Aatmaj Amol Salunke

Boston, MA | (617) 708-9926 | salunke.aa@northeastern.edu
aatmaj28.github.io/Portfolio | linkedin.com/in/aatmaj-salunke-7106041b0 | github.com/aatmaj28

EDUCATION

Northeastern University, Boston, MA — Sept 2024 – Dec 2026
M.S. in Artificial Intelligence (Khoury College of Computer Sciences) — CGPA: 3.78/4.0
Courses: Reinforcement Learning, Foundations of AI, Algorithms, NLP, Generative AI, Applied Programming for AI

Manipal University Jaipur, India — Oct 2020 – May 2024
B.Tech in Computer Science and Engineering — CGPA: 3.62/4.0
Courses: NLP, Deep Learning, Design & Analysis of Algorithms, Database Management Systems, Computer Vision

EXPERIENCE

AI Software Engineer Research Assistant | DASH Lab | Boston, MA, USA — Sept 2025 – Present

- Built the document ingestion and RAG pipeline for an on premise AI Teaching Assistant serving 500+ students, extracting text from Canvas course files with Docling and indexing them via LlamaIndex semantic chunking, nomic-embed-text-v1.5 embeddings, and Qdrant hybrid search.
- Developed AuditBOT with Next.js, TypeScript, and the Claude API for accounting students, plus an admin dashboard giving faculty researchers exportable quantitative and LLM based qualitative analytics (decision scoring, confidence calibration) across 500+ user sessions.

AI Training Data Engineer | LiquidPixels, Inc. | Billerica, MA, USA — May 2026 – Aug 2026

- Benchmarked 5 model configurations (Qwen2.5-Coder-14B, Qwen3-Coder-30B, GPT-5.6 at three reasoning tiers) on a 25 case suite via controlled ablation, showing prompt scaffold ablation cost 48 render coverage points, 2.4× any model upgrade, and open weight models matched frontier APIs within noise.
- Protected confidential data by closing a credential egress path where 127 of 582 corpus chunks carried AWS keys, customer identifiers, and internal hostnames into third party model prompts, shipping shape preserving redaction across 6 formatter sites at 0.05% text cost.

AI Engineer | Borealis Global Advisory LLC | South Carolina, USA — May 2025 – Dec 2025

- Engineered custom Python pipelines to process, classify, and structure 2TB+ of academic papers and business case studies into searchable datasets, enabling rapid knowledge retrieval across all downstream applications.
- Developed a Google ADK multi agent forecasting system coordinating 15 agents over 5 data sources, including macroeconomic indicators and PostgreSQL sentiment databases, across 100+ countries, with DDL based audit logging of every workflow decision for full traceability.

ACADEMIC PROJECTS

Void AI – Alpha Discovery in Under-Covered Equities — *CrewAI, RAG, Haystack, PostgreSQL, MCP*

- Built a multi factor coverage gap scoring engine across ~2,900 stocks using yfinance, SEC EDGAR, Finnhub, and Tavily data, with an ML weight optimizer (Linear, Ridge, Random Forest, XGBoost) and a quintile backtest showing a +1.56% Q1–Q5 spread at a 30 day horizon.
- Engineered a Haystack v2 RAG assistant over SEC EDGAR filings using pgvector on PostgreSQL and BAAI/bge-small-en-v1.5 embeddings, with an LLM as a judge validation layer catching false positives such as penny stock traps and volume anomalies.

ResilienceAI – Global Supply Chain Risk Intelligence Platform — *Google BigQuery, Mapbox*

- Engineered data pipelines in Python, Pandas, and NumPy unifying heterogeneous sources across 266 countries and 25 years: 72K+ GDELT conflict events via Google BigQuery, ERA5 climate data via Google Earth Engine, EM-DAT disasters, FAOSTAT, WHO alerts, and World Bank indicators.
- Modeled multi hop supply chain disruption with NetworkX dependency graphs and Monte Carlo simulation for probabilistic risk scoring, reaching 82% directional accuracy, and surfaced results in a Next.js and Mapbox dashboard with exportable PDF risk briefings.

SKILLS

Languages: Python, SQL, TypeScript, JavaScript, C++, Java
AI/ML: pandas, NumPy, scikit-learn, PyTorch, Hugging Face Transformers, LLMs, RAG, Embeddings, Prompt Engineering, LlamaIndex
Backend & Databases: PostgreSQL, MySQL, SQLite, Google BigQuery, pgvector, Qdrant, FAISS, FastAPI, Selenium, Beautiful Soup
Cloud & DevOps: Azure, AWS (EC2, S3, Lambda), Git, Jupyter Notebooks, Docker, GitHub Actions, CI/CD, Linux
Core CS & Data: Data Cleaning, Data Validation, Exploratory Data Analysis, Statistical Modeling, Model Evaluation, Feature Engineering, Unit Testing (pytest), Code Review, Technical Documentation, Reproducible Workflows