

SUJENDRA JAYANT GHARAT

Boston, MA | (857) 930-1933 | gharat.su@northeastern.edu | [linkedin.com/in/sujendra-gharat](https://www.linkedin.com/in/sujendra-gharat) | github.com/suju297

EDUCATION

Northeastern University , Boston, MA	Dec 2025
Master of Science in Information Systems	GPA : 3.68/4.0
University of Mumbai , Mumbai, India	May 2018
Bachelor of Engineering, Electronics Engineering	

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, Java, Bash

Backend/Frameworks: FastAPI, Django, Flask, Node.js, Express, React

ML/AI: PyTorch, Transformers (Hugging Face), LangChain, OpenAI SDK (GPT-4o, o4, GPT-5), Llama3

Cloud/DevOps: AWS, GCP, Docker, Kubernetes, Terraform, Git, GitHub Actions

Databases: SQL (MySQL, MSSQL), NoSQL (MongoDB, Firebase)

EXPERIENCE

Software Engineer Intern (Co-op)	Jan 2025 – Aug 2025
Xellar Biosystems	<i>Boston, MA</i>
<ul style="list-style-type: none">Developed a FastAPI microservice with Redis lookups that converts 16-bit images on first load, stores derived results, and serves repeat requests from cache, achieving 0.4 second median load timeBuilt an S3 uploader in Python (boto3) with parallel uploads, retries, and conflict-handling rules (overwrite, rename, skip), ingesting terabyte-scale data in a non-versioned bucketStandardized raw microscopy data into a canonical S3 structure and auto-generated a per-file manifest used by the FastAPI service for Redis lookups and reliable retrievalDesigned a template-driven experiment tool with a React front end and Django backend for 8 and 32 chip organ-on-chip plates, enabling reusable layouts and automatic drug and dose assignment for consistent study setup	
Graduate Research Assistant – AI-CARING IMWUT submission (First Author, Nov 2024)	Feb 2024 – Aug 2025

- Deployed a **multi-stage pipeline** that extracts reminder intent from caregiver conversations and compiles it into AST-validated Python functions.
- Designed a feasibility checker using the OpenAI **o4 mini reasoning model** that validates reminders against available sensors and a **home layout graph** before generation, **preventing non feasible reminders**
- Deployed a low-latency **AWS IoT Core** to **gRPC** middleware streaming data from over 120 sensors (about 3,000 events per second) and achieved a 50% reduction in end-to-end latency.

Senior Software Engineer

Feb 2022 – Aug 2023

Capgemini

Mumbai, Maharashtra

- Orchestrated RESTful API calls for Multi-Modality AI using **Python** and **Flask**, achieving 40% improvement in response times
- Maintained 99.9% uptime, boosted scalability, and reduced resource costs by deploying applications on **Kubernetes** clusters with Agile methodologies
- Authored automation scripts using **Batch** and **Bash**, utilized by 50+ team members, reducing support dependencies by 40%

Software Engineer

Aug 2018 – Feb 2022

LTIMindtree

Mumbai, Maharashtra

- Led integration of 30+ third-party RESTful and SOA APIs in **Node.js**, collaborating with cross-functional teams and vendors
- Enhanced **REST API** performance by leveraging advanced concurrency and asynchronous patterns in **JavaScript/TypeScript**, and implementing **MongoDB** caching in Node.js, resulting in a 15% boost in response times and a 30% reduction in API calls
- Implemented **graceful shutdowns** and **process monitoring** in Node.js, reducing recovery time from errors by 50%, preventing resource leakage by 20%, and maintaining uninterrupted application operation with 99.9% uptime

ACADEMIC PROJECTS

Quant Stack Exchange Chat Assistant

Sep 2024 – Oct 2024

- Developed FinRobot, a generative AI chat assistant for quantitative finance using AutoGen with RAG, delivering context-aware answers with approximately 85 percent response accuracy.
- Engineered a scalable knowledge layer with a daily Python scraping pipeline that ingests Quant Stack Exchange into Neo4j (knowledge graph) and Pinecone (vector index) to enable fast, entity-aware retrieval.

Cloud Native Web App

Jan 2024 – Apr 2024

- Provisioned **Packer** and **Terraform** to provision pre-configured machine instances, resulting in a 75% reduction in configuration time and facilitating swift deployment of infrastructure changes
- Implemented **serverless** user verification system with **Cloud Function** for email verification and tracking in **Cloud SQL**