Gaurav Prakash

+1 (510)-980-1988 | gauravprakash2104@gmail.com | gaurav-prakash | Gauravp2104

Education

University of Maryland — College Park, MD
*Master's in **Data Science*** — *Sep 2025 – Present*

National Institute of Technology Karnataka — Surathkal, India
*Bachelor of Technology in **Information Technology*** — *Dec 2021 – May 2025*

The Indian School — Bahrain
*Higher Secondary XII CBSE — **96.2%*** — *Apr 2019 – June 2021*

Experience

Analytics Intern | *GlobalFoundries* — Apr 2025 – Aug 2025

- Built a **RAG-based chatbot** integrated with GPT-4o to automate employee Q&A across 50+ policies and workflows.
- Developed an **ML-driven candidate success model** using NER and cosine similarity to compute fit scores for job applications.
- Implemented data pipelines using **Python (pandas, SQLAlchemy)** for structured and unstructured data ingestion.
- Collaborated with the People Analytics team for ROI modeling and predictive insights on new hire retention.

Data and AI Intern | *IBM India* — May 2024 – July 2024

- Developed a **sales optimization RAG system** for Parle using LangChain and HuggingFace embeddings.
- Performed **geospatial data mining** to identify retail clusters and forecast sales using regression and clustering.
- Built a **FastAPI backend** for LLM queries integrated with Twilio WhatsApp API for real-time communication.
- Explored text summarization and entity extraction from 10K+ retailer PDFs using NLP pipelines (spaCy, NLTK).

Projects

Refining LLMs with RL for Human-like Text Generation | *Research Paper* — Jan 2024 – Apr 2024

- Used **Proximal Policy Optimization (PPO)** to fine-tune FLAN-T5 for dialogue summarization, achieving 28% reduced AI detectability.
- Implemented **LoRA** for parameter-efficient fine-tuning on limited compute; optimized inference throughput by 35%.
- Explored RLHF-style reward modeling for coherence, aligning with Experian's focus on generative AI research.

Twitter Bot Detection via GraphSAINT — Jan 2025

- Applied **Graph Neural Networks** (GNNs) for classifying fake accounts using heterogeneous edge relations.
- Implemented edge-type normalization and achieved 87.38% F1-score on Twibot-20 dataset.

Driver Drowsiness Detection System | *TensorFlow, OpenCV* — Jan 2024 - Feb 2024

- Designed a robust real-time drowsiness detection system that continuously monitors drivers' eye states to detect fatigue by identifying patterns of prolonged eye closures.
- Integrated OpenCV to capture real-time webcam video, analyzed frames to classify eye states, and implemented a scoring system to trigger an alarm for prolonged eye closures, enhancing driver safety.

Technical Skills

Languages: Python, C++, R, SQL, JavaScript
Frameworks: PyTorch, TensorFlow, Scikit-learn, LangChain, Django, FastAPI
Tools: Qiskit, HuggingFace, Matplotlib, Pandas, NumPy, Docker, Git
Domains: Data Mining, Predictive Modeling, Quantum Computing, Generative AI, NLP, Regression, Ensemble Learning