

# Shubh Gupta

[shubhgupta.dev](https://shubhgupta.dev) | [shubhguptasg07@gmail.com](mailto:shubhguptasg07@gmail.com) | [linkedin.com/in/shubhngupta](https://linkedin.com/in/shubhngupta) | [github.com/ShubhG7](https://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

---

### 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

### LegAllity (Legal Document Assistant) | Next.js, FastAPI, Gemini 2.5 Flash, python-docx/docxtpl

- Created a split-screen **.docx** copilot with a **3-step** flow (Upload, Fill, Download), AI-first placeholder extraction (Gemini JSON with regex fallback), and **2** one-click actions per field (“Explain”, “Give example”)
- Delivered format-preserving generation with a **3-stage** pipeline (MailMerge, docxtpl, post-processing) and deterministic normalization for money, dates, emails, and names, enabling multi-field updates and **1-click .docx** output

### DocEdit AI | Next.js, TipTap, Supabase, Claude (Tool Calling), Zustand

- Developed a Notion-style editor with TipTap and **Claude tool calling** that produces **structured edits** (insert, replace, append, suggest, find-and-replace, delete) and applies them client-side with HTML sanitization
- Shipped persistence in **24 hours**: debounced auto-save, per-document chat history stored as JSONB, unload-safe saves via **sendBeacon**, and Supabase Postgres (RLS) **version history with rollback**

## 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](#), [Google Scholar](#)