Shreshth Rai

+91 9863265347 shreshthrai.com | LinkedIn | GitHub work.shreshth25@gmail.com

EDUCATION

Year	Degree / Examination	Institute	CGPA / %
2028	B.Tech in Mathematics and Computing	Delhi Technological University, Delhi	8.3 CGPA
2024	CBSE Class XII	Sanskriti School, Chanakyapuri, Delhi	93%
2022	CBSE Class X	G.Rio School, Kohima, Nagaland	98.8%

EXPERIENCE

Bitrix Innovations Pvt. Ltd. | Full Stack AI Intern *Jan 2026 - Present*

- **Deployed a containerized web platform** on AWS EC2 using Docker and Nginx, serving S3-hosted videos via a dynamic **React UI backed by MongoDB metadata**, with SSL-secured routing for all application components.
- Built an **LLM-driven video synthesis pipeline** using React (Remotion) and a **multi-agent interview simulation engine**, reducing API token usage by 33% and redundant calls by 68% for scalable course content and low-latency interview simulations.

Dept. of Applied Mathematics, DTU | AI Research Intern *July 2025 - October 2025*

- Participated in the **SAND (Speech Analysis for Neurodegenerative Diseases) Challenge** at ICASSP, implementing deep learning and machine learning techniques for voice activity detection and biomarker extraction from speech signals.
- Developed **BiLSTM and Vision Transformer** models to analyze prosodic features such as intonation, rhythm, and pauses for distinguishing neurodegenerative conditions including Alzheimer's disease.

PUBLICATIONS

Inference-Time Expert Subsets for Gradient-Misaligned Robustness *(Under Review)*

Improved adversarial robustness of image classifiers via an **inference-time stochastic ensembling** method, reducing attack success without retraining. Built from independently adversarially trained models with random subset selection. Achieved **SOTA 74.66% / 43.5%** robust accuracies on CIFAR-10/100 under **AutoAttack**, the standard worst-case adversarial robustness benchmark.

SELECTED PROJECTS

DECIBEL: Unified Audio Perception and Multilingual Reasoning LM | **Mixture-of-Experts, Multimodal Reasoning**

- Redesigned audio-language paradigm using a novel **Mixture-of-Experts** approach with **7 expert models** and distributed weights.
- Built **distilled and quantized expert models** with **10.3% WER**, **97% accuracy**, **0.439 mAP**, **+20 BLEU**, using **<6GB VRAM**.
- **Finetuned Falcon-7B** for unified Audio+Text modeling using **FT-Adapters, LoRA, and Wav2Vec2 embeddings**.

Fashionate: a Multimodal Fashion Recommendation System | **Multimodal Learning, GANs**

- Built a **multimodal fashion recommendation system** using **vision–language embeddings** and large-scale similarity search.
- Developed APIs for **natural-language product search**, **user preference modeling**, and personalized outfit recommendations.
- Implemented a **virtual try-on pipeline** using **SpadeGAN** and DensePose-based garment warping, achieving **SSIM 0.77**.

Vision-SSL: Self-Supervised Visual Representation Learning | **Self-Supervised Learning, Computer Vision**

- Implemented pipelines for **contrastive** and **masked image modeling** approaches to SSL, based on *SimCLR* and *MAE*.
- Benchmarked custom SSL models on ImageNet-100 (~130K images), studying representation quality under limited supervision.
- Achieved **98% accuracy with SimCLR** and **89% accuracy with MAE** on downstream classification.

SKILLS

- **Machine Learning & AI:** Transformers, LLMs, SFT, PEFT, RAG; CNNs, ViTs, Diffusion Models, GANs, VAEs, UNets; Self-Supervised Learning, Multimodal Models (VLMs, ALMs)
- **ML Frameworks & Data:** PyTorch, TensorFlow, Pandas, NumPy, FAISS, ChromaDB, Weights & Biases (wandb)
- **Backend & Databases:** FastAPI, PostgreSQL, MongoDB, Redis, Supabase, SQL
- **Cloud & DevOps:** AWS, GCP, Docker, Kubernetes, Nginx, Gitea/Gitlab CI/CD
- **Programming:** Python, C++, Java, Kotlin, JavaScript, C, Bash
- **Certifications:** Machine Learning Specialization, Deep Learning Specialization – Stanford University

ACHIEVEMENTS

2nd Runner-up | Hack4Delhi (Government of Delhi)

- Built a **vision-language pipeline** for **railway tampering detection** using sliding-window inference and streaming attention.
- Optimized video processing with **FlashAttention 2.0** and temporal sampling for efficient multimodal reasoning.

2nd Runner-up | FinAgent (IIT Bombay)

- Built a **browser-based autonomous financial agent** using Salesforce AI Research's **xLAM** for multi-step action execution.
- Introduced a **human-in-the-loop recovery mechanism** (*Conscious Pause*) for long-horizon planning.

Grand Finalist | Smart India Hackathon 2025 (DRDO)

- Developed a **unified Audio Language Model** for joint audio understanding, representation learning, and reasoning in mission-critical settings, with emphasis on robustness under noise and low-latency.
- Focused on **on-prem inference**, **model robustness**, and aspects like privacy relevant to defense and surveillance applications.

POSITIONS OF RESPONSIBILITY

Co-Head, *AIMS DTU* *September 2024 - Present*

- Conduct knowledge transfer sessions for freshmen, teaching a comprehensive machine learning curriculum and mentoring teams for technical hackathons.