

SUSHANT SHAMBHARKAR

+91-8149040621 ◊ sam1064max@gmail.com ◊ sushant.ai

About Me

AI Architect and Tech Lead with 10+ years of experience in architecting and scaling production AI platforms across finance, adtech, analytics, and cloud ecosystems. Specialized in GenAI, Agentic AI, LLM infrastructure, RAG, distributed inference, and data-to-AI platform engineering. Proven track record of transforming complex business workflows into intelligent products and autonomous AI capabilities. Technical leader in driving AI strategy, platform architecture, engineering excellence, and enterprise-scale AI adoption while mentoring teams and leading cross-functional initiatives.

Work Experience

- **AngelOne** (Jan'24 - Ongoing)
(Project: *AI Factory* | Department: *AI* | Role: *AI Architect*)
Built an enterprise Agentic AI Platform that standardized agent orchestration, memory, RAG, and tool integration for production AI applications. The platform enabled multiple business solutions including AI-Powered Portfolio Intelligence, Auto Portfolio Rebalance Suggestions, and RM Enablement workflows for HNI clients, improving customer engagement by 35%, reducing advisor response latency by 60%, and cutting RM onboarding time by 50%.
- **Amagi** (May'23 - Jan'24)
(Project: *Audience Intelligence Platform* | Department: *Tech* | Role: *TechLead*)
Architected a distributed AI platform combining low-latency edge inference for audience understanding and ad personalization with centralized NLP and LLM-powered services for content classification, audience insights, and campaign optimization. Leveraging PyTorch, ONNX Runtime, MLflow, Kubernetes, Snowflake, and Delta Lake, improved CTR by 30% and increased viewer retention by 20% across large-scale streaming advertising workloads.
- **Slintel** (Oct'21 - Mar'23)
(Project: *Analytics Copilot* | Department: *Data and ML* | Role: *Lead MLE*)
Pioneered an AI-powered analytics platform that transformed BI users into self-service data consumers through NLP2SQL, dashboard generation, and intelligent query optimization. Leveraging fine-tuned LLMs and automated aggregation strategies, improved dashboard creation efficiency by 200% and reduced dashboard load times by 3x.
- **Google** (Dec'19 - Sep'21)
(Project: *Mendel* | Department: *Google Cloud* | Role: *Sr. MLE*)
Developed an AI-powered software delivery platform that combined CodeBERT-based code intelligence, deployment telemetry, and historical incident analysis to predict release risk, detect deployment conflicts, and recommend canary rollout strategies. Leveraging historical code changes, rollback events, and production failures as training signals, reduced code-induced failures by 25% and minimized hotfix cycles through proactive risk mitigation.
- **Goldman Sachs** (Aug'18 - Dec'19)
(Project: *Email Analytics* | Department: *Client Experience Engineering* | Role: *Sr. MLE*)
Developed an AI-assisted client communication platform leveraging BERT trained on enterprise data and fine-tuned on SME-curated golden datasets. Enabled contextual response recommendations, semantic search, and intelligent prioritization using historical communication patterns, reducing response latency and improving service consistency. Established automated retraining, evaluation, and deployment pipelines for continuous optimization.
- **Persistent Systems** (Jul'13 - Aug'15, Sep'17 - Aug'18)
(Project: *Mobile Workforce Management* | Department: *Location Product* | Role: *MLE*)
Engineered predictive workforce optimization systems for route planning, task duration forecasting, and resource allocation, improving field productivity and reducing operational downtime. Migrated legacy SAS workloads to scalable PySpark ML pipelines, achieving a 10x reduction in processing latency while eliminating licensing costs.

Technical Skills

Python, SQL, Java, LangGraph, CrewAI, MCP, A2A, LangSmith, AgentOps, DeepEval, Ragas, Guardrails, LiteLLM, vLLM, MLflow, Vertex AI, Bedrock, AI Foundry, OpenAI, Claude, Gemini, Semantic Kernel, FastAPI, Kafka, Spark, Flink, Ray, Qdrant, PostgreSQL, Docker, Kubernetes, Terraform, Temporal, Harness, AWS, GCP, Azure, Databricks, Prometheus, Grafana, OpenTelemetry, Presidio, CI/CD.

Education

Master of Technology, Computer Science, IIT Bombay (2015 - 2017)
Bachelor of Engineering, Computer Science, Nagpur University (2009 - 2013)

Achievements, Activities and PORs

- Recognized as a top performer at Google, Goldman Sachs, and Persistent Systems.
- Mentored 30+ junior engineers, conducted 100+ interviews, and led a team of 6 engineers.
- Speaker and mentor on GenAI, Agentic AI, NLP, and distributed AI systems.
- Built and operate a self-hosted AI lab for agentic workflows, private LLM infrastructure, distributed inference, and AI platform experimentation.
- Published open-source templates and reference architectures for enterprise-grade agentic AI systems.
- Scored 99.51 percentile in GATE 2015 among 115425 registered students.

Portfolio

[LinkedIn](#) ◊ [GitHub](#) ◊ [Medium](#) ◊ [Substack](#) ◊ [HackerRank](#) ◊ [AtCoder](#) ◊ [LeetCode](#) ◊ [CodeChef](#)