

Satyam Chandrakant Chatrola

+1 (973) 905-1864 sc10247@nyu.edu linkedin.com/in/satyamchatrola github.com/Nightshade14 Portfolio GCP Profile

EDUCATION

New York University

Master of Science in Computer Science (CGPA: 3.61 out of 4.00)

New York City, NY

September 2023 – May 2025 (tentatively)

Gujarat Technological University

Bachelor of Engineering in Computer Engineering (CGPA: 3.79 out of 4.00)

Gujarat, India

June 2018 – June 2022

SKILLS

Languages and DBs: Python, SQL, C, C++, Java, PostgreSQL, MySQL, MongoDB, Apache Solr, Elasticsearch, Qdrant, Pinecone.

AI and ML: SVMs, Gradient Boosted Trees, Clustering, **Computer Vision**, NLP, CNN, RNN, **Recommendation and Search Systems**, **Transformers**, **Prompt Engineering**, Generative AI (GANs, Diffusion models), **LLMs (RAG, PEFT, LoRA)**, Differential Privacy.

Data Science: NumPy, Pandas, Polars, Matplotlib, Seaborn, SciPy, Scikit-learn, **PyTorch**, **TensorFlow**, **Transformers**, **Optimum**, **Bitsandbytes**, **Ollama**, **Langchain**, Optuna, Tableau, Statistical Modeling, Stochastic Methods, **Hypothesis Testing**, **A/B Testing**.

MLE, MLOps and Data Eng.: **AWS SageMaker**, **ONNX**, **TensorRT**, MLflow, Evidently, Apache (Hadoop, Spark, Airflow, Kafka).

SWE and Others: **AWS**, **Google Cloud Platform (GCP)**, REST APIs, Git, **Docker**, **Kubernetes**, **CI/CD (CircleCI, GitHub Actions)**, **Prometheus**, Flask, **Django**, **FastAPI**, **Redis**, **Pytest**, System Design, HTML, CSS, JavaScript, Node.js.

EXPERIENCE

Rapidops

Machine Learning Engineer + Data Scientist

Ahmedabad, India

January 2022 – June 2023

Face Authentication and Authorization System (Python, PyTorch, YOLO, MTCNN, FaceNet, Triplet Loss, Qdrant, OpenCV, dlib)

- Spearheaded a face authentication and authorization system with **4 environments and 2 access points** while **leading a team of 3**, from **data collection through on-premises deployment and monitoring**; enhanced productivity tracking and premises security.
- Architected a **scalable microservice** utilizing YOLO, MTCNN, fine-tuned FaceNet model, and Qdrant to achieve **low-latency** authentication and authorization on **live video streams**, delivering a **0.93 F1 score** without need to retrain for new individuals.

AI-Powered Search and Recommendation System (Python, PyTorch, Hybrid Recommenders, Apache Solr, Docker, FastAPI)

- Boosted **conversion rate** by **23%** and **click-through rate** by **42%** with **collaborative, content, and market-basket** recommenders.
- Engineered search & recommendations with **custom taggers & LTR** techniques with **Apache Solr** to serve results in **25ms**.

Natural Language (English) to SQL query generation (Python, TensorFlow, Keras, T5, BERT)

- Researched and benchmarked **State of the Art (SOTA) AI models** generating SQL from Natural Language with **76% EMA**.
- Experimented with Transformers (**T5, BERT**) generating **73% Exact Match Accuracy (EMA)** with **36%** faster inference.

RESEARCH EXPERIENCE

Approaches to Type 2 Diabetes Mellitus Prediction with Machine Learning and Deep Learning

- Researched AI techniques for Type-2 Diabetes Mellitus classification with **95.8% precision, 95.8% recall**, and **99.4% specificity**.

PROJECTS AND OPEN-SOURCE CONTRIBUTIONS (with Embedded GitHub Repository URLs)

RAG Microservice: Research-mate chatbot (Python, FastAPI, PyTorch, Transformers, GCP, Pinecone, Prompt Engg, Llama 3.2)

- Engineered a **context-aware RAG-based** chatbot leveraging **Pinecone vector database**, enabling semantic search across **2,700 research papers** with **95% query relevance** by **Anthropic AI's Contextual Retrieval** technique with fast inference.
- Optimized model performance with **Binary Quantization**, achieving **7x speedup** in **inference time** and **85% reduction** in **memory**.

Microservices webapp: LLM Essay Evaluator (Python, PyTorch, ONNX, TensorRT, FastAPI, AWS, MLflow, Evidently, Frontend)

- **Calibrated** Transformers (**BERT**) and LLMs (**GPT-2**) with **PEFT techniques** (quantization), cosine-annealed learning rate and warm-up to assess essays, attaining a **Kappa Score of 81.7%** and **surpassing the Benchmark** score by **5.7%**.
- Orchestrated **2 fault-tolerant microservices** & inferred with **ONNX models** and **TensorRT** via **asynchronous REST APIs**.

Open Source Project: mAigic (Python, OpenAI Function Calling, Prompt Engineering, CircleCI, Pytest, MyPy, Ruff, uv)

- Architected an AI-powered email management system leveraging **OpenAI's function calling API**, achieving **95% accuracy in task extraction** and automated Trello board updates, **reducing manual email processing time by 70%**.
- Engineered a **production-grade Python API** with **100% test coverage** through **CircleCI pipeline**, implementing comprehensive **static type checking** with mypy, and leveraged **SQLite-based** conversation tracking system.

Open Source Project: ETL pipeline migration to Spark (Python, PySpark, AWS, MinIO, Databricks.Koalas)

- Migrated **ETL pipeline** to **Spark** for NYU's **Open-Source** project with **160% speedup** in feature extraction for AI tasks.

CERTIFICATIONS

- Graduated from **Udacity's AWS Machine Learning Engineer Nanodegree** with top remarks.