

Shubh Gupta

shubhgupta.dev | shubhguptasg07@gmail.com | linkedin.com/in/shubhngupta | github.com/ShubhG7 | +1 617 982 4304

Experience

BNY Mellon

Software Developer

Remote, USA | Jun 2024 – Present

- Engineered a Kafka + Spring Boot monitoring pipeline to flag unusual payments, taking detection from **minutes to seconds**, boosting speed by **40%**, and lowering false positives
- Developed event-driven settlement and reconciliation services (Spring Boot, JPA/Hibernate), exposing secure APIs through AWS API Gateway and WAF for high-reliability transaction flows
- Implemented GPT-4 workflows with LangChain and Pinecone to summarize emails, chats, and documents for compliance, saving **hours per week** and improving audit readiness by **50%**
- Deployed services on AWS EKS/Fargate with Docker and Helm; added CloudWatch, X-Ray, and SNS alerts to shorten detection time and reduce on-call overhead

Ernst & Young LLP

Software Engineer

Mumbai, India | Jul 2021 – Jul 2023

- Streamlined fraud pipelines to reduce training time by **40%** and lower false positives by **25%**, speeding reviews
- Lowered cloud spend by **20%** with tagging, auto-shutdown, and cleanup automation, removing idle resources
- Automated ML deployments on AWS EC2 with scripted rollouts, reducing release time and lowering launch by **30%**
- Established Splunk dashboards and alerts to surface failures earlier, improving detection by **35%** and speeding triage

Kings Expomedia Ltd.

Software Engineer

Mumbai, India | Dec 2019 – Apr 2021

- Re-platformed the flagship site from PHP to **Next.js** with a mobile-first redesign, increasing session time by **30%**
- Optimized Core Web Vitals (LCP/FID) by **40%** using lazy loading, bundling, and route-based code splitting
- Integrated Strapi CMS with automated publishing, reducing turnaround by **50%** and saving hours of manual edits
- Led releases end-to-end from design to deployment, increasing delivery speed and reducing rework for stakeholders

Projects

[Hivemail](#) | Privacy-First AI Email CRM (LangGraph, pgvector, DuckDB, Parquet, Cloud Run)

- Built an **AI email memory layer** that connects threads into shared context for chat and meaning-based search, with **3** privacy redaction modes and **BYOK keys** so users control cost and data sharing
- Orchestrated a LangGraph workflow that routes about **70%** of emails through rules at **zero LLM cost**, and cut analytics latency from **5–30s** to **under 1s** with a **Parquet-backed query layer** on Cloud Run

[PatchPilot](#) | GitHub Maintenance Agent (Python, LangGraph, GitHub Actions, Pydantic, Gemini)

- Delivered an agent that converts GitHub issues into ready-to-review pull requests by scanning the repo, selecting relevant files, posting a concise plan, and producing a **git apply**-ready unified diff patch
- Hardened reliability with guardrails and checks: allowlisted paths, capped diff size, up to **3** repair retries, and CI verification using **ruff** and **pytest**, with JSONL run logs uploaded as Actions artifacts

[Tubot](#) | LLM Tutor, Interest-Based Personalization, Microservices (Go, Python, gRPC, OIDC, PostgreSQL/pgvector, Prometheus/Grafana)

- Designed an LLM tutor that teaches with **interest-based analogies** and short quizzes, using an HTTP gateway and gRPC quiz engine to keep responses fast and sessions convenient
- Implemented personalization with **10** categories on a **1–5** scale; preseeded analogy data, validated **5/5** API tests across **3** subjects, and kept usage costs low with cached responses

Education

Boston University

Masters of Science in Computer Science

Boston, MA

Aug 2023 – Jan 2025

Mumbai University

Bachelors of Engineering in Information Technology

Mumbai, India

Aug 2019 – Jul 2023

Skills/Publications

Languages: Python, Java, Rust, C, C++, C#, Go (Golang), Kotlin, JavaScript/TypeScript

ML & DB: PyTorch, TensorFlow, vLLM, LoRA, RAGs, LLMs, GANs, VAEs, MLOps, MySQL, PostgreSQL/pgvector

Systems & DevOps: React, Microservices, gRPC, Protocol Buffers, CI/CD, Docker, Kubernetes, REST API, Git

Cloud/Big Data: GCP, AWS, Spark, Airflow, Kafka, GraphQL, MinIO/S3, NATS, Prometheus, Data Structures, OOP

Certifications/Publications: [Azure AI Engineer Associate](#), [Springer Paper: SVM, ANN, and GARCH](#)