Ganesh Epili

Master's Student || NLP Research Scholar

✉ ganeshepili1998@gmail.com 📞 8984445951 📍 Bengaluru, India ⬡ /gnsepili in /gnsepili ⬦ Dev Portfolio

PROFESSIONAL EXPERIENCE

NLP Researcher   05/2022 – present | Gandhinagar, India
IRLP Lab, DAIICT
- Conducted research on information retrieval systems and natural language processing (**NLP**), as well as generative AI models like **LLM**s specializing in the domain of software code, to enhance the understanding and processing of code-related queries and information.
- Served as a technical organizing member for the FIRE conference organized by IRLP Lab, contributing to the development of a platform for conference participants and managing the overall event operations.

Backend/ML Intern   07/2019 – 08/2020 | Chandigarh, India
Jungleworks / Jugnoo [Awarded as best employee] ⧉
- Created a platform for building custom chatbots on client's data. Uses **fastify/ Nodejs** for the Backend and **RASA** for the NLP server.
- Developed Interactive dashboards for e-commerce websites (clients) using the KATO (in-house) platform and optimized the codebase to reduce the graph loading time by **70%.**

EDUCATION

M.Tech   08/2022 – present | Gandhinagar, India
Dhirubhai Ambani Institute of Information and Communication Technology
- Specialized in Machine Learning (ML) and enrolled in fundamental and advanced ML and Deep Learning (DL) courses.
- Engaged in applied research at IRLP Lab, contributing to real-world projects that bridge theoretical concepts with practical applications.

B.Tech   07/2016 – 07/2020 | Phagwada, India
Lovely Professional University
- Explored fundamentals of computer science and applied mathematics and gained substantial exposure to software development.

PROJECTS

RepoRAG: Enhancing LLMs for Repository-Level Code Generation ⧉   01/2024 – present
(Paper under review)
- Developed/Researched a RAG-based solution to overcome the context length limitations of LLMs, enabling efficient code generation and debugging at the repository level by leveraging metadata, hierarchical structures, and framework documentation.

IndicTranse: Speech-to-Speech Translation for Indic Languages   03/2023 – present
- Developed a voice-to-voice translation platform for Indian languages using a fine-tuned LLM, enabling real-time voice translation with integrated **ASR** (Automatic Speech Recognition) and **TTS** (Text-To-Speech).
- For under-the-hood Indic language translation, I've fine-tuned the **phi-2 2.7B** model on the **AI4Bharat** dataset.

NaturalSQL: Enabling Natural Language Queries for SQL Databases ⧉   12/2022
- Developed a model based on the Transformers architecture, trained from scratch on the WikiSQL dataset, to convert natural language queries into SQL queries, allowing users to interact with SQL databases with natural language.

IntellectRAG: Enriched LLM Q&A with Document Insights   11/2023 – 12/2023
- Developed a RAG-enhanced LLM system using **Mistral 7B**, enabling advanced Q&A capabilities with **documents**, **webpages**, and **YouTube** video transcripts for comprehensive information retrieval.

Leaderboard for FIRE Conference   08/2023 – 10/2023
- Developed the backend server for a leaderboard platform at the MTIL conference, utilizing **FastAPI** and **MySQL**. This enabled the evaluation of participant models and the generation of leaderboards.
- The platform offers features such as team collaboration, managing numerous leaderboards across various domains, and real-time results updates.

DetectGPT: AI-Generated Text Detection ⧉   02/2024 – 03/2024
- Developed a system to distinguish between AI-generated and human-written text using a zero-shot approach based on probability curvature, implementing the methodology outlined in the "**DetectGPT** ⧉ " research paper.

SKILLS & TECHNOLOGIES

Languages:
Python, Javascript, C, C++

Frameworks & Libraries:
scikit-learn, PyTorch, LangChain, TensorFlow, NodeJS, FastAPI, Apache Spark

Tools & Technologies:
Spacy, NLTK, AWS, Git, SQL, Docker