Yash Thakkar

thakkaryash21@gmail.com • yash-thakkar.com • linkedin.com/in/yashthakkar21/ • 412-699-0335

EDUCATION

Carnegie Mellon University, Pittsburgh, PA — Dec 2025
Master of Information Systems Management — **GPA - 4.04**
Relevant Coursework: Cloud Computing, Distributed Systems, LLM Applications, Machine Learning, MLOps
Teaching Assistant: Engineering Data-Intensive Scalable Systems, Data Science & Big Data, Agile Methods

University of Mumbai, Dwarkadas J. Sanghvi College of Engineering — May 2022
Bachelor of Engineering, Mechanical Engineering — **GPA - 3.92**

SKILLS

Languages & Libraries: Python, TypeScript, JavaScript, Java, SQL, NumPy, Pandas, Bash/Shell
AI & ML: PyTorch, Hugging Face, LangChain, Prompt Engineering, RAG. Scikit-learn, MLflow
Infra & DB: AWS, GCP, Azure, Docker, Kubernetes, Terraform, Kafka, MySQL, MongoDB, Redis, Neo4j
Web: FastAPI, Flask, React.js, Redux, Node.js, Next.js, GraphQL, REST API, Pytest, Tailwind, WebRTC, HTML/CSS

EXPERIENCE

Newfront Insurance, San Mateo, CA — May 2025 – Aug 2025
Software Engineering Intern

- **LLM Document Processing:** Built automated pipelines to extract & review benefits plan data using **FastAPI, GraphQL, PostgreSQL**, & OpenAI APIs, **reducing processing time by over 90% & saving $175k annually**
- **GraphQL Performance Optimization:** Improved GraphQL performance by optimizing data retrieval paths & streamlining database queries, **reducing server load & improving latency by 450 ms**
- **AI Pipeline Enhancement**: Improved text extraction accuracy through enhanced **prompt engineering** & context-aware document **chunking**, **achieving 7% higher recall & 5% better precision**
- **System Monitoring**: Implemented comprehensive monitoring with **over 20 Datadog alerts** & an **analytics dashboard** to track system adoption, usage patterns & cost savings, **boosting operational reliability**

Advect Systems, Mumbai, India — Jan 2022 – Aug 2024
Founding Software Engineer

- **Full-Stack Product Development**: Led end-to-end development of Nanofactory, a decentralized 3D printer management & automation app using **Svelte, Flask, & WebRTC**, to enable **remote manufacturing workflows**
- **CI/CD & Automation**: Built comprehensive CI/CD pipelines with **Docker, GitHub Actions**, Cloudflare & Python Integrations, **reducing time-to-market by 40% & improving team productivity by 34%**
- **Test Coverage:** Built end-to-end & unit test suites with **Playwright & PyTest**, **achieving 80% code coverage** across critical application workflows

PROJECTS

Real-Time Air Quality Forecasting System — Oct 2025

- **Kafka Streaming Inference:** Developed a real-time inference pipeline using **Kafka with over 100 engineered temporal features** to predict CO levels, achieving an **R^2 accuracy of 0.96**
- **Model Registry & Drift Monitoring**: Integrated **MLflow** for versioned deployments & **Evidently** for daily/weekly drift detection, **improving model governance**

AI-Powered Learning Portal — Oct 2025

- **RAG Pipeline:** Engineered a modular retrieval system with **query reformulation & ChromaDB** vector search, improving multi-turn grounding & quiz generation across 20 curated sources
- **CoT Evaluation:** Implemented a structured 5-step Chain-of-Thought, evaluated via LLM-as-a-Judge, **achieving 4.47/5 reasoning score**

Twitter Social Graph Analysis — Sept 2025

- **Graph Processing:** Analyzed a Twitter social graph using **Spark's iterative execution** model and PageRank algorithm over a **10.4 GB dataset**, identifying influential users with a 15-minute end-to-end job runtime
- **Performance Debugging:** Diagnosed Spark job bottlenecks by inspecting **YARN logs** and **Spark Web UI**, applying targeted tuning under constrained cluster resources to **improve execution efficiency by 74%**