Naresh Kumar Devulapally

United States | +1 (716) 256-7463 | devulapa@buffalo.edu | LinkedIn | naresh-devula.github.io

Summary

- PhD Candidate in Computer Vision, specializing in **Multimodal Video Understanding, Temporal Modeling and Vision-Language Models**, seeking full-time Research Scientist roles, starting **Dec 2026**.
- Strong publication record at **CVPR, ICCV, AAAI, CoLM, ACM MM**, with a proven track record of developing and productionizing AI systems across Generative AI and Multimodal Learning.

Education

The State University of New York at Buffalo May 2024 – May 2027 (Expected)
Ph.D., *Computer Vision and Machine Learning* | PhD Thesis Slides Buffalo, New York, USA
Advisor: Dr. Vishnu Lokhande, Committee: Dr. Siwei Lyu, Dr. Junsong Yuan

- ***Chair's Fellowship for Research Excellence**, University at Buffalo, SUNY, 2024*
- ***CSE Graduate Teaching Award for Teaching Excellence**, University at Buffalo, SUNY, 2025*

The State University of New York at Buffalo August 2022 – May 2024
Master of Science in Computer Science and Engineering Buffalo, New York, USA
MS Thesis: *Towards Generalizable and Robust Multimodal ML Models*

National Institute of Technology, Tiruchirappalli June 2016 – June 2020
Bachelor of Technology in Mechanical Engineering Tiruchirappalli, Tamil Nadu, India

Work Experience

Computer Vision and Machine Learning Lab | *Research Assistant, Ph.D.* **May 2024 – Present**
Video-Language Models, Multimodal Learning, World Foundation Models Buffalo, NY, USA

- Watermarking **NVIDIA Cosmos world foundation models** in collaboration with NVIDIA: LoRA on the generation stream of Cosmos3-Nano, parameter-matched against Meta's Pixel Seal, wins **all 18 attack conditions at 5 matched quality levels**, by +10.5 points clean and +19.2 under VAE regeneration.
- Developed a prompt-inversion framework in diffusion LLMs giving interpretable, edit-friendly prompts with a **∼10× reduction in inversion time**. *(CVPR 2026, with Adobe Research)*
- Built object-level, in-generation watermarking for text-to-image latent diffusion by fine-tuning only the text token embeddings: **99% bit accuracy at 48 bits with $10^5\times$ fewer trained parameters**. *(ICCV 2025, with Adobe Research)*

Adobe | *Research Scientist Intern* | *Manager: Dr. John Collomosse* **May 2026 – Present**
Multimodal Video-Audio Learning, Cross-Modal Fusion, Distributed Training San Jose, CA, USA

- **Trained a multimodal video-audio model from scratch** for content authenticity: bidirectional cross-attention at encoder and detector fuses 8-frame 256×256 video with 48 kHz audio into a **shared 384-dim space**, read out by **256 learned queries**.
- Added a coalition token spanning both-present, video-only and audio-only states for **missing-modality tolerance**, at **100% clean bit accuracy** on AVA and VGGSound.
- Pre-trained on paired **Kinetics-700** clips with distributed multi-node training and mixed precision: **97.2% / 96.3%** across 24 paired codec conditions (H.264 CRF 28, H.265 CRF 26, AAC/MP3 16 kbps, 8 kHz resampling) at ∼48 dB video PSNR, **+3.2 / +5.1 points** over VideoSeal and HierMark.

Visual Computing Lab | *Research Assistant, MS Thesis* **August 2022 – May 2024**
Multimodal Learning, Transformers, Audio-Video-Text Modeling Buffalo, NY, USA

- Built **multimodal transformer and knowledge-distillation** architectures fusing **audio, video and text** for conversational emotion recognition, on datasets of **∼1M utterances**, with up to **∼10% improvement** over unimodal baselines. *(ACM MM 2023, BigMM 2023)*
- Handled **missing and misaligned modalities** at inference via joint embedding learning, so recognition degrades gracefully when a stream drops. *(ICME 2024)*

Carbynetech, Inc. | *Senior Data Scientist* **August 2020 – July 2022**
Production AI, Object Detection, Forecasting Models, Optimization Hyderabad, India

- Deployed a real-time object detection pipeline using Flask APIs on Azure Web Services for export-pallet inspection; **achieved 94% mAP** and reduced manual inspection time by over 60%.

Publications

10. Mahesh Bhosale†, **Naresh Kumar Devulapally**†, Abdul Wasi, Chau Pham, Vishnu Suresh Lokhande, David Doermann. "Score-Control for Hallucination Reduction in Diffusion Models." *Preprint*, 2026.
9. Yash Shah, Abhijit Chakraborty, **Naresh Kumar Devulapally**, Vishnu Lokhande, Vivek Gupta. "OSCAR: Orchestrated Self-verification and Cross-path Refinement." *Conference on Language Modeling (CoLM)*, 2026.
8. **Naresh Kumar Devulapally**, Shruti Agarwal, Vishal Asnani, Vishnu Lokhande. "Interpretable Prompts made Edit-Friendly: Token-to-Token Similarity Reduction in dLLMs for Edit-Friendly Hard Prompt Inversion." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.
7. Arjun Ramesh Kaushik, **Naresh Kumar Devulapally**, Vishnu Suresh Lokhande, Nalini Ratha, Venugopal Govindaraju. "Forget Less by Learning Together through Concept Consolidation." *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2026.
6. Arjun Ramesh Kaushik, **Naresh Kumar Devulapally**, Vishnu Suresh Lokhande, Nalini Ratha, Venugopal Govindaraju. "Forget Less by Learning from Parents Through Hierarchical Relationships." *AAAI Conference on Artificial Intelligence (AAAI)*, 2026.
5. **Naresh Kumar Devulapally**, Mingzhen Huang, Vishal Asnani, Shruti Agarwal, Siwei Lyu, Vishnu Lokhande. "Your Text Encoder can be an Object-Level Watermarking Controller." *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.
4. **Naresh Kumar Devulapally**, Shruti Agarwal, Tejas Gokhale, Vishnu Lokhande. "Latent Diffusion Unlearning: Protecting Against Unauthorized Personalization Through Trajectory Shifted Perturbations." *ACM International Conference on Multimedia (ACM MM)*, 2025.
3. **Naresh Kumar Devulapally**, Sidharth Anand, Sreyasee Das Bhattacharjee, Junsong Yuan. "Adaptive Missing-Modality Emotion Recognition in Conversations via Joint Embedding Learning." *IEEE International Conference on Multimedia and Expo (ICME)*, 2024.
2. **Naresh Kumar Devulapally**†, Sidharth Anand†, Sreyasee Das Bhattacharjee, Junsong Yuan. "Multilabel Emotion Analysis in Conversation via Multimodal Knowledge Distillation." *ACM International Conference on Multimedia (ACM MM)*, 2023.
1. **Naresh Kumar Devulapally**, Sidharth Anand, Sreyasee Das Bhattacharjee, Junsong Yuan. "AMuSE: Adaptive Multimodal Analysis for Speaker Emotion Recognition in Group Conversations." *IEEE International Conference on Multimedia Big Data (BigMM)*, 2023.

Technical Skills

- **Languages & Frameworks:** Python, C++, CUDA, PyTorch (DDP/FSDP), JAX/Flax, TensorFlow, Hugging Face (Transformers, Diffusers, Accelerate, PEFT), OpenCV
- **Multimodal & Video Understanding:** Video Representation Learning, Temporal Modeling, Long-Context Video, Audio-Video-Text Fusion, Vision-Language Models (VLM/MLLM), Multimodal Retrieval & Embedding, Missing-Modality Robustness, CLIP
- **Generative & Foundation Models:** Diffusion Models, Latent Diffusion, Text-to-Image/Video Generation, World Foundation Models (NVIDIA Cosmos), Large-Scale Pre-training, Fine-tuning (LoRA/PEFT)
- **Systems & Scaling:** Distributed / Multi-Node Training (Slurm), Inference Optimization, Docker, Kubernetes, AWS, GCP, Azure