Bindita Chaudhuri

bindita91@gmail.com bindita.github.io binditachaudhuri U.S. Green Card holder

Summary

Senior Applied Research Scientist with 10+ years of experience, including 5+ years in industrial R&D, designing, scaling, and productionizing multimodal Generative AI and Computer Vision systems. Expertise spans audio-visual foundation model post-training, multimodal video generation, synthetic audio, 3D human reconstruction (speech-driven facial animation) and real-time neural rendering. Led technical roadmaps and mentored researchers at Flawless AI and Meta Reality Labs; Ph.D. in Computer Science from University of Washington, with publications at top-tier venues (CVPR, ECCV).

Technical Skills

Generative AI: Multimodal Foundation Models, VLM, Diffusion Transformers, Autoregressive video synthesis, GAN, VAE
GenAI Training: Large-Scale Data Curation, Distributed Training, Post-Training (PEFT/DPO), Step Distillation
Core Languages & Deep Learning: Python, PyTorch, TensorFlow, C/C++, CUDA, Swift
Research Domains: Audio-Visual learning, Multimodal video generation/editing, Digital avatars, 3D Computer Vision

Work Experience

Flawless AI Los Angeles, CA
Senior Applied Research Scientist Oct 2023 - Present

- Architected a **multimodal video generation** network by LoRA finetuning of **large audio-visual foundation models** for audio-driven facial animation. Optimized the inference latency using distribution matching distillation.
- Built an end-to-end **video scene text editing pipeline** integrating VLM-based OCR, SAM2 segmentation, masked video-to-video diffusion/flow-matching models (VACE) and VLM-as-judge evaluation, **unlocking a new product capability** for automated global localization.
- Redesigned and shipped a **neural rendering GAN architecture** for **real-time lip-sync inference**, improving perceptual accuracy by 23% under challenging production conditions - occlusions, film grain, and extreme poses. Deployed in multiple feature films including ***Watch the Skies* on Amazon Prime Video**.
- Engineered a **multilingual synthetic audio** generation system with voice and emotion preservation by orchestrating ASR transcription, context-aware LLM translation and Qwen3-TTS.
- **Led a team of 5+ scientists** and coordinated priorities across engineering and product teams, defining the following technical roadmaps for visual dubbing from foundational research to production deployment:
  – Improved 3D face modeling/tracking for beards using high-dimensional **Gaussian Splatting**.
  – Deployed speaker diarization and crosstalk separation for **audio refinement** capabilities.
  – Adapted deepfake detection techniques into **automatic QA** metrics for corpus filtering and model evaluation.
- Supervised an intern project on **text-controlled 3D human** body part motion generation and editing (*ICCV 2025*).

Meta Reality Labs Redmond, WA
Research Scientist Jul 2021 - Aug 2023

- Scaled and optimized masked generative transformers for text- and image-conditioned video synthesis using **FSDP and mixed-precision training across multi-node GPU clusters**, improving training throughput by 11%.
- Designed **TalkinNeRF**, an animatable NeRF for reconstructing full-body talking humans with fine-grained hand and face details from text and speech inputs (*ECCVW 2024*).
- Mentored an intern on a novel multimodal coarse-to-fine approach for **detailed 4D face geometry reconstruction from in-the-wild videos** for AR/VR applications (*CVPR 2023*).
- Improved the realism of **3D video calling in Quest headsets** by integrating neural processing of emotion and non-speech vocalizations in speech-driven 3D facial animation.
- Architected a spatio-temporal dynamic video accumulation network enabling self-occlusion-free face texture completion. Spearheaded enhanced video texture recovery for synthetic faces using a **customized StyleGAN**.

Facebook Reality Labs Research Sausalito, CA
Research Intern Summer 2020

- Created a region-adaptive VAE to synthesize photorealistic texture maps for 3D humans with independent geometry and appearance controllability targeting **virtual try-on applications** and semi-supervised synthetic data generation (*CVPR 2021*).

Microsoft Cloud & AI Redmond, WA
Student Researcher Oct 2018 - Dec 2019

- Developed a high-fidelity personalized 3D face avatar generation pipeline from user videos, reducing identity drift (*ECCV 2020 Spotlight*). Stabilized **real-time face tracking** algorithms optimized for challenging, unconstrained in-the-wild video streams.

Microsoft Research Redmond, WA
Research Intern Summer 2018

- Designed a real-time multi-task learning framework for joint 2D face detection and 3D facial motion retargeting (*CVPR 2019*); successfully **shipped the model** to production for **SwiftKey Puppets**, scaling to millions of active Android users for immersive virtual communication.

Intel Labs Santa Clara, CA
Research Intern Summer 2017

- Implemented a Spatial Transformer Network for optical flow prediction and image super-resolution to enable high-quality frame interpolation from HD multi-camera arrays. U.S. **patent** granted.

Education

Ph.D. **University of Washington**, *Computer Science* 2016 – 2021
- *Honors*: People's Choice Award at UW Research Day (*GPA: 3.72/4.0*)
- *Thesis*: Deep Facial Expression Modeling and 2D-to-3D Motion Retargeting

M.E. **Indian Institute of Technology Bombay**, *Electrical Engineering* 2014 – 2016
- *Honors*: Department Academic Excellence Award (*GPA: 9.83/10.0*)
- *Thesis*: Region-based Remote Sensing Image Retrieval using Graph Theory

B.E. **Jadavpur University**, *Electronics and Telecommunication Engineering* 2010 – 2014
- *Honors*: University Gold Medal, The Supriya Basu Scholarship (*GPA: 9.65/10.0*)
- *Thesis*: Low Cost Low Bandwidth Virtual Education Platform Design

Selected Publications from

KinMo: Kinematic-aware Human Motion Understanding and Generation [link] ICCV 2025
TalkinNeRF: Animatable Neural Fields for Full-Body Talking Humans [link] ECCVW 2024
Semi-supervised Synthesis of High-Resolution Editable Textures for 3D Humans [link] CVPR 2021
Personalized Face Modeling for Improved Face Reconstruction and Motion Retargeting [link] ECCV 2020 (Spotlight)
Joint Face Detection and Facial Motion Retargeting for Multiple Faces [link] CVPR 2019
Learning to Generate 3D Stylized Character Expressions from Humans [link] WACV 2018
Deep Learning for Multi-Camera View Interpolation [link] U.S. Patent *10,547,823*, 2020

Professional Activities

- **Conference Reviewer**: CVPR, ECCV, ICCV, NeurIPS, SIGGRAPH Asia, BMVC, ICLR, ICPR, IEEE VR, AAAI
- **Journal Reviewer**: ACM TOG, IEEE TPAMI, CVIU, IEEE TGRS, IEEE GRSL
- **Invited Talk**: Beyond PhD Conference, University of Southern California
- **Area Chair (Student)**: CSE Graduate Admissions Committee, University of Washington (2020).
- **Teaching Assistant**: University of Washington (CSE) and IIT Bombay (EE).