

Govind Thakur

📍 Phoenix, AZ | 📞 213-643-9098 | ✉ govind.thakur11@gmail.com | 🌐 govindthakur.com | 🌐 govind-thakur2 | 🐙 GThakur26

AI Systems Engineer building **production LLM and multimodal systems** at enterprise scale. I design evaluation and data frameworks, deliver AI APIs serving **100M+ users**, and drive release decisions through measurable reliability, latency, and quality outcomes.

Education

Master of Science, Computer Science–Artificial Intelligence, University of Southern California Aug 2022–May 2024
Bachelor of Engineering, Computer Engineering, Dwarkadas J Sanghvi College of Engineering Jul 2018–Jun 2022

Technical Skills

- **Core Competencies:** Agentic AI, AI Engineering, ML Systems, LLMs, Semantic Search, Generative AI, RAG, Evaluation, Deep Learning
- **Stack:** Python, Java, C++, SQL, React, FastAPI, Flask, PyTorch, TensorFlow, Docker, AWS, GitHub Actions, Kubernetes, CI/CD

Experience

AIML Engineer–II, GenAI Search Team Sept 2024–Present
IntraEdge – Client Partner American Express, Phoenix, AZ

- Led delivery of American Express’ **first production Agentic AI voice bot** for fraud charge verification, driving agent design, prompt strategy, and cross-functional execution.
- Owned **security, risk, and governance approvals** (threat modeling, model risk, SaaS compliance), unblocking launch in a regulated financial environment.
- Designed and optimized production GenAI Search APIs (Python, Java) serving 100M+ users, improving retrieval and latency through pipeline parallelization.
- Built **AGX** (React + Python), an evaluation platform to simulate production behavior, measure ranking quality, and catch regressions pre-release.
- Architected automated validation and monitoring pipelines and led LLM red-teaming, reducing regression defects by **35%**.
- Defined large-scale search quality metrics (**NDCG@k, precision/recall, containment, latency**) and translated experimentation into product and executive decisions.

Machine Learning Engineer Jun 2023–Sept 2024
Northern Lights, Los Angeles, CA

- Led architecture and implementation of a multimodal analytics MVP with distributed ingestion pipelines and production backend integrations.
- Built containerized ML workflows on AWS SageMaker and Docker for automated training, evaluation, and CI/CD delivery.
- Designed Airflow pipelines processing **millions of records** and deployed LLM systems to extract structured insights from enterprise data.

Software Engineering Intern Jun 2021–Aug 2021
JP Morgan Chase & Co, Mumbai

- Improved NLP accuracy of JPMC’s E-Trading Assistant Bot by 20% through Rasa model tuning and advanced intent classification.
- Refactored Athena bot architecture with modular microservices, improving performance and maintainability.

Patents & Projects

System and Method for Estimating Intrinsic Popularity of Content — Patent Pending (US Serial No. 63/642,980)
Speech to Code: Voice-Driven Developer Environment Dec 2020–Jun 2022

- Built a PyTorch-based voice-driven coding interface for visually impaired developers with autonomous error correction support.

Research & Publications Google Scholar

Thakur, G. (Under Review, 2025). “Emergent Ethics in Agentic Simulations: Moral Incubation and Conscious Development across GPT, Claude, Gemini, and XAI.” *Springer Nature – AI and Ethics Journal*.

Pathare, A., Mangrulkar, R., Thakur, G., et al. (2023). “Comparison of Tabular Synthetic Data Generation Techniques using Propensity and Cluster Log Metric.” *Elsevier – International Journal of Information Management Data Insights*.

Certifications

IBM AIML Specialization | Udacity ML Nanodegrees (3) (Credential)

Honors & Extracurriculars

- Course Producer for USC CSCI 401 & 102; mentored 50+ students in ML and GenAI projects.