

Srishti Agarwal

+91 9650353197 | agarwal11srishti@gmail.com | Gurugram, India | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | [LeetCode](#)

SUMMARY

Founding engineer building backend systems for AI/LLM infrastructure & production deployment; 2 years of experience shipping at a fast-moving AI startup.

PROFESSIONAL EXPERIENCE

Founding Engineer

GrowthZ AI

Full Time

July 2024 – Present

- Core contributor to [GrowthZ.ai](#): an AI platform that unifies ad creative generation, campaign automation, and performance analytics for B2B clients, from system architecture to production deployment.
- Designed scalable REST APIs across 6 microservices powering a multi-tenant campaign automation engine that auto-generates keywords, ad groups, creatives, and extensions; trained and tuned the LightGBM model driving real-time bid optimization.
- Built an LLM-powered ad copy generation pipeline that scrapes Play Store/App Store/website data, extracts sentiment and themes, and generates structured, keyword-researched ad copies with a RAG workflow, embedding brand content with Gemini embeddings and Google Search grounding for context-aware ad group and phrase generation, producing static, animated, and AI-generated video creatives at sub-60s latency.
- Architected an automated video-to-Meta-ad pipeline for one of India's largest Meta advertisers, processing ~1,000 video uploads/day across 60+ ad accounts. Moved uploads to direct-to-object-storage streaming, cutting cloud egress costs ~60%.
- Maintained AWS infrastructure with Docker and Pulumi (IaC), cutting costs 30% via NAT optimization, resource rightsizing, and scheduled shutdowns, while sustaining 99% uptime.
- Built a performance analytics layer aggregating multi-platform campaign data into a unified marketing dashboard, surfacing real-time performance signals to the marketing team.
- Designed and built Creative Suite, an AI ad-copy-to-creative generator, as sole engineer end-to-end. Adopted by 9 clients within 3 months, opening a new monetization channel. Recognized as Top Performer for 3 consecutive months (140% of quarterly variable).
- Independently built a Google Ads audit service running performance & hygiene checks across account categories, cut manual audit time from 3-4 hours to under 3 minutes per account, accelerating client onboarding.

PROJECTS

LLM Sentinel | LLM Observability & Governance Platform

Live

Tech Stack: NestJS, TypeScript, Gemini, Redis, PostgreSQL, SQS

- Built a drop-in reverse proxy between any app and an LLM provider, running synchronous pre-call guardrails without adding latency to the client's request path.
- Designed async LLM-as-judge quality scoring (hallucination, toxicity, faithfulness) off the critical path, feeding a tamper-evident, hash-chained audit log with EU AI Act Article 12-compliant reporting.
- Implemented provider fallback with per-provider circuit breakers, measured at p50 ~29ms / p95 ~46ms gateway overhead under load.

CallSense AI | AI Call-Intelligence Platform

Live

Tech Stack: FastAPI (Python), Next.js, Gemini, WebRTC VAD

- Built a transcription pipeline with a 3-provider chain (Whisper, Smallest Pulse, Sarvam) and real speaker diarization using VAD + voice embeddings + clustering, not heuristic speaker alternation.
- Architected LLM rubric-based QA scoring using schema-validated structured outputs, plus sentiment analysis, keyword topic tracking, and semantic search via Gemini embeddings + cosine similarity.
- Streamed live transcription job status via Server-Sent Events, with each pipeline stage degrading gracefully when its provider dependency is unavailable.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL, C++

Frameworks & Libraries: Node.js, NestJS, Express.js, FastAPI, SQLAlchemy

AI/ML: LangChain, LangGraph, Gemini Vision AI, RAG, LLM-as-Judge Evaluation, Gemini 2.5 Flash (+ Embeddings), Imagen 3, Veo 3

Infra & Cloud: AWS (S3, SQS, Lambda, EC2, API Gateway), GCP (Cloud Run, Cloud SQL), Cloudflare R2, Docker, Pulumi (IaC)

Databases & Systems: PostgreSQL, MongoDB, DuckDB, System Design (HLD/LLD), DSA, Microservices, Git

EDUCATION

SRM Institute of Science and Technology

Bachelor of Technology in Computer Science and Engineering

9.25/10 CGPA

Sept 2020 – Jun 2024