Ananya Agarwal

+1 (973) 769-2819 | aa13549@nyu.edu | linkedin.com/in/ananya-agarwal03/ | github.com/happyananya

Education

New York University – M.S in Computer Science; CGPA: 3.78/4 — Sept. 2025 – May 2027
Vellore Institute of Technology – B.Tech in Computer Science; CGPA: 3.7/4 — Sept. 2021 – May 2025

Technical Skills

Languages – Python, JavaScript/TypeScript, SQL (PostgreSQL), C/C++, Java, HTML/CSS, Node.js
AI & Machine Learning – PyTorch, TensorFlow, HuggingFace, YOLOv8, OpenCV, Scikit-learn
ML & LLM Systems – Prompt Engineering, Retrieval-Augmented Generation (RAG), Model Evaluation, Incremental Learning, CNN, Vision Transformers (ViT), NLP
Full-Stack & Data – React, Angular, Django, FastAPI, Firebase, D3.js, Fabric.js, Pandas, NumPy
Cloud & DevOps – AWS, CI/CD (Travis CI), Git, Linux, Firebase
Coursework – Data Structures & Algorithms, Big Data Analytics, Operating Systems, Machine Learning

Experience

New York University - VIDA Lab — **Feb. 2026 – Present**
Software Engineer — *New York, USA*

- Modernized a large-scale image intelligence platform from Angular 4 to Angular 20, migrating 31 components, 9 services, and 8 data models to a modular standalone architecture integrated with Firebase (Auth, Realtime DB, Cloud Storage).
- Designed backend data processing pipelines and Cloud Functions (Node.js) for image transformation, metadata ingestion, and access control, enabling scalable handling of high-volume media and structured data.
- Implemented evaluation-driven workflows for validating annotation consistency and system outputs, improving reliability of human-in-the-loop data labeling systems.

Universiti Teknologi PETRONAS — **Jan. 2025 – May 2025**
Machine Learning Engineering Intern — *Seri Iskandar, Perak, Malaysia*

- Developed a scalable code analysis pipeline to detect AI-generated C programs using 9+ software engineering metrics, including Cyclomatic Complexity, Halstead metrics, and entropy-based features.
- Engineered feature extraction and clustering workflows on structured code datasets, achieving a Silhouette Score of 0.48 and identifying distinguishable structural patterns in generated vs human-written code.
- Built automated parsing and preprocessing pipelines for unstructured source code data, enabling downstream ML modeling and evaluation.

Samsung Research — **June 2024 – Dec. 2024**
Machine Learning Engineering Intern — *Remote*

- Engineered an incremental learning pipeline in PyTorch to mitigate catastrophic forgetting in evolving classification systems, retaining 85% accuracy after integrating new data classes.
- Evaluated CNN and Vision Transformer (ViT) architectures under continual learning constraints, analyzing trade-offs in knowledge retention, generalization, and model stability.
- Designed experimental evaluation frameworks to benchmark model performance across sequential datasets, improving robustness of model selection decisions.

Aidash — **June 2024 – July 2024**
Data Science Intern — *Bangalore, Karnataka, India*

- Implemented YOLOv8 architecture to detect alligator cracks on roads, achieving a precision of 91.26%.
- Boosted model recall by 50% using Canny edge detection and CLAHE (adaptive histogram equalization) preprocessing on a 25,000-image dataset.

National Aerospace Laboratories — **Oct. 2023 – Dec. 2023**
AI/ML Research Intern — *Remote*

- Spearheaded VGG19 algorithm implementation, attaining 70% accuracy in symbol and shape recognition using a dataset of 90,000 images.
- Utilized innovative techniques to develop a grid-based display system, achieving an 80% accuracy rate in identifying coordinates for various shapes and symbols.

Projects

InsightFlow | *Python, FastAPI, React, RAG, PostgreSQL, Docker, Prompt Engineering*

- Built a full-stack NL-to-SQL application enabling non-technical users to query relational databases conversationally, with a React chat interface and Dockerized FastAPI backend serving generated SQL and tabular results.
- Engineered a RAG pipeline using FAISS vector search and Sentence Transformers (all-MiniLM-L6-v2) to semantically match user queries to database schema metadata, providing targeted context to Google Gemini for accurate SQL generation.
- Designed an automated schema introspection module that dynamically extracts table and column metadata from Supabase (PostgreSQL), enabling zero-configuration onboarding for new databases without manual schema definitions.

nomz - Restaurant Recommendation App | *Django, React, TypeScript, AWS, PostgreSQL, REST APIs, CI/CD*

- Built a full-stack data platform integrating public health datasets with user-generated signals across 18 data models, enabling large-scale search, ranking, and filtering across NYC restaurants.
- Designed scalable backend APIs and automated data ingestion pipelines, supporting real-time query processing and consistent synchronization of structured datasets.
- Implemented CI/CD pipelines and modular service architecture, enabling production deployment and iterative feature integration.

Student Performance Visualization | *Python, SHAP, UMAP, Regression Modeling*

- Developed an end-to-end ML pipeline on 1,000+ records with 25 features, training and evaluating 10 regression models to predict student performance, achieving a best test R^2 of 0.88.
- Integrated SHAP-based interpretability and UMAP-based dimensionality reduction to analyze feature importance and latent data structure, enabling explainable insights.

Hand Gesture & Facial Movement Controlled Virtual Mouse | *Python, OpenCV, MediaPipe, PyAutoGUI*

- Developed a real-time computer vision system enabling cursor control via hand gestures and facial movements, processing live video streams with low-latency inference.
- Engineered dual-mode interaction pipelines using OpenCV and MediaPipe, supporting dynamic switching between gesture-based and facial tracking input systems.

Next Word Prediction | *Python, RNN*

- The model uses Recurrent Neural Network (RNN) to predict the next word in apps, search engines, virtual assistants, and autocorrect features on smartphones.
- User needs to input base text (a blank string is also accepted) and the number of words they want the model to generate after the base text.

Certifications & Leadership

AWS Certified Solutions Architect Associate - SAA C03

Head of Human Resources, Centre for Social Entrepreneurship and Development, VIT (Jan. 2023 – Jan. 2024)

Optimized club operations by leading training sessions, improving member performance by 15%, and collaborating on strategic initiatives.