Master of Engineering, Bachelor of Arts in Computer Science | Minor in Business | GPA: 4.0

Ithaca, NY

Relevant Coursework: Object-Oriented Programming (OOP) and Data Structures, Machine Learning, Foundations of AI, Analysis of Algorithms, Functional Programming, Operating Systems, Computer Networks, Backend Development

Honors: [CodePath](#) Fellow, 2x 1st Place [JPMorgan Tech for Good](#), 3rd Place [TSA](#) National Software Development

SKILLS

- **Programming Languages:** Python, JavaScript, TypeScript, C++, C, Java, SQL, HTML & CSS, OCaml
- **Libraries & Frameworks:** React & Next.js, Express & Node.js, Flask, Spring Boot, Django, PyTorch, TensorFlow
- **DevOps & Cloud:** Docker, AWS, Microsoft Azure, GCP, DigitalOcean, Jenkins, Kubernetes
- **Databases & Tools:** PostgreSQL, MongoDB, MySQL, GraphQL, Git, Postman, Linux & Unix

EXPERIENCE

Google

May 2025 - August 2025

Software Engineer Intern, Google Cloud Platform Vertex AI

Sunnyvale, CA

- Built a synthetic benchmark generator in Python utilizing critique agents, reducing benchmark creation time by 95%
- Integrated an autorater evaluation framework within a RAG system, leading to a 39% increase in generation quality
- Developed an agentic AI system to generate LLM prompts, resulting in an 8x increase in prompt engineering output

Paramount

Jun 2024 - Aug 2024

Machine Learning Engineer Intern, AI/ML Global Supply Chain Operations

New York City, NY

- Implemented a CNN in Python with TensorFlow to classify ephemeral text within media with 94% accuracy
- Utilized parallel computing via Spark and Jupyter Notebooks to train neural networks, reducing train time by 80%
- Deployed microservice on AWS Lambda and Jenkins using CI/CD pipelines, processing 9,000+ media hours yearly

Johnson & Johnson

May 2023 - Aug 2023

Software Engineer Intern, Intelligent Automation

Skillman, NJ

- Developed a RAG LLM in React, LangChain, and Django with 10,000+ users, achieving annual savings of \$20,000+
- Applied NLP algorithms to create text-based vector embeddings, optimizing document semantic search by 70%
- Integrated Milvus vector database to store/retrieve embeddings and single sign-on (SSO) through Microsoft Azure

PROJECTS

AI Dev | *TypeScript, React, Next.js, MongoDB, Firebase, Docker, GCP Compute Engine*

[GitHub](#)

- Fine-tuned LLaMA3.2 on 2GB of organization data, creating a personalized LLM to use for coding help and ideation
- Engineered a RAG system, enabling real-time, context-aware lookups of GitHub repositories and Notion databases
- Integrated Firebase authentication and role-based permissions, ensuring secure user access and tailored features

Eatery | *Python, Django, PostgreSQL, Docker, AWS EC2*

[GitHub](#)

- Developed a Cornell Dining app with 40,000+ downloads, being used by 10,000+ students and faculty
- Optimized backend response times by 300% via route caching and refactored models to remove redundancies
- Implemented user authentication and authorization, allowing for personalized notifications and protected routes

MusicMaster | *JavaScript, Python, React, Flask, PostgreSQL*

[GitHub](#)

- Built a ML music recommendation engine with React and Flask, using Spotify's REST API and a 2,000+ song dataset
- Created a kNN recommendation model employing cosine similarity with Pandas and Sklearn to recommend songs
- Developed OAuth 2.0 compliant user authentication and utilized PostgreSQL for storing favorited content records

Other Projects:

- **Volume** | *TypeScript, Python, MongoDB, GraphQL, Docker, AWS EC2 & S3* | A Cornell student publication app with 1,000+ users, facilitating the exploration and sharing of content and campus life [GitHub](#)
- **All In** | *Java, Spring Boot, MySQL, Docker, DigitalOcean Droplet* | A sports betting app based on Cornell Athletics, allowing users to place bets on athletes and analyze data collected by a web scraper [GitHub](#)

LEADERSHIP EXPERIENCE

Cornell AppDev

Sep 2023 - Present

Vice Team Lead | Previously Backend Lead, Developer, & Course Instructor

Ithaca, NY

- Lead all development and delivery for an open-source team of 60+ students building apps with 15,000+ users
- Collaborate with iOS, Android, Backend, Design, and Marketing teams using Agile to create and maintain apps
- Taught weekly sessions to 100+ students focusing on client-server architecture, database design, and DevOps

CS 4780 (Machine Learning), CS 4700 (Artificial Intelligence)

Jan 2025 - Present

Teaching Assistant

Ithaca, NY

- Host office hours for 250+ students, teaching traditional ML, neural networks, deep learning, and generative AI
- Work closely with students to prepare them for exams and guide them through complex projects