ADITYA AGARWAL

New York Metropolitan Area | adiagarwal1509@gmail.com | www.adityaagarwal.me
LinkedIn: @adityaagarwal1999 | GitHub: @adiagarwalrock

TECHNICAL SKILLS & CORE COMPETENCIES

Languages & Frameworks: Python, Django, Flask, Fast-API, Java, Node.js, PyTorch, CUDA, Vue.js
Generative AI & MLOps LLMs, MLOps, Data Science, RAG Pipelines, Agentic Systems, Vector Databases, LangChain, Llama-Index, AutoGen
Security & Compliance: Data Governance, Automated Monitoring, Security Controls, Content Analysis, Threat Detection
Cloud & Infrastructure: AWS, GCP, Docker, Terraform, Docker-Compose
Data & Distributed Systems: MongoDB, MySQL, PostgreSQL, ETL Pipelines, Real-time Processing, Microservices, Redis
Development Practices: CI-CD, Software Engineering, SDLC, Computer Architecture, Git, DevOps, Event-Driven Architecture

WORK EXPERIENCE

Founding Machine Learning Engineer | Soopra.ai, San Francisco **Aug 2024 - Jan 2026**

- Led the **system design** and implementation of **Generative AI** workflows using **Multi-Agent Architecture**, increasing site interactivity
- Established **MLOps** best practices for deploying large-scale **LLM agents**, including automated **evaluation frameworks** to monitor model drift and hallucination rates.
- Built resilient data processing systems creating Agent-first social media platform with high availability and distributed computing principles.
- Applied advanced **Data Science** methodologies to optimize recommendation engines, aligning technical output with executive **decision-making** goals.
- Optimized **inference latency** and token consumption for conversational AI, achieving a 20% reduction in user onboarding time through automated validation scripts.

Software Developer Intern (Full-time) | Soopra.ai, San Francisco **Jun 2023 - May 2024**

- Engineered Hybrid-Search with Keyword and Semantic matching on multi-source ingestion pipeline with recency-biased re-ranking, achieving 95% embedding accuracy.
- Designed event-driven chat systems utilizing domain-driven architecture to dynamically determine conversation flow, increasing user engagement by 25% through intelligent response routing
- Optimized background task processing using **Celery** and Redis for real-time social media ingestion, cutting data loading latency by 15%.

Full-Stack Developer | SA Consultant, Bangalore **Aug 2021 - Jul 2022**

- Managed the full **Software Development Life Cycle (SDLC)** for enterprise web applications, applying **System Design** patterns and **Software Engineering** best practices.
- Implemented billing and monitoring systems with PostgreSQL backend, enhancing project documentation accuracy and providing audit trails for compliance.
- Deployed cloud-native solution on Heroku with RDS integration, ensuring high availability, data security, and seamless scalability for distributed teams.

EDUCATION

Master of Science in Machine Learning **Sep 2022 - May 2024**
Stevens Institute of Technology, NJ
Bachelor of Engineering in Computer Science **Aug 2017 - Jul 2021**
Alliance University, Bangalore

PROJECTS

Atlantis – Web Notes | GitHub **March 2023**

- Developed a self-hosted web platform for Mermaid diagrams and Markdown notes with live preview and export features.
- Built Docker-ready editor enabling interactive diagram creation, local persistence, resulting in 4K+ Docker pulls and strong developer engagement.

Heathcliff - Personal Agent | GitHub **Jan 2025**

- **Architected a voice-activated agentic system** using **LangChain** and **Gemini** that autonomously orchestrates external integrations (Gmail, Calendar, Spotify) to execute complex, multi-step user workflows with real-time latency.
- **Implemented a persistent memory layer** using **ChromaDB** and **Mem0** for Retrieval-Augmented Generation (RAG), enabling the assistant to retain long-term user context and preferences across sessions while minimizing hallucinations.

SOAR - Managed RAG System | Product Page **Jul 2024**

- Built an AI framework using Django, Llama-Index, and Qdrant to turn unstructured documents into a semantic, searchable knowledge graph with low-latency vector search.
- Developed customization tools for live prompt engineering, model swapping (Claude, GPT-4, etc.), and parameter tuning to optimize AI outputs.
- Productized solutions such as embedded widgets & RESTful APIs with automated PDF/CSV reports, Slack alerts, and ticket creation pipelines.
- Deployed on GCP with Docker/Kubernetes, enabling auto-scaling, security controls, and a roadmap for multimodal (image/table) support.
- Implemented ML-based governance controls highlighting malicious spans within content for enhanced attribute and security monitoring

Insight - Enterprise Project Management Platform **Oct 2021 - Jun 2022**

- Led development of Django-based full-stack platform with comprehensive project monitoring, billing estimates, and automated reporting capabilities for enhanced operational efficiency
- Deployed production system on Heroku with AWS RDS, implementing infrastructure-as-code principles for reliable data management and security compliance

Fantastic Computing Machine - ML Platform **Dec 2020 - Jun 2021**

- SaaS platform on **Flask framework** for dynamic ML model deployment, enabling secure model sharing and collaborative development.

CERTIFICATIONS & PUBLICATIONS

- Neural Networks and Deep Learning - Coursera (deeplearning.ai) | April 2020
- Published Research: "*Text Mining Approach Based on TF-IDF and SVM for Text Classification*" in *Latest Innovation for Future Education 2021*
  ISBN: 978-81-947019-0-3 | January 2021