
EDUCATION

University of Massachusetts Amherst

Aug. 2023 – May 2025

*Master of Science in Computer Science; CGPA 3.92/4**Amherst, MA, USA**Relevant coursework: Advanced NLP, Systems for Data Science, Systems for Deep Learning, Distributed and Operating Systems***University of Mumbai (S.P.I.T.)**

Aug. 2019 – May 2023

*Bachelor of Technology in Information Technology, Minor in Management; CGPA 3.72/4**Mumbai, MH, India**Relevant coursework: Data Structures, Algorithms, Distributed Computing, Databases, Computer Networks*

EXPERIENCE

Intellect Design Arena – Software Engineer (Product)

NJ, USA | July 2025 – Present

- Engineered **10+** core features for [Xponent](#), a commercial insurance underwriting workbench, automating compliance checks, risk analysis, and premium calculations across broker and MGA workflows.
- Developed internal tooling and admin module with **6+** interfaces for configuring underwriting rules and system settings with client-side caching, reducing configuration load times by **60%**.
- Architected an asynchronous audit trail using Spring Boot and AWS SQS to track state changes across UI, offloaded payload comparison to background services to maintain low API latency and cutting audit review time by **65%**.
- Integrated [Risk Analyst \(RA\)](#) into underwriting workflows, enriching and validating **8,000+** submission data fields to improve risk assessment accuracy for Commercial and Specialty lines.
- Owned production reliability for [Magic Submission](#), an AI-driven submission ingestion pipeline, sustaining **99%+** uptime via log-based monitoring and fast RCA across distributed ingestion pipelines using Grafana, AWS services and MongoDB.
- Tools:** Java, Spring Boot, Angular, TypeScript, AWS SQS, PostgreSQL, MongoDB, Camunda, Kibana, Grafana, Loki, Jenkins

Priro Systems – Software Engineer (Product)

TX, USA | June 2024 – Dec 2024

- Built [SpecGenie](#), an AI-powered B2B workflow automation platform from concept to production for large-scale EPC projects, compressing setup time from **40 hours to 2 hours (95% reduction)** and enabling deployment across 3D modeling tools.
- Architected agentic document extraction leveraging AWS Textract and LLMs, converting unstructured piping specs to structured enterprise data with **98% accuracy** (processing **1000+** components in **under 2 minutes**).
- Designed compound AI systems combining LLM orchestration and rule-based validation to generate **35+** bulk-loadable specification and catalog Excel sheets, eliminating **95%** of manual data entry.
- Implemented dashboards with React and Material UI, integrating **15+** RESTful endpoints, enabling material engineers to review, edit, and version-control extracted data in a collaborative cloud environment.
- Tools:** React, Material UI, AWS (Textract, S3), LangChain, Celery, Redis, Python, Django, JavaScript

LokaVidya Technologies – Software Engineer Intern (Backend)

Mumbai, IN | July 2021 – Jan 2022

- Developed [LokaVidya Meet](#), a video conferencing platform using Node.js and BigBlueButton APIs to deliver **6+** meeting features, supporting secure virtual meetings across educational institutions.
- Migrated RBAC from in-memory storage to Redis for session and role caching, reducing access latency by **87%** and offloading Azure SQL queries on subsequent requests.
- Tools:** JavaScript, Node.js, Express.js, React, Microsoft Azure, SQL, MongoDB

HIGHLIGHTS

Salient Entity Linking (SEL) - Apple

- Architected Salient Entity Linking system in collaboration with Apple's Siri and Search team, extending [ReFinED](#) framework with fine-tuned salience prediction models to enable efficient indexing and retrieval across web-scale news datasets for Safari.
- Built and evaluated bi-encoder, cross-encoder, and late-interaction architectures for entity salience prediction, achieving **+24 F1** over prior benchmarks. Submitted to ACL 2026.

Automatic Short Answer Grading (ASAG) - CMinDS

- Developed ASAG system for free-form student responses combining stacked Flair embeddings, alignment scoring, and lexical similarity features; achieved **0.667 Pearson correlation** on the Mohler dataset, outperforming prior baselines by **13%**.
- Implemented preprocessing and per-question training pipeline with custom spell correction, word normalization, and cosine similarity variants, improving semantic matching accuracy by **15%**.

Reflective Prompting: Self-Refine RAG for Piazza QA

- Implemented [Reflective Prompting](#), an iterative RAG pipeline using FAISS and Llama-3-8B, applying self-refinement over retrieved context to improve factual QA on private Piazza dataset, achieving **+8%** accuracy over vanilla RAG.

SKILLS, TOOLS & LANGUAGES

Languages: Python, Java, C/C++, JavaScript, TypeScript**Full Stack:** Django, Flask, FastAPI, Node.js, Spring Boot, Next.js, React, Angular, Svelte, Postgres, MongoDB, SQL, Supabase**Data Science & AI/ML:** PyTorch, TensorFlow, NumPy, Scikit-learn, Matplotlib, Seaborn, spaCy, MLflow, Unsloth, vLLM**DevOps:** Docker, Docker Compose, AWS, Azure, Kubernetes, Redis, CI/CD, Jenkins, Grafana, Loki

LEADERSHIP

- Led team of **20+** as **Chairperson, CSI**, raising **\$6k+** for **9** tech events including S.P.I.T Hackathon (**150+** teams).
- Managed **100+** members as **Vice-Chairperson, Oculus**, raising **\$10k+** and coordinating operations for **5000+** attendees.