Nirmalya Mallick Thakur

+91 77973 95326 | nirmalya23@iiserb.ac.in | Website | Github

Education

Indian Institute of Science Education and Research, Bhopal 2023 – Present
Bachelor of Science in Computer Science *GPA: 8.6/10*

Technical Skills and Interests

Programming Languages: Python, C, C++, Java Script
Areas of Interest: World Models, Video Diffusion, Graphics Programming, Simulation, Neural Audio Codecs, Computer Vision
Coursework: Linear Algebra, Data Structures and Algorithms, Introduction to Programming in C, Probability and Statistics
Libraries: Pytorch, WanDB, Matplotlib, OpenGL, DearImGUI, Librosa, Pandas, Numpy

Publications

Analysis of Speaker Verification Performance Trade-offs with Neural Audio Codec Transmission
Nirmalya Mallick Thakur, Jia Qi Yip and Eng Siong Chng (Aug 2025) [APSIPA 2025]

Experience

Research Intern | Speech and Language Lab Feb, 2025 – Present
Nanyang Technological University, Singapore

- Performed literature review of state-of-the-art neural audio codec (NAC) papers like DAC, Encodec, SD-Codec and SpeechTokenizer
- Conducted semantic analysis and fine-tuning of a NAC on Vietnamese ASR datasets using frozen discrete-code representations and seq-to-seq loss
- Worked on bidirectional semantic-to-acoustic (S2A) and acoustic-to-semantic (A2S) mapping of embeddings in NACs
- Analyzed the impact of NAC compression on performance of speaker verification (SV) systems on the Voxceleb1 dataset

Summer Research Intern | LEAP Lab May, 2025 – July, 2025
Indian Institute of Science, Bangalore

- Examined state-of-the-art blind source separation (BSS) models and evaluated them through SI-SDR and WER metrics using SpeechBrain and Asteroid toolkits
- Investigated single and dual channel audio-based source distance estimation (SDE) using engineered phase, magnitude and ILD based features
- Initiated behavioral pilot study for spatial perception in a dichotic listening setup
- Conducted LLM based complexity analysis in multilingual narrative tasks using fine-tuned models

Projects

DiffuseNet | ***Generative models trained from scratch (DDPM, DiT, VAE)*** *Github Link*

- Studied research papers such as DDPM, Flow matching, Vision Transformer, Diffusion Transformer and LDM
- Created a U-Net architecture with a linear noise scheduler based on the original "DDPM" paper
- Trained a variational autoencoder (VAE) from scratch using a combination of LPIPS, reconstruction and adversarial losses inspired by SD-VAE
- Implemented a 1.2B parameter DiT architecture from scratch for unconditional denoising and image generation
- Trained the DiT on extracted frames from Minecraft and Pokemon gameplay on 4x A100s and achieved high fidelity scene generations
- Added text conditioning through cross-attention with classifier free guidance using CLIP encoder and Qwen2.5-VL for extracting captions

Sim3D | ***Physics Engine for real-time Cloth and particle simulation*** *Github Link*

- Tools and technologies used: C++, OpenGL, DearImGUI, GLSL(OpenGL Shading Language)
- Developed a 3D physics engine written entirely from scratch in C++ using OpenGL to simulate Newtonian Mechanics, Particle-particle interactions, and real-time cloth simulation using mass-spring systems
- Particle system with efficient collision detection using Uniform grid partitions
- Interactive scene editor using ImGui for easy configuration and gizmos for intuitive scene manipulation
- Applied OOP concepts to abstract and integrate different engine components, improving engine design

ENCODEC | ***High Fidelity Neural Audio Compression (Paper Reproduction)*** *Github Link*

- Tools and technologies used: PyTorch, Google Colab, Librosa
- Reproduced a simple version of Meta AI's Encodec paper from scratch inspired by their codebase
- Created a convolution based encoder-decoder architecture with residual vector quantization (RVQ) for compressing latent audio embeddings into discrete codes
- Implemented a multi-scale STFT discriminator with adversarial loss to enhance output quality
- Trained the model using several loss components like reconstruction loss, perceptual loss, commitment loss and generator loss on the LibriSpeech development dataset
- The model can handle multiple target bandwidths like 1.5, 3, 6, 12, 24 kbps and output in both streaming and non-streaming fashion

DeepLLMs | ***BPE Tokenizer and Decoder-only Transformer from scratch*** *Github Link*

- Tools and technologies used: PyTorch, Google Colab
- Implemented a custom Byte Pair Encoding(BPE) tokenizer from scratch and trained it on 100K characters achieving a compression ratio of 3.61x
- Developed a 13M-parameter decoder-only Transformer model inspired by the paper "Attention is All You Need"
- Implemented Multihead Self Attention from scratch with 6 single heads
- Trained the model using the custom tokenizer on two datasets of roughly 500K and 1.5M tokens and generated human like outputs

Positions of Responsibility

Core Committee: CyberHeathens, Coding Club, IISER Bhopal 2023 – 2024