Shubham Gulati

shubham53.sg@gmail.com | gulati.sh | linkedin.com/in/shubhamgulati02 | github.com/shubhxm02

Experience

Google Jun 2024 – Present
Software Engineer — Cloud Support Platform *Bengaluru, India*

- **Distributed Data Pipelines:** Built and owned distributed data pipelines and services that keep customer support running for **50M+ Google Cloud and Workspace organizations** at **99.9% uptime**.
- **Multi-Agent Harness:** Designed and deployed a **multi-agent engineering harness** to proactively triage production alerts, catch data drift, and mitigate incidents **before** they escalate.
- **Reliability & Testing:** Strengthened production reliability by setting up end-to-end integration tests covering **15+ services** that run in presubmits in **under 5 minutes**, catching breaking regressions before release.
- **Forward-Deployed AI:** Partnered with **Google Pixel Marketing in Germany** (20% project) to forward-deploy **10+ agentic workflows** into **GChat**, automating campaign operations directly in their team environment.

Mindcase Sep 2023 – Feb 2024
AI Systems Engineer *Remote*

- **Legal OCR & Indexing:** Built an **OCR and indexing pipeline** over **50,000+ statutory clauses and Supreme Court cases**, turning dense legal PDFs into searchable records.
- **Citation-Grounded Search:** Partnered with corporate lawyers to build a legal search system delivering **sub-2s answers** with **exact inline citations**, cutting manual verification time by **70%**.

Google May 2023 – Jul 2023
Software Engineer Intern *Bengaluru, India*

- **Async Worker Productionization:** Hardened asynchronous verification workers for production rollout by configuring **ACLs, service permissions, and staged deployment pipelines**.
- **Observability & E2E Reliability:** Built **monitoring dashboards, automated alerts, and end-to-end integration tests** in staging, resolving cross-team infrastructure migration blockers.

Ezyfi Jan 2023 – Apr 2023
Software Engineer Intern *New Delhi, India*

- **Text-to-SQL Engine:** Built schema-pruned **BigQuery** query generation across **40+ warehouse tables**, reaching **91% first-pass accuracy** while cutting prompt tokens by **60%**.
- **Query Safety Rails:** Gated generated SQL behind pre-execution dry-run validation and auto-charting, resulting in **zero broken warehouse runs**.

Projects

Stratos (stratosapp.co) 2024 – Present
Creator & Engineer — Multi-LLM Brand Visibility Platform *Remote*

- **AI Visibility Platform:** Built **Stratos**, a web app that tracks how **ChatGPT, Claude, and Perplexity** rank, describe, and cite brands in AI search answers.
- **Multi-LLM Auditing:** Built automated **Python** pipelines that run **1,800+ prompts per audit** across major LLMs and extract structured brand mentions, citations, and sentiment.
- **Self-Serve Dashboard:** Built the **Next.js, TypeScript, and PostgreSQL** dashboard and scheduled audit reports so teams can compare brand visibility against competitors over time.

Technical Skills

Agentic Systems Infrastructure: Multi-Agent Systems, Agentic Workflows, AI Harness, Tool Calling for Agents, Eval Pipelines & Guardrails, Prompt Engineering, Structured Output, Context Management

Backend and Distributed Systems: Distributed Data Pipelines, Microservices, Event-Driven Architecture, Cloud Spanner, PostgreSQL, BigQuery, Google Cloud Platform (GCP), Amazon Web Services (AWS)

Programming & Engineering: Go, Python, TypeScript, C++, SQL, Next.js, REST APIs, CI/CD, Observability & Alerting, System Design, Data Structures & Algorithms

Education

Delhi Technological University (DTU) 2020 – 2024
*Bachelor of Technology in Information Technology — **CGPA: 8.92 / 10.0*** *New Delhi, India*