ABHISHEK JAIN

8764724760 abhishekjain18138@gmail.com abhishek-jain-594j abhishek4922

Education

Indian Institute Of Information Technology, Dharwad **Nov. 2022 – Present**
Bachelor of Technology in Data Science and Artificial Intelligence *Dharwad, Karnataka*

Experience

Vocab-ai **Jan 2025 - Apr 2025**
Project Intern *Remote*

- Built a multilingual conversational AI pipeline using AI4Bharat ASR, Mistral LLM, and Indic TTS, enabling seamless end-to-end voice-based interactions.
- Fine-tuned Wav2Vec2.0 on 50+ hours of noisy telephonic data, reducing Word Error Rate (WER) to 20% and improving robustness under real-world audio conditions.
- Engineered a RAG-based retrieval system to ground LLM responses in domain-specific documents, enhancing accuracy and contextual relevance in generated outputs.

Skills

- **Programming:** Python
- **Machine Learning:** Scikit-learn, TensorFlow, Keras, PyTorch
- **Data Analysis:** Pandas, NumPy, Matplotlib, Seaborn
- **Big Data Tools:** Hadoop, PySpark
- **Database:** MySQL, MongoDB
- **Visualization:** Tableau, Power BI
- **Generative AI:** LangChain, LlamaIndex, HuggingFace
- **Cloud:** AWS, GCP

Projects

MediBot-AI | Context-aware medical chatbot powered by RAG

- Built a medical chatbot using a **RAG** pipeline with **LangChain**, **Groq's LLaMA**, and **Hugging Face models**.
- Integrated **Flask** and **Pinecone** for real-time vector-based retrieval.
- Automated PDF preprocessing and chunking for semantic embedding.

Vision to Text: Image Captioning System from Scratch | TensorFlow, Keras

- Built an image captioning model using **CNN** for feature extraction and **LSTM** for generating descriptions.
- Preprocessed and structured the **Flickr8k dataset** for efficient model training and evaluation.
- Developed an end-to-end captioning system with saved model and **tokenizer** for real-time inference.

Image Feature Extraction and Analysis | Amazon ML Challenge 2024

- Extracted product text using **PaddleOCR** from image datasets, identifying key attributes using regex.
- Standardized data by mapping and normalizing measurement units across diverse inputs.

Intelligent Brain Tumor Detection System | Python, Keras

- Built a CNN model to classify brain tumors using MRI images, enhancing accuracy via preprocessing.
- Implemented using **Conv2D**, **Dense layers**, and optimized with Scikit-learn.
- Visualized model predictions and performance using **Matplotlib**.

Achievements & Certifications

- **Top 17 Semi-Finalist**, SBI Hack-AI-Thon 2025
- **Rank 138**, Amazon ML Challenge 2024
- **Google Cloud: Foundations and Generative AI Certification**