

MINGFEI CHEN

+1(206) 945-4355 ◊ lasiafly@uw.edu ◊ Seattle, WA
◊ [Personal Website](#) ◊ [Google Scholar](#) ◊ [LinkedIn](#)

EDUCATION

University of Washington	Sep 2023 -
Ph.D. in Electrical and Computer Engineering, GPA: 3.98/4.00	Seattle, WA
Google PhD Fellowship 2025.	
University of Washington	Sep 2021 - Jun 2023
M.S. in Electrical and Computer Engineering, GPA: 3.98/4.00	Seattle, WA
Huazhong University of Science and Technology	Sep 2016 - Jun 2020
B.S. in Computer Science and Technology, GPA: 3.96/4.00	Wuhan, China
Undergraduate excellent student and outstanding undergraduate graduation thesis awards.	

PUBLICATIONS

- [1] **Mingfei Chen**, Yue Liu, Yifan Wang, Zhengqing Li, Homanga Bharadhwaj, Yujin Chen, Chuan Qin, Yuan Tian, Ziyi Kou, Eric Whitmire, Raj Sodhi, Eli Shlizerman, Hrvoje Benko. “**Flowing from Reasoning to Motion: Learning 3D Hand Trajectory Prediction from Egocentric Human Interaction Videos.**” Under review.
- [2] **Mingfei Chen***, Zijun Cui*, Xiulong Liu*, Jinlin Xiang, Caleb Zheng, Jingyuan Li, Eli Shlizerman. “**SAVVY: Spatial Awareness via Audio-Visual LLMs through Seeing and Hearing.**” Neural Information Processing Systems (NeurIPS), 2025 (**Oral - Acceptance Rate < 0.4%**).
- [3] **Mingfei Chen**, Israel D. Gebru, Ishwarya Ananthabhotla, Christian Richardt, Dejan Markovic, Steven Krenn, Todd Keebler, Jacob Sandakly, Eli Shlizerman, Alexander Richard. “**SoundVista: Novel-View Ambient Sound Synthesis via Visual-Acoustic Binding.**” Computer Vision and Pattern Recognition (CVPR), 2025 (**Highlight - Acceptance Rate < 2.9%**).
- [4] **Mingfei Chen**, Eli Shlizerman. “**AV-Cloud: Spatial Audio Rendering Through Audio-Visual Cloud Splatting.**” Neural Information Processing Systems (NeurIPS), 2024.
- [5] **Mingfei Chen**, Kun Su, Eli Shlizerman. “**Be Everywhere - Hear Everything (BEE): Audio Scene Reconstruction by Sparse Audio-Visual Samples.**” International Conference on Computer Vision (ICCV), 2023.
- [6] **Kun Su***, **Mingfei Chen***, Eli Shlizerman. “**INRAS: Implicit Neural Representation for Audio Scenes.**” Neural Information Processing Systems (NeurIPS), 2022.
- [7] **Mingfei Chen**, Jianfeng Zhang, Xiangyu Xu, Lijuan Liu, Yujun Cai, Jiashi Feng, Shuicheng Yan. “**Geometry-Guided Progressive NeRF for Generalizable and Efficient Neural Human Rendering.**” European Conference on Computer Vision (ECCV), 2022.
- [8] **Mingfei Chen***, Yue Liao*, Si Liu, Zhiyuan Chen, Fei Wang, Chen Qian. “**Reformulating HOI Detection as Adaptive Set Prediction.**” Computer Vision and Pattern Recognition (CVPR), 2021.
- [9] Jingyuan Li, Trung Le, Chaofei Fan, **Mingfei Chen**, Eli Shlizerman. “**Brain-to-Text Decoding with Context-Aware Neural Representations and Large Language Models.**” *Journal of Neural Engineering*, Oct. 2025.
(* equal contribution.)

RESEARCH PROJECTS

Research Interest: Multi-modal research for spatial reasoning in 3D scenes, as well as the related applications on multi-modal LLMs, XR devices and robotics.

Meta Reality Labs Research	Jun 2025 – Present
Mentors: Yue Liu, Homanga Bharadhwaj, Yifan Wang, Zhengqing Li	Redmond, WA
Project: Flowing from Reasoning to Motion: Learning 6DoF Hand Trajectory Prediction from Egocentric Human Interaction Videos	

- Built EgoMAN, the first large-scale egocentric benchmark for *stage-aware 3D hand trajectory prediction* with structured visual–linguistic–motion reasoning supervision.
- Proposed a reasoning-to-motion architecture that connects VLM intent reasoning with 6DoF trajectory generation via structured token interfaces and progressive training.
- Achieved state-of-the-art 6DoF motion forecasting, with improved interpretability and cross-task generalization on diverse, unseen egocentric manipulation scenarios.

NeuroAI Lab, University of Washington, Seattle

Advisor: Eli Shlizerman

Dec 2021 - Present
Seattle, WA

Project: Multi-modal for Spatial Reasoning in 3D Scenes (Ongoing)

- Develop comprehensive models for precise 3D egocentric perception and interpretation of Audio-Visual Large Language Models (AV-LLMs) in real-world environments, integrating visual, acoustic, and additional sensory data for an efficient perceptual system. Focus on smart multi-modal data usage for optimal capture sampling and efficient memory storage and representation.
- Explore how various modalities can benefit AV-LLMs and other downstream tasks for AR scenes and robotics, including 3D scene understanding and reconstruction, egocentric acoustic modeling, and 3D interaction understanding.
- SAVVY: We introduce SAVVY-Bench, the first benchmark for 3D spatial reasoning in dynamic audio-visual scenes. We also propose SAVVY, a novel training-free pipeline integrating egocentric spatial tracks and dynamic global maps, which significantly enhances AV-LLM performance for improved audio-visual spatial awareness in such environments. (NeurIPS25 Oral)

Project: 3D Audio-Visual Representation and Reconstruction from Sparse Audio-Visual Samples

- Explored audio-visual scene representation for 3D world to support the cross-modal tasks such as spatial audio scene reconstruction.
- AV-Cloud: Proposed the Audio-Visual Cloud, a set of sparse AV anchor points derived from camera calibration, to learn the audio-visual scene representation and generate spatial audio for arbitrary listener locations. Developed a novel approach for rendering high-quality spatial audio in 3D scenes that is synchronized with the visual stream but does not rely on explicit visual cues. (NeurIPS24)
- BEE: Developed a real-time end-to-end integrated rendering pipeline, for dynamic audio reconstruction from sparse audio-visual samples. (ICCV23)
- INRAS: Proposed a method for high-fidelity impulse responses modeling at any emitter-listener positions. (NeurIPS22)

Meta Codec Avatars Lab

Mentors: Israel D. Gebru, Alexander Richard

Jun 2024 - Dec 2024
Pittsburgh, PA

Project: Visually-guided Ambient Sound Field Modeling for Navigable Codec Spaces (CVPR25 Highlight)

- Developed a novel approach, SoundVista, to reconstruct target binaural sound using reference ambisonic microphones at arbitrary locations, enhancing ambient sound modeling in diverse environments.
- Created benchmarks in both simulated Soundspace environment and real-world captured room, conducting research from synthetic agents to real.
- Pioneered the integration of 3D visual capture data with acoustic parameters to create a pre-trained visual representation model, facilitating improved acoustic scene analysis. Utilized the model to sample optimal reference microphone positions, which can significantly boost performance within the same microphone budget.
- Implemented adaptive reweighting of input audio channels using the pretrained visual model and room geometry, optimizing reconstruction across varying room sizes and layouts without constraints on microphone quantity.

National University of Singapore & Sea AI Lab

Advisor: Shuicheng Yan, Jiashi Feng

Jun 2021 - Nov 2021

Singapore

Project: Controllable High-Fidelity 3D Human Rendering Under Sparse Views (ECCV22)

- Developed a geometry-guided progressive NeRF for efficient human body rendering from sparse views (3 cameras, 120° apart), with a multi-view feature integration approach for enhanced occluded information accuracy using geometry guidance.
- Curated a large-scale dataset for pretraining and validated the method's real-world robustness.
- Achieved 70% rendering time reduction, taking just 175ms on RTX 3090, outperforming industry standards.

Sensetime & University of Washington, Seattle

Advisor: Jenq-Neng Hwang

Nov 2020 - Jun 2021

Beijing, China

Project: Online Multi-Object Tracking (MOT)

- Proposed an innovative online MOT framework for tailored feature aggregation in detection and association.
- Crafted a track association module for combined location and appearance matching using motion and appearance embeddings.
- Achieved state-of-the-art online MOT tracking on MOT17, enhancing association effectiveness and maintaining top detection accuracy.

Sensetime Research

Advisor: Si Liu

Jul 2020 - Mar 2021

Beijing, China

Project: Human-Object Interaction (HOI) Detection (CVPR21)

- Formulated HOI detection as a set prediction problem, addressing instance-centric and location limitations.
- Developed a one-stage HOI framework with a transformer for optimal feature aggregation and introduced an instance-aware attention module.
- Achieved 31% accuracy improvement over leading one-stage methods on the HICO-DET dataset without additional features.

AWARDS & HONORS

[2025] Google PhD Fellowship 2025 in Machine Perception (North America).

[2023] International Conference on Computer Vision Diversity, Equity & Inclusion (DEI) Award.

[2020] Huazhong University of Science and Technology Outstanding Undergraduate Graduation Thesis.

[2018, 2019] Huazhong University of Science and Technology Merit Student Scholarship.

[2018] Meritorious Winner of Mathematical Contest In Modeling (MCM/ICM).

[2017] Huazhong University of Science and Technology Undergraduate Excellent Student (Top 1% in 35000).

TEACHING & LEADERSHIP

UW EE497/498/598: Engineering Entrepreneurial Capstone

Graduate Teaching Assistant

Dec 2022 - Present

Seattle, WA

- Lead TA of the Engineering Innovation and Entrepreneurship (ENGINE) capstone program ([program website](#)).
- Mentored more than 100 students in all stages of preparation of their Capstone. Responsibilities include: assist in defining project scope, monitor progress, coordinate resources.
- Oversee 7 TAs in organizing the large-scale program.

UW SP22 EE596: Introduction to Deep Learning Applications and Theory

Graduate Teaching Assistant

Mar 2022 - Jun 2022

Seattle, WA

- Held lab sessions, developed examples and instruction notes of weekly lab assignments.

SERVICE

- Reviewer for CVPR, NeurIPS, AAAI, ICLR, ICML, IEEE TPAMI.
- Student organizer of University of Washington NeuroAI Seminar.
- Session chair of ICCV-23 AV4D workshop.