



GAURAV DEWAN

FULL-STACK + AI ENGINEER

VISIT MY SITE >>

CONTACT

+44 7436 291430

gauravdewan01@gmail.com

London, UK

[Online CV // Dev Portfolio](#)

SKILLS

- Python (FastAPI, pytorch)
- React/Native/NextJS
- LLMs & genAI Applications (quantisation, RAG, deepeval, deployments, fine-tuning)

EDUCATION

UNIVERSITY OF WARWICK

2:1 in Mathematics (BSc)

MERCHANT TAYLORS' SCHOOL

A* Maths | A* Further Maths

A Physics | A Comp Sci

LANGUAGES

- English | Fluent
- Hindi | Fluent
- Panjabi | Fluent

PROFILE

- 6 years as a full-stack developer
- 3 years specialising in AI engineering
- Experience in strategy consulting & project management
- 15+ projects combining Python & JavaScript
- Deep knowledge of generative AI, such as RAG, fine-tuning LLMs & quantisation
- Successfully delivered projects for industry leaders like Barclays, Deloitte & Stripe
- Spearheaded 8 projects as a tech lead

PROJECTS

genAI Chatbot | **FAB Bank [via Deloitte]** 12 months

Engineered a sophisticated genAI-powered chatbot for enterprise use for 2 Fortune 500 companies, totalling 100k+ DAUs.

- Included RAG; document uploads & custom knowledge bases; multimodal support; agentic workflows (including API actions); guardrails & logging; model routing for cost optimisations; multi-model support (GPT-series, Claude & Gemini).
- Tech lead for 2 backend microservices, while also contributing to front-end implementation, genAI integrations & API/micro-services design.
- Responsible for v1 frontend (React); heavily involved in subsequent releases.

Urban Planning Assistant | **NEOM** 12 months

Developed an urban planning dashboard which classifies & scores architectural submissions (renderings, plans & briefs), using a custom genAI model.

- Included a data dashboard to visualize and interact with the AI's urban planning insights.
- Powered by cutting-edge AI, including a quantised image generation model & self-supervised learning.
- In-use by NEOM [Saudi Arabia].

Triaging Assistant | **eConsult [via Deloitte]** 8 months

Developed a multimodal genAI powered assistant that triages incoming healthcare requests & classifies their urgency.

- Includes a real-time visualisation dashboard for medical practitioners.
- Established strategic partnership with a private healthcare consultancy for PoC implementation.
- Collaborated on business strategy to align the solution with consultancy's long-term goals.

AI Institute | **Monitor Deloitte** 5 years

Selected to rotate into the AI Institute, a fast-paced incubator for AI PoCs and client projects.

- Onboarded as a full-stack & genAI developer across 8 projects spanning 12 clients.
- Responsible for ongoing research and horizon scanning within the field of genAI, including workshoping with multiple Fortune 500 companies, to architect and discuss genAI solutions.
- Led 1 week-long genAI training for graduate intake at Deloitte Delivery Institute in Cairo.
- Presented in various talks and seminars on recent genAI trends, while leading the genAI community of practice.

Full-Stack Engineering Intern | **Stripe** 3 months

Built an internal credit platform dashboard and integrated backend systems with Open Banking APIs for real-time financial data management.

- Developed an internal credit platform data dashboard that streamlined reporting and decision-making for financial stakeholders.
- Integrated backend systems with Open Banking APIs to enable real-time data access, ensuring seamless connectivity with external financial institutions.