Girish

girish.research.pr@gmail.com +91-9588703587 Website girish-b794092a1 gir-ish

Education

UPES **B.Tech**, Computer Science (AI-ML Hons.) — Aug 2022 – May 2026

- GPA: 7.5/10.0

Experience

IIIT-Delhi, Under-graduate Research Associate (Hybrid) — New Delhi, India, June 2024 – Present

- Researched deepfake detection, speech emotion recognition, multimodal AI, and Audio Language Models (ALM) to advance speech and audio technology.
- Working with experts to improve machine learning (ML) and deep learning (DL) models for real-world applications.

Reliance Jio, ML Engineer (Remote) — Bengaluru, India, May 2024 – June 2024

- Evaluated Speech Foundation Models (SFMs) for stress recognition, showing that multilingual SFMs perform better than raw data models in understanding physiological signals.
- Worked with interdisciplinary teams to test AI models in different domains, proving that advanced speech models can be used beyond speech tasks for better generalization.

ARTVIEWINGS LLC, AI Research Engineer (Remote) — San Jose, California, July 2024 – Aug 2024

- Developed a scalable AI system for multilingual audio-visual question answering.
- Integrated the MERA framework with ensemble techniques and multilingual data to enhance performance across eight languages.

Suratec Co., LTD, Computer Vision Intern (Remote) — Bangkok, Thiland, June 2024 – July 2024

- Built an application to detect and analyze golf swing phases using video processing and machine learning.
- Designed a real-time feedback interface and optimized the system for different video formats to enhance accuracy and user experience.

Ulster University, Under-graduate Research Associate (Remote) — London Derry, UK, Dec 2023 – May 2024

- Conducted NLP and speech analysis research for autism detection using code-switched speech.
- Improved algorithms, analyzed multilingual datasets, and enhanced model performance for early diagnosis.

Projects

Helix: Versatile AI Assistant GitHub

- AI-Powered Smart Responses: Uses OpenAI GPT, LangChain, and Pinecone for real-time, context-aware interactions, improving customer engagement and automation.
- Versatile & Scalable: Easily deploys across industries like healthcare, retail, and finance with minimal code changes, ensuring seamless adaptability.
- Omnichannel & Fast: Connects via WhatsApp (Twilio) and phone calls, with a FastAPI backend and async MongoDB for quick, efficient responses.
- Tools Used: LangChain, FastAPI, MongoDB, Twilio.

TwinVerify: Secure Encryption with Two-Factor Audio and Text Authentication Framework GitHub

- Created a unique audio encryption system to securely store answers in files, allowing access only for verified users.
- Developed a two-step authentication process using voice verification and text matching for safe encryption and decryption.
- Built a secure platform with random challenge questions, encrypted data storage, and easy file management.
- Tools Used: Speech-to-Text, MongoDB, Natural Language Processing (NLP), Flask, C++

Multimodal Personality Prediction Using Contrastive Learning GitHub

- Built a neural network with two processing paths that uses contrastive learning to predict personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) from audio and video data.
- Trained the model on both speech and visual cues, improving how it understands and learns personality traits from different sources.
- Optimized the system for better integration of multimodal data, ensuring accurate and reliable personality predictions from real-world audio and video inputs.
- Tools Used: TensorFlow, Hugging Face Transformers.

Golf Phase Detection and Analysis Application. GitHub

- Developed algorithms to detect and classify golf swing phases (setup, backswing, downswing, ball impact, and follow-through) with high accuracy.
- Designed a user-friendly interface for players and coaches, enabling easy review and real-time analysis across different video formats.
- Tools Used: OpenCV, MediaPipe, FFmpeg, Streamlit

Technical Skills

Frameworks and Libraries: TensorFlow, Keras, PyTorch, OpenCV, NLTK, SpaCy, LangChain

Programming Languages: Python, C, Java, SQL

Web Technologies: HTML, JavaScript, FastAPI, Flask

Methodologies: Retrieval-Augmented Generation (RAG)

Technologies: Git, Linux, AWS

Publication & Pre-print

Strong Alone, Stronger Together: Synergizing Modality-Binding Foundation Models with Optimal Transport for Non- Verbal Emotion Recognition

Orchid Chetia Phukan, Mohd Mujtaba Akhtar*, ***Girish****, Swarup Ranjan Behera, Sishir Kalita, Arun Balaji Buduru, Rajesh Sharma, SR Mahadeva Prasanna

ICASSP 2025

NeuRO: An Application for Code-Switched Autism Detection in Children

Mohd Mujtaba Akhtar*, ***Girish****, Orchid Chetia Phukan*, Muskaan Singh*

INTERSPEECH Show & Tell 2024

SOURCE TRACING OF SYNTHETIC SPEECH SYSTEMS THROUGH PARALINGUISTIC PRE-TRAINED REPRESENTATIONS

Girish*, Mohd Mujtaba Akhtar*, Orchid Chetia Phukan*, Drishti Singh, Swarup Ranjan Behera, Pailla Balakrishna Reddy, Arun Balaji Buduru, Rajesh Sharma

Submitted

ARE MULTIMODAL FOUNDATION MODELS ALL THAT IS NEEDED FOR EMOFAKE DETECTION?

Mohd Mujtaba Akhtar*, ***Girish****, Orchid Chetia Phukan*, Swarup Ranjan Behera, Jaya Sai Kiran Patibandla, Pailla Balakrishna Reddy, Arun Balaji Buduru, Rajesh Sharma

Submitted

ARE MAMBA-BASED AUDIO FOUNDATION MODELS THE BEST FIT FOR NON-VERBAL EMOTION RECOGNITION?

Mohd Mujtaba Akhtar*, Orchid Chetia Phukan*, ***Girish****, Swarup Ranjan Behera, Sanjib, Arun Balaji Buduru, Rajesh Sharma

Submitted

SYNERGIZING NEURAL AUDIO CODEC AND SPECTRAL REPRESENTATIONS FOR DEPRESSION DETECTION

Mohd Mujtaba Akhtar*, ***Girish****, Orchid Chetia Phukan*, Swarup Ranjan Behera, Pailla Balakrishna Reddy, Arun Balaji Buduru, Rajesh Sharma

Submitted

Beyond Speech and More: Investigating the Emergent Ability of Speech Foundation Models for Classifying Physiological Time-Series Signals

Orchid Chetia Phukan*, Swarup Ranjan Behera*, ***Girish****, Mohd Mujtaba Akhtar*, Arun Balaji Buduru, Rajesh Sharma

Submitted

Representation Loss Minimization with Randomized Selection Strategy for Efficient Environmental Fake Audio Detection

Orchid Chetia Phukan*, ***Girish****, Mohd Mujtaba Akhtar*, Swarup Ranjan Behera*, Nitin Choudhury, Arun Balaji Buduru, Rajesh Sharma, SR Mahadeva Prasanna

Submitted

Volunteering

ISCA Team Volunteer:

- Managing script development and co-host episodes on speech technology and AI advancements.