

# Dilip Venkatesh

Email: dipsivenkatesh@gmail.com  
Linkedin

Github  
Google Scholar

EDUCATION	<b>Birla Institute of Technology and Science, Pilani</b> B.E., Computer Science and Engineering, 2024	GPA: 7.57
PROJECTS	<p><b>LLM for Indic languages:</b> Developed a custom decoder-based large language model for Indic languages, including the creation and implementation of a specialized SentencePiece tokenizer</p> <p><b>GPT-4 Multimodal for medical image analysis:</b> Utilized GPT-4V multimodal model to classify the severity of diabetic retinopathy (0-4) and generated detailed diagnostic reports</p> <p><b>Fine-tuning LLMs for sentiment analysis:</b> Fine-tuned IndicBERT v2 model for sentiment analysis in Tamil using Hugging Face, achieving an F1 score of 0.85 on the test set.</p> <p><b>Machine Learning for Probability of success of a startup:</b> Employed Bayesian networks to analyze and predict the probability of startup success, enhancing decision-making processes and investment strategies, and identify critical aspects for the success of a startup.</p> <p><b>Database systems - Recursive DNS resolver:</b> Engineered a recursive DNS resolver comprising a root server, TLD, and ANS to retrieve IP addresses from domain names, integrating SQL, Python, and HTML/Bootstrap for robust functionality.</p> <p><b>Machine learning in medical research:</b> Implemented machine learning and deep learning techniques for the analysis and diagnosis of diabetic retinopathy, utilizing the Relief algorithm for feature selection on the DR dataset. Employed U-Nets for semantic segmentation of retinal images, enhancing the accuracy and precision of medical image analysis. Developed the project using PyTorch and Julia (Flux framework),</p>	
PUBLICATIONS	<p><b>Cross-Lingual Auto Evaluation for Assessing Multilingual LLMs:</b> Developed the Cross Lingual Auto Evaluation (CIA) Suite, an innovative framework for multilingual NLP evaluation, including Hercule (a cross-lingual evaluator LLM) and Recon (a novel test set with 500 human-annotated instructions across 6 languages). Submitted to NAACL 2025.</p> <p><b>BITS Pilani at SemEval-2024 Task 10: Fine-tuning BERT and Llama 2 for Emotion Recognition in Conversation:</b> Fine-tuned LLaMA2 and BERT models on a Hindi language dataset for sentiment analysis using PyTorch, Hugging Face's Transformers, Datasets, and Evaluate libraries; achieved a maximum weighted F1 score of 0.42. Published in SemEval at NAACL 2024.</p> <p><b>BITS Pilani at SemEval-2024 Task 9: Prompt Engineering with GPT-4 for Solving Brainteasers:</b> Leveraged OpenAI's GPT-4 Turbo API for prompt engineering, addressing brain teasers to test the reasonability of language models. Published in SemEval at NAACL 2024.</p> <p><b>BITS Pilani at SemEval-2024 Task 1: Using text-embedding-3-large and LaBSE embeddings for Semantic Textual Relatedness:</b> Obtained sentence embeddings using sentence-transformers/LaBSE and OpenAI text-embedding-large, for identifying semantic textual relatedness analysis across nine regional languages globally. Published in SemEval at NAACL 2024.</p>	

## SKILLS

**Technical Proficiency** : Python, PyTorch, HuggingFace, Julia, C++, Java, Flux.jl, SQL, Git/Github, Pandas, numpy, scikit-learn, Git/Github, L<sup>A</sup>T<sub>E</sub>X.

**Subjects / Electives**: Natural Language Processing, Operating Systems, Design and Analysis of Algorithms, Data Structures and Algorithms, Database Systems

## EXPERIENCE

### Project Associate

**AI4Bharat**

August 2024 - Present

IIT Madras

- Working on creating a cross lingual evaluator LLM for multilingual NLP under Prof. Anoop Kunchukuttan and Prof. Raj Dabre

### Software Development Intern

**BlokTrek Technologies Pvt Ltd**

July 2023 - December 2023

Bangalore

- Developed and maintained backend REST APIs using Python, ensuring robust and scalable integration for core company operations.

- Engineered high-quality, production-level Python code for critical functionalities, including pricing, payments, subscription management, and tear sheets, contributing to the company's operational efficiency and service reliability.

- Implemented a Named Entity Recognition (NER) system utilizing Azure OpenAI NER API and spaCy, enhancing the company's data processing and analytics capabilities.

- Crafted and managed backend Python code for CRUD operations on MongoDB, streamlining database interactions and data management processes.

- Leveraged AWS technologies, including Lambda, S3, DocumentDB, and EC2, to build, deploy, and manage scalable, secure, and cost-efficient applications, driving the company's cloud-based solutions and infrastructure.

### Intern

**APS Lifetech**

May 2022 - July 2022

Remote

Worked in the web development team, utilizing Python for backend development and HTML/Bootstrap for frontend design, contributing to the creation of responsive and functional web applications.