Balu Harshavardan Koduru

+1 (716) 907-9338 | github.com/BaluHarshavardan99 | balu.koduru99@gmail.com | linkedin.com/in/balu-koduru/

Experience

Machine Learning Researcher, SUNY RF *Sep 2022 - May 2024*

Tools: Python, NumPy, PyTorch, Pandas, Scikit-Learn, Spacy, NLTK, GPT-4 API, Whisper API, OpenAI CLIP, HuggingFace

- Pioneered the first ever image-caption dataset in endoscopy and engineered a contrastive learning-based Visual Question Answering (VQA) model, **EndoAssistant** for endoscopic surgery analysis. (***Submitted: NeurIPS 2024***)
- Fine-tuned **LLaVA (LLM)** to create Endoscopy AI assistant that answers questionnaires on endoscopy surgical procedures.
- Utilized **OpenAI's CLIP** (multimodal) model to generate descriptive captions for endoscopic medical imagery.
- Implemented Automatic Speech Recognition (ASR) with **Whisper API** to generate captions, employed Spacy and **ChatGPT (GPT-4)** API for text correction, reducing errors from 40% to 7%.
- Led an interdisciplinary team of data scientists and doctors to develop a **Speech Therapy AI Agent**, integrating clinician expertise with deep learning to assist Speech-Language Pathologists.
- Developed an NLP pipeline using transformer models and fine-tuned **large language models (LLMs)** to extract structured data from patient health records.

Machine Learning Researcher, ACPS Group, Norway *Jan 2021 - Jun 2022*

Tools: Python, MATLAB, TensorFlow, Keras, PyTorch, Matplotlib, Git, AirSim, Scikit-Learn, Linux

- Implemented Multi-Agent Reinforcement Learning algorithm to control drone fleets in AirSim simulator using Unreal Engine.
- Developed a CNN-based deep learning model in **PyTorch** to classify UAVs based on Radio Signals, achieving 99.09% classification accuracy. Integrated a custom MATLAB signal processing pipeline reducing processing time by 30%. ***(IEEE))***
- Developed CNN algorithm in TensorFlow to analyze human speech signals and classified background environment achieving 95.2% accuracy. ***(Publication: JASA)***
- Fine-tuned TensorFlow machine learning models for seamless deployment on edge devices (Nvidia Jetson), meeting tight latency requirements.

Data Analyst, JSW Energy *Jun 2019 - Dec 2019*

Tools: Python, MySQL, Pandas, NumPy, Scikit-Learn, Seaborn, MS Excel, Tableau

- Engineered a deep learning model in TensorFlow to manage and predict Un-Burnt Carbon (UBC) levels in 240 MW thermal power plant combustion chambers and provided **data-driven recommendations** enhancing boiler efficiency by 15%.
- Leveraged Exploratory Data Analysis (EDA) to identify plant parameters influencing UBC, and developed interactive dashboards in Tableau to visualize key performance indicators and track UBC levels and plant efficiency metrics.
- Executed ETL pipeline, merging sensor data with plant parameters, facilitating seamless data integration, & standardization.

Relevant Skills and Competencies

- **Programming Languages:** Python, C++, SQL, PostgreSQL, MATLAB.
- **Libraries/Modules:** PyTorch, TensorFlow, Keras, OpenCV, Scikit-learn, Pandas, Hugging Face, NLTK, Spacy, LangChain.
- **Machine Learning Skills:** MLOps, Generative AI, Natural Language Processing (NLP), Large Language Models (LLM), Prompt Engineering, Computer Vision, Graph Neural Networks, Reinforcement Learning, Exploratory Data Analysis (EDA), Data Visualization, ETL.
- **Tools/Cloud:** Git, Tableau, Llama, ChatGPT API, PySpark, AWS, MS Office Suite, Anaconda, ROS, AirSim, Linux.

Projects

Prompt Engineering: Hallucination mitigation in Chatbots

- Enhanced LLM accuracy by implementing prompt structuring and iterative refinement; decreased hallucinations by 40%, leading to a 25% increase in factual, verifiable information output.
- Formulated transformer model built on **RoBERTa** for hallucination detection in chat-bot responses achieving **87.4%** accuracy.

Film Verse: Movie Database and Query System

- Led the design and development of a comprehensive movie information system using Python and SQL. Managed and maintained a large database of over 10,000 movie records using MySQL and PostgreSQL
- Performed database performance analysis and optimization, achieving a 60% improvement through strategic indexing and query tuning, resulting in faster query execution and an improved user experience

Education

University at Buffalo, The State University of New York *May 2024*
Master of Science (MS) in Artificial Intelligence

Birla Institute of Technology and Sciences Pilani, India (BITS Pilani) *June 2022*
Master of Science (M.Sc.) in Physics
Bachelor of Engineering (B.E.) in Electronics and Instrumentation

Publications (ORCID ID: 0000-0002-2502-723X)

- Classification of UAVs Using Time-Frequency Analysis of Remote Control Signals and CNN - ***IEEE iSES 2022***
- Light Weight Deep CNN for Background Sound Classification in Speech Signals - ***JASA***