

Pinxin Liu (Andy)

andypinxinliu@gmail.com ◊ +1-585-540-7309 ◊ [Homepage](#) ◊ [Google Scholar](#)

Job Interest

Seeking *Research / Applied Scientist / ML Engineer* positions in *Multimodal Language Models, Human Video Generation (3D / Diffusion), Video Understanding (MLLM), Audio-Visual Learning, and 3D Perception & Understanding*. My master's research focused primarily on language models and human-centered perception/generation.

Education

University of Rochester

M.S., Computer Science

B.S., Computer Science; Highest Honor Distinction in Research; GPA: 3.83/4.0

Rochester, NY

Jan 2025 – May 2026 (expected)

Jul 2020 – May 2024

Experience

Meta, Reality Lab — Research Scientist Intern

May 2025 – Dec 2025

Report to: [A. Richard](#), [D. Markovic](#)

- *LLM-based Motion Understanding & Generation*: Integrated **3D body mesh representations** into an **LLM tokenization pipeline** with a **diffusion/flow-matching generation head**, enabling unified cross-modal understanding and controllable fine-grained body motion generation; unlocking instruction-following avatar animation at scale for interactive digital human experiences on Meta Quest and AR devices.
- *Automatic Data Annotation at Scale*: Designed an automated annotation pipeline converting raw 3D pose and mesh sequences into structured semantic language descriptions without human labelers; reducing data curation cost by orders of magnitude and enabling training on **internet-scale motion corpora** that was previously infeasible.

Flawless AI — Research Scientist Intern

Jun 2024 – Dec 2024

Report to: [P. Garrido](#), [A. Shapiro](#)

- *Foundation Human Video Generation*: Built a **large-scale diffusion-based foundation model** with **multi-stage training** on **100M+ video-text pairs** using FSDP across multi-node GPU clusters, establishing the core generative backbone for the company's film production pipeline.
- *Speech-to-Gesture Alignment*: Developed **cross-modal alignment** between pixel-level motion and **speech semantic embeddings**; shipped a upper-body co-speech gesture animation system in production for film post-production clients.
- *Pixel-Level Artifact Refinement*: Shipped a **pixel-space refinement module** suppressing boundary artifacts across diverse backbones without retraining, adopted as a plug-and-play quality layer across video generation products.

Selected Publications (* Equal contribution)

- **Pinxin Liu***, L. Song*, et al. Bridging Facial Understanding and Animation via Language Models. *CVPR*, 2026.
- **Pinxin Liu**, et al. GestureLSM: Latent Shortcut based Co-Speech Gesture Generation. *ICCV*, 2025. [[Paper](#)] [[Code](#)]
- **Pinxin Liu***, et al. KinMo: Kinematic-aware Human Motion Understanding and Generation. *ICCV*, 2025. [[Paper](#)]
- **Pinxin Liu***, et al. Contextual Gesture: Co-Speech Gesture Video Generation. *ACM MM*, 2025. [[Paper](#)] [[Demo](#)]
- **Pinxin Liu***, Y. Tang*, et al. MMPerspective: Do MLLMs Understand Perspective? *NeurIPS*, 2025. [[Paper](#)] [[Code](#)]
- **Pinxin Liu***, L. Song*, et al. GaussianStyle: Gaussian Head Avatar via StyleGAN. *3DV*, 2025. [[Paper](#)] [[Demo](#)]
- **Pinxin Liu***, et al. Tri²-plane: Volumetric Avatar Reconstruction with Feature Pyramid. *ECCV*, 2024. [[Paper](#)] [[Demo](#)]

Honors & Awards

[National Student Employee of the Year](#) (2024)

NVIDIA Academic Grant: \$110,000 (2025)

Skills

Systems & Engineering: Large-scale distributed training & optimization (multi-node clusters, DDP/FSDP); Cloud & DevOps (AWS, Docker, Git); Model compression & acceleration (distillation, AMP).

Programming Languages: Python, C++, CUDA, JavaScript, SQL, Bash, Java.

Libraries & Tools: PyTorch, JAX, TensorFlow, HuggingFace, NumPy, NLTK, SpaCy, W&B.