

Samarth Borade

Los Angeles, CA | (213) 476-2975 | borade@usc.edu | linkedin.com/in/samarthborade | github.com/samarthborade007




EDUCATION

- **University of Southern California** *Los Angeles, CA*
Master of Science in Computer Science *Aug 2025 – Present*
- **SVKM's Mukesh Patel School of Technology Management** *Mumbai, India*
Bachelor of Technology in Computer Engineering | Merit Scholar, Top 5% of Cohort *Aug 2022 – May 2025*



EXPERIENCE

- **Ai-Horizon.io (Nexasprint Technologies Pvt Ltd)** *Mumbai, India*
AI Engineer Intern & Founding Engineer *Jun 2024 – Sep 2024*
 - **Shipped 40+ production AI agents** using **large language model (LLM)** pipelines, **retrieval augmented generation (RAG)**, and **reinforcement learning from human feedback (RLHF)**; drove a **\$2.4M seed valuation in 3 months**.
 - **Architected** the **neural network-backed** agentic framework and **LangChain LLM** layer from scratch with **prompt engineering**, **tool use**, and **vector store** retrieval; the scalable library became the company's **primary product offering**.
 - **Built DataSculptor**, a **machine learning** dataset cleaning pipeline with **PII detection**, **toxicity filtering**, and **embedding-based** deduplication across **20+ datasets / 500K+ records**, cutting preprocessing time by **60%**.
- **Allcargo Logistics Ltd** *Mumbai, India*
Data Science Intern *Dec 2022 – Jun 2023*
 - **Designed Python ETL pipelines** and **Power BI dashboards** for transportation and warehousing data across **5+ hubs**, driving a **10% gain in delivery efficiency** and **15% reduction in operational delays**.
 - **Developed** a scalable 8-module **Simulation Operations Dashboard** with **ML-based demand forecasting** (MAE: 8.2%) and drill-down visualizations, enabling real-time planning and monitoring across distribution hubs.
 - **Modeled a dynamic pricing optimization engine** on **10,000+ transactions** using demand elasticity and market indicators; deployed to production, improving revenue per shipment by **5–8%** with no churn increase.

KEY PROJECTS

- **NeuralOps – Multi-Agent LLM Orchestration Framework**  *Dec 2024 – Mar 2025*
Python, FastAPI, LangGraph, GPT-4o, Claude 3.5, Qdrant, Redis, Celery, Docker, Prometheus
 - **Architected** a stateful **multi-agent system** via **LangGraph** with dynamic tool-routing and persistent **long-term memory** in **Qdrant**; agents decompose tasks into sub-agent calls with shared memory and automatic rollback on failure.
 - **Shipped Celery + Redis** async queue handling **1,000+ concurrent agent runs** and a **Prometheus + Grafana** observability stack tracking token spend, p95 latency, and hallucination flags; full **MLOps** discipline for agentic workflows.
- **DashExAI – Intelligent Business Analytics Platform**  *Jan 2025 – Apr 2025*
Python, Next.js, TypeScript, Scikit-learn, XGBoost, LLMs, LangChain, Pandas, Recharts
 - **Built** a full-stack **machine learning** platform supporting **6+ model families** (ARIMA, Prophet, XGBoost, K-Means, DBSCAN) with automated feature engineering, **model training**, **model evaluation**, and one-click deployment.
 - **Implemented** an automated **exploratory data analysis (EDA)** pipeline with **SHAP-based feature importance** across **20+ chart types**, cutting analyst data prep from hours to **under 5 minutes**.
- **OMNI.ai – AI-Powered Website Generator**  *Nov 2024 – Mar 2025*
React, Node.js, WebAssembly, GPT-4, Claude, Groq, Ollama
 - **Built** a **natural language processing (NLP)** pipeline converting prompts → live websites via **WebAssembly WebContainers** for real-time in-browser execution; delivers **end-to-end prototypes in under 2 minutes** with zero-reload deployment.
 - **Designed** a **model-agnostic LLM routing layer** unifying **GPT-4, Claude, Groq, and Ollama** with per-request provider selection and streaming, enabling A/B testing of generation quality and **inference performance**.

PUBLICATIONS

- **Co-Author – Age of Agentic AI: Building Machines That Think, Decide, & Act**  *Jan 2026*
Amazon, Technical Writing *Agentic AI Systems*
 - Co-authored practitioner guide on **intent-driven agent execution**, **tool-use orchestration**, and multi-agent collaboration patterns for production autonomous systems; targeted at working ML and backend engineers.
- **Research Paper – Automated Classification of Musical Genres**  *Dec 2024*
Springer ICRTCIS 2024, Algorithms for Intelligent Systems Series
 - **Engineered** a **Random Forest** classifier with audio feature engineering (MFCCs, RMS, spectral centroid, zero-crossing rate) achieving **80.12% accuracy** across 10 genres on GTZAN, outperforming CNN-RNN (56.4%) and ANN (64.06%) baselines.

SKILLS

- **Expert:** Python, PyTorch, LangChain, LlamaIndex, RAG Pipelines, Transformer Fine-tuning, Model Training, Model Evaluation, Inference, Scikit-learn, NumPy, Pandas, SQL
- **Proficient:** TensorFlow, Hugging Face, FastAPI, React, Node.js, Docker, MongoDB, REST APIs, WebAssembly, ETL Pipelines, Data Pipelines, Monitoring
- **Familiar:** C++, JavaScript/TypeScript, R, Bash, AWS, Google Cloud, Power BI, CI/CD, A/B Testing, Kubernetes

ACHIEVEMENTS

- **Certification:** NVIDIA Deep Learning Institute – Transformer-Based NLP (attention mechanisms, BERT/GPT pretraining, fine-tuning)
- **Hackathons:** **1st Prize** – DronaHQ Hackathon 2020 (Android app, paperless logistics, 200+ teams); **Finalist** – Cyber Cypher 3.0 2024 (18-hour agentic AI challenge)