

Satyam Chandrakant Chatrola

satyamchatrola14@gmail.com [linkedin.com/in/satyamchatrola](https://www.linkedin.com/in/satyamchatrola) github.com/Nightshade14

Experience

Machine Learning Engineer, Rapidops – Ahmedabad, India

January 2022 – June 2023

Personalized Search and Recommendations

(Python, PyTorch, Apache Solr, Docker, Kubernetes, FastAPI)

- Engineered a hyper-personalized search system for **5M+ SKUs** using hybrid search (semantic + lexical), Learning-to-Rank (LTR) models, and **multi-stage re-ranking**, achieving **92%** relevance accuracy.
- Spearheaded **6 recommendation strategies** leveraging real-time behavioral signals to drive **7.2% conversion lift**, **34% CTR** surge, within 6 months through hyper-personalized profile generation and geo-contextual targeting.

Biometric Access Management System

(Python, PyTorch, YOLO, MTCNN, Triplet Loss, FaceNet, Qdrant)

- Architected a facial recognition authentication system utilizing YOLO, MTCNN, fine-tuned FaceNet model, and Qdrant on **realtime** video streams, identifying individuals with **0.96 F1** score across **750+** individual profiles.
- Designed **real-time attendance** and **blacklist alert system** with **RBAC**, reducing manual efforts by **80%** and enhancing security and integrated with existing HRIS platforms.

Research Experience

- Benchmarking Fine-Tuned Transformers, LLMs and LSTM Networks for Automated Essay Scoring ([Link](#))
- Approaches to Type 2 Diabetes Mellitus Prediction with Machine Learning and Deep Learning ([Link](#))

Skills

Languages and DBs: Python, JavaScript, TypeScript, Java, SQL, MySQL, MongoDB, Apache Solr, Qdrant, Pinecone

AI and ML: Computer Vision, Natural Language Processing, Transformers, **Recommendation Systems**, **Search Systems**, **Large Language Models (LLMs)** (RAG, PEFT, QLoRA), Mixture of Experts (MoE), Model Context Protocol

Data Science: NumPy, Pandas, Scikit-learn, **PyTorch**, **TensorFlow**, HuggingFace, MLflow, Tableau, A/B Testing

Others: REST APIs, Flask, FastAPI, **AWS**, **Google Cloud**, **Azure**, LangChain, Ray, Nvidia (**TensorRT**, **Triton**, **Dynamo**), Evidently, Redis, Apache (Hadoop, Spark, Kafka), **Docker**, **Kubernetes**, **CI/CD** (CircleCI, GitHub Actions)

Projects

RAG WebApp: Research-mate ([Link](#))

(Python, FastAPI, PyTorch, RAG, GCP, Pinecone, Llama 3.2, JavaScript)

- Engineered a **RAG**-based chatbot and hybrid search across **2,700** research papers with **95%** query relevance by leveraging **Anthropic AI's Contextual Retrieval** technique while resolving cold-start issues with warm-up.
- Optimized system with Quantization, achieving **7x speedup** in inference time and **85% reduction** in memory.

Microservice: LLM Essay Evaluator ([Link](#))

(Python, PyTorch, ONNX, FastAPI, AWS, MLflow, Evidently, JavaScript)

- Fine-tuned Transformers (BERT) and LLMs (**GPT-2**) with PEFT techniques (quantization), cosine-annealed learning rate and warm-up, attaining a Kappa Score of **81.7%** and surpassing the Benchmark score by **5.7%**.
- Designed **2 microservices** and leveraged low-latency inference techniques like **ONNX** models and **TensorRT**.

Open Source Project: mAIgic ([Link](#))

(Python, SQLite, OpenAI Function Calling, CircleCI, Pytest, MyPy, Ruff, uv)

- Architected an email management python package with 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 API for the package with **80% test coverage**, automated through **CircleCI**.

Certifications and Achievements

- Secured **1st Runner Up** in **Qualcomm x Microsoft on-device Edge AI** Hackathon.
- Graduated from Udacity's **AWS Machine Learning Engineer Nanodegree** ([Link](#)).

Education

New York University – MS in Computer Science (CGPA: 3.67)

September 2023 – May 2025

Relevant Coursework: MLOps, Efficient AI and Hardware Accelerator Design, High Performance ML, Deep Learning

Gujarat Technological University – BE in Computer Engineering (CGPA: 3.79)

June 2018 – June 2022