

Arsh Jafri

856-509-9650 | jafri.ar@northeastern.edu | arshjafri.com | [LinkedIn](#) | [GitHub](#) | U.S. Citizen

EDUCATION

Northeastern University

Boston, MA

Bachelor's Degree in Computer Science and Economics

May 2027

Coursework: Algorithms & Data Structures, Object-Oriented Design, Artificial Intelligence, Database Design, Discrete Math

Honors: University Honors Program, Honors Scholarship, Dean's List (Fall 2024, Spring 2025)

TECHNICAL SKILLS

Languages: Python, Java, Go, JavaScript, TypeScript, SQL, Swift, R

Frameworks & Libraries: Node.js, Flask, FastAPI, React, LangChain, PyTorch, Pandas, NumPy, HTML/CSS

Databases: PostgreSQL, MongoDB, MySQL, PgVector

Developer Tools: AWS (Lambda, EC2, CloudWatch), GCP, Docker, Linux, CI/CD, Git, PyTest, REST APIs, Celery, RabbitMQ, Redis

Concepts: Full Stack, Backend, System Design, Machine Learning, LLM Applications, Distributed Systems

EXPERIENCE

PwC (PricewaterhouseCoopers)

Sept 2025 – Present

AI Engineering Co-op

Boston, MA

- Developed a multi-agent system using **LangGraph** and PwC's Code Intelligence **MCP server** to automate security design documentation and architecture diagram generation for enterprise codebases.
- Reduced security documentation delivery time by **71%** across **4** client engagement teams.
- Optimized agent orchestration for **1M+** LOC codebases using **Celery** with RabbitMQ/Redis for parallel processing.
- Deployed agent workflows on **AWS Lambda** with **Docker** and **CI/CD** pipelines for rapid iteration and scalability.

PlateMate (Startup)

Apr 2025 – Sept 2025

Software Engineering Intern

Boston, MA

- Engineered **RESTful APIs** and backend pipelines in **Go** and **Node.js** to deliver real-time restaurant recommendations.
- Improved recommendation accuracy by **37%** developing an embedding-based content filtering algorithm in **Python**.
- Reduced response times by **25%** via parallel async execution and optimized **MongoDB** connection pooling.
- Deployed distributed backend on **AWS Lambda** and **EC2** with caching and CloudWatch, achieving **<150ms** latency.

Disrupt FinTech Consulting

Dec 2024 – Apr 2025

Software Lead

Boston, MA

- Built a data analytics platform for a **\$140M** private equity firm, cutting time-to-insight by **65%** for executives.
- Boosted non-technical user adoption by **25%** by integrating **Gemini 1.5 Pro** for natural language SQL querying.
- Developed RESTful APIs and microservices with input validation, **error logging**, and role-based access controls.
- Deployed platform on **GCP** using Dockerized services with Python, React, TypeScript, and PostgreSQL.
- Directed roadmap and delivered **technical demos** to stakeholders, aligning development with client needs.

PROJECTS

StarkBot - RAG Chatbot | Python, FastAPI, React, PostgreSQL, PgVector

Sept 2025

- Built a **RAG chatbot** to answer domain-specific questions about Iron Man with grounded citations and attribution.
- Implemented **PgVector** for **semantic search** and fine-tuned embedding models to improve retrieval relevance.

Tripful - AI Travel Planner | React, Node.js, TypeScript, REST APIs

May 2025

- Developed an AI-powered full-stack personalized itinerary generator with **real-time** flight, hotel, and activity data.
- Integrated **OpenAI**, Amadeus, and Google Places APIs with re-ranking logic to tailor results to user preferences.
- Cut third-party API latency by **35%** through request batching and caching.

Clearview - Bias Detection Chrome Extension | JavaScript, NLP, Sentiment Analysis, Web Scraping

Feb 2025

- Built a Chrome extension with **200+ installs** to detect political bias in news articles using NLP and sentiment analysis.
- Achieved **87% accuracy** against expert-labeled datasets across 2,000+ articles.
- Engineered DOM parsers and background logic to maintain **<1s load** times on dynamic news sites.

Econostats - Economic Data Visualization Platform | Python, AWS, Flask, Pandas, NumPy, Plotly

Jan 2025

- Developed a real-time visualization platform integrating **FRED API** with interactive charts and custom dataset uploads.
- Optimized data pipelines for **40%** lower latency, enabling high-throughput access during peak load.