KAMAL UTLA

+91 8179298095 • kamalutla@gmail.com • https://www.linkedin.com/in/kamal-utla/ •

EDUCATION

B.Tech. Engineering Physics Graduating May 2024
Indian Institute of Technology, Roorkee 8.415 / 10

RESEARCH & TECHNICAL SKILLS

Programming & ML: Python, SQL, PyTorch, NumPy, Pandas
LLMs & Representation Learning: HuggingFace, vLLM, LangChain, OpenAI, Anthropic, Gemini, ChromaDB
Systems & Infrastructure: Docker, GitHub, FastAPI, Flask, PostgreSQL, WebRTC, Google Cloud Platform

RESEARCH & INDUSTRY EXPERIENCE

BorderPlus, Bangalore: AI Engineer Nov 2025 — Present

- **Knowledge Engineering & Semantic Ingestion**: Architected a centralized knowledge ingestion and semantic indexing platform for large healthcare corpora with document-specific preprocessing, chunking strategies, and embedding pipelines. Built LLM extraction pipelines that convert long regulatory documents (e.g., clinical standards and policy texts) into structured, taxonomy-tagged rules and knowledge.
- **Agentic Workflows & Information Extraction**: Developed multi-step agentic pipelines orchestrating LLM reasoning across tasks such as document classification, rule abstraction, and spoken clinical transcript processing. Systems detect entities (e.g., patients), extract tasks from care plans, and generate structured notes for downstream healthcare systems.
- **Retrieval-Augmented Reasoning Systems**: Built multilingual RAG architectures combining dense and sparse retrieval, query reformulation, intent classification, and user-memory integration. Enables grounded question answering over healthcare literature with inline citations and bibliography generation.
- **LLM Evaluation, Alignment & Data Infrastructure**: Designed automated compliance evaluation pipelines that score clinical reports against extracted regulatory rules. Improved production conversational agents through prompt orchestration, guardrails, and token optimization (15% reduction), and built tools for voice/chat data collection and prompt A/B testing.

Convai Technologies Inc., Bangalore: Data Scientist — Applied Research Team June 2024 — Oct 2024

- **LLM Infrastructure**: Deployed and served large language models using vLLM on custom cloud infrastructure; built unified backends supporting OpenAI, Anthropic, Gemini, and LLaMA-family models with function/tool calling and safety guardrails.
- **Retrieval & Representation Learning**: Proposed and implemented a hybrid **semantic similarity + BM25 + LLM** chunking and retrieval method, reducing end-to-end latency by **27%** and memory usage by **3–4×** in RAG pipelines.
- **Model Fine-Tuning**: Fine-tuned Llama-3-70B and Llama-3.2-90B-Vision models using curated and synthetic datasets to improve controllability, factuality, and response consistency.
- **Systems Prototyping**: Built research prototypes including multi-agent conversational systems, browser-based autonomous agents via gRPC, real-time WebRTC plugins enabling contextual QA over live websites, and multimodal avatar-matching pipelines using weighted embedding fusion.
- **Evaluation Methodology**: Developed automated test pipelines using LLM-as-judge frameworks to generate, score, and regress prompt-template performance across large-scale test sets.

INTERNSHIP EXPERIENCE

Convai Technologies Inc., Bangalore: Machine Learning Intern Feb 2024 — Apr 2024

- Designed a probabilistic framework for dynamic personality modeling in LLMs using Big Five personality traits.
- Improved long-horizon behavioral consistency using few-shot and constraint-based prompting.
- Built synthetic data generation pipelines to fine-tune smaller LLMs for emotion and context modeling.
- Developed evaluation metrics and testing frameworks for conversational agent performance.

McMaster University, Hamilton, Ontario, Canada: MITACS Globalink Research Intern May 2023 — July 2023

- Developed multi-objective optimization models for electric vehicle recycling center placement, jointly minimizing financial cost and environmental impact using non-linear optimization techniques in Python.
- Conducted a comprehensive synthesis of recycling supply chain literature to inform sustainable infrastructure planning.

ACADEMIC PROJECTS

Using Novel Deep Learning Techniques for Modelling Batteries to Predict the State of Charge of Lithium-Ion Batteries in Electric Vehicles Aug 2023 - May 2024

Created a Novel battery dataset to meet the Indian Environmental Conditions and Presented a poster in the IMESD Conference

- Developed and validated a novel deep learning architecture for SoC prediction, achieving 99.6% accuracy through innovative feature engineering and signal processing techniques.
- Designed and executed experimental protocols for generating India-specific battery performance data using advanced laboratory instrumentation.
- Implemented post-processing algorithms incorporating low-pass filters to optimize prediction stability and reduce noise in real-time applications.

Stylising Images using Deep learning and Style transfer November 2023

Achieved decent style transfer on a custom dataset and low performance computer.

- Implemented and evaluated VGG-19 based style transfer architectures, comparing performance between custom-trained and pretrained ImageNet weights.
- Developed a curated dataset of artistic styles, implementing preprocessing pipelines to optimize training efficiency.
- Conducted comparative analysis of style transfer quality metrics across different model configurations and training approaches.

Using Siamese Networks for Fake Signature detection April 2023

- Architected a Siamese CNN implementation using TensorFlow, optimizing network parameters for signature verification tasks.
- Developed a training pipeline incorporating contrastive loss functions, achieving 81% accuracy on the ICDAR 2011 dataset.
- Conducted systematic evaluation of model performance across various signature types and forgery scenarios.

Monte-Carlo Modelling of Light Transport Aug 2022 - Nov 2022

- Developed a physics-based Monte Carlo simulation framework for modeling photon transport in human epidermal tissue.
- Implemented non-uniform probability distributions based on biological parameters to accurately model photon propagation patterns.
- Conducted validation studies comparing simulation results with published literature, achieving strong correlation with experimental data.

Exploring Classification Models to Predict Students' Expenditure Problems March 2023

- Designed and implemented a comprehensive survey methodology to collect financial behavior data from IIT Roorkee students.
- Developed feature engineering pipelines to transform survey responses into meaningful predictive variables.
- Conducted comparative analysis of various classification algorithms including SVM with multiple kernel configurations and logistic regression models.

HONORS & AWARDS

- MITACS Globalink Research Fellow (2023)
- Winner — Qiskit Fall Fest 2021, IIT Roorkee Quantum Hackathon (Designed a Quantum Music Composer)
- IIT Roorkee Encore Award (2023) for all-round excellence