

Zeyang Zhang

zeyangz@andrew.cmu.edu | [Google Scholar](#) | [zeyang.ai](#)

EDUCATION

Carnegie Mellon University (CMU)

Master of Science in Artificial Intelligence Engineering – Information Security

Thesis: Cross-Modal Concept Learning Through Hierarchical Quantized Embeddings

Advisor: Prof. Carlos Busso

Aug. 2024 – May 2026

GPA: 3.94/4.00

University of Illinois at Urbana-Champaign (UIUC)

Bachelor of Science in Computer Science

Sep. 2020 – May 2024

GPA: 3.94/4.00

RESEARCH INTERESTS

Multimodal Machine Learning, Natural Language Processing, Multimodal Representation Learning, Signal Processing, Multimodal Large Language Models, Wearable Sensing, Human-Centered AI, AI Agents

PUBLICATIONS

[1] Speaker Diarization in Static and Egocentric Videos via Speech-Face Alignment

Zeyang Zhang, Xavier Yin, Zhaobo Zheng, Kumar Akash, Teruhisa Misu, and Carlos Busso.

Accepted, ACM International Conference on Multimodal Interaction (ICMI), 2026. [\[doi\]](#)

[2] Hierarchical Quantized Cross-modal Masked Autoencoders for Audio-Visual Emotion Recognition

Zeyang Zhang and Carlos Busso.

Accepted, ACM International Conference on Multimodal Interaction (ICMI), 2026. [\[doi\]](#)

[3] Contrastive Gated Fusion for Multilingual Speaker Verification

Zeyang Zhang, Katsuhiko Naito, and Hajer Dahmani.

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2026. [\[doi\]](#) [\[code\]](#)

RESEARCH EXPERIENCE

Multimodal Speech Processing (MSP) Laboratory, CMU

Research Assistant, advised by Prof. Carlos Busso

Aug. 2025 – Present

Multimodal Modeling of Individual and Team Performance in Collaborative Settings Jan. 2026 – Present

- Collaborated with Honda Research Institute (HRI) on multimodal sensing and modeling of human behavior, individual performance, and team dynamics in collaborative work settings.
- Created synchronized multimodal recordings from 8 static and 5 egocentric cameras, integrating speech, video, IMU, gaze, and other biometric signals.
- Designed multi-view video annotation questionnaires for labeling performance, engagement, and leadership.
- Developing multimodal models that fuse facial behavior, speech, head-motion IMU, and gaze to predict individual and team performance, engagement, and leadership in dynamic, multi-party interactions.

Audio-Visual Distillation for IMU Representation Learning

May 2026 – Present

- Developed an IMU representation learning framework that combines IMU-only self-supervised learning, audio-visual knowledge distillation, and Mantis-8M for robust human activity recognition across domains.
- Adapted a gravity-aligned equivariant (GAE) front-end to encode rotation-invariant IMU features.
- Designed complementary pretraining strategies for in- and out-of-domain generalization, outperforming all existing baselines across human activity recognition benchmarks.
- Demonstrated complementary audio-visual supervision: video performed best in-domain, audio transferred best across devices, and joint distillation outperformed both. Plan to submit this work to ICASSP 2027.

Speaker Diarization in Static and Egocentric Videos via Speech-Face Alignment

Jan. 2026 – May 2026

- Collaborated with Honda Research Institute (HRI) to develop SEAD, a static-egocentric audio-visual speaker diarization framework that identifies “who spoke when” in dynamic, multi-party interactions, [1].
- Built a multi-stage pipeline combining DiariZen-based audio diarization, face detection/tracking, cross-camera identity matching, cross-audio speaker matching, and audio-visual speaker ID alignment.
- Fine-tuned an audio-visual active speaker detection model that fuses facial and audio cues to improve audio-only diarization, recovering missed speech, suppressing false alarms, and reducing speaker confusion.
- Improved diarization performance over an audio-only fusion baseline, reducing Diarization Error Rate from 28.65% to 15.83% and Jaccard Error Rate from 29.84% to 17.05%.

Hierarchical Quantized Cross-modal Masked Autoencoders for Audio-Visual Emotion Recognition

Aug. 2025 – May 2026

- Developed HQC-MAE, a self-supervised audio-visual emotion recognition framework that combines vector quantization, contrastive learning, and masked autoencoding to learn robust audio-visual representations, [2].
- Designed hierarchical discrete representation learning with 8 learnable VQ codebooks across audio, video, and fusion encoders to improve cross-modal alignment and suppress modality-specific noise.
- Evaluated HQC-MAE on MSP-IMPROV, MAFW, and CREMA-D, achieving consistent improvements over strong self-supervised and fully supervised audio-visual emotion recognition baselines.
- Demonstrated HQC-MAE's robustness under noisy test conditions and cross-corpus generalization, improving over the HiCMAE baseline by +11.62% UAR / +5.26% WAR on corrupted MSP-IMPROV inputs and by +3.50% UAR / +4.05% WAR when training on MSP-IMPROV and testing on CREMA-D.

Accounting AI Research Lab, CMU

Aug. 2025 – Present

Research Assistant, advised by Prof. Jinghong Liang

Predicting Future Earnings Changes Using Graph Representations

- Designed a novel graph representation of quarterly firm-level financials, encoding accounting statements as nodes (financial accounts) and edges (inter-account flows).
- Developed a temporal graph neural network (GNN) combining graph message passing with a sequence encoder that learns from historical sequences of quarterly financial graphs to predict future earnings direction, outperforming all existing baseline models across all forecast periods.
- Demonstrated the value of graph structure by showing that the full GNN consistently outperformed corresponding node-only variants across multiple ML methods.
- Plan to submit this work to the IEEE Special Session on Machine Learning on Big Data (MLBD 2026).

Coordinated Science Laboratory, UIUC

Jun. 2023 – Aug. 2023

Machine Learning Intern, advised by Prof. Dušan Stipanović & Dr. Srikar Annamraju

Machine Learning Algorithms for Healthcare Purposes

- Built an AI-based virtual therapist for post-stroke rehabilitation using time series models (LSTM, Informer) integrated with a Geomagic Touch haptic device.
- Developed an end-to-end data collection pipeline capturing multivariate time-series data, including 3D kinematics (position, orientation) and force vectors from therapist-guided rehabilitation sessions.
- Trained two models based on LSTM and Informer to predict 3D motion trajectories and assistive forces; achieved 98% accuracy with <10ms inference latency.
- Deployed production-ready models with ONNX Runtime on a dedicated server and integrated real-time inference into a Unity app, enabling accessible, personalized stroke-rehab therapy.

INTERSHIP EXPERIENCE

Artificial Intelligence for Artistic Institution Focused on Contemporary Art

Institution Lab Inc.

AI Research Intern

Supervisor: Tom Cebic (Director of Initiatives)

May 2025 – Aug. 2025

- Designed and built an Emotionally Aware AI Companion for contemporary-art exhibitions.
- Led the creation of a multimodal dataset of digital artworks; developed a web app for uploading digital artworks (Flask backend; HTML/CSS/JS + Bootstrap 5 frontend) and deployed it on AWS.
- Fine-tuned VideoLLaMA-3 with a semantic-affective fusion layer on our created dataset; implemented a graph-aware retrieval-augmented generation (GraphRAG) pipeline, achieving >90% test accuracy.
- Deployed the fine-tuned model on AWS and launched a real-time chat web app based on OpenWebUI.

Developing an AI Chatbot Agent for Carnegie Mellon University Libraries

AI Chatbot Agent Developer

Supervisor: Haoyong Lan (CMU Librarian)

Jan. 2025 – May. 2026

- Built a production chatbot with Google Vertex AI to answer library-related questions.
- Fine-tuned Gemini on two years of conversational data; built the conversational agent with Vertex AI Agent Builder/Playbooks and an OpenAI tool suite; implemented RAG pipeline that pulls real-time content from library.cmu.edu, reducing hallucinations by > 90%.
- Led the web frontend (HTML/CSS/JS); built a Google BigQuery-backed feedback capture and analytics app to guide new tools and knowledge base expansion.
- Serving 1K+ weekly sessions; live via the chat icon on the lower-right of library.cmu.edu.

SELECTED COURSE PROJECTS

Contrastive Gated Fusion for Multilingual Speaker Verification

11-775 Large-Scale Multi-Media Analysis (2025 Fall)

Supervisor: Prof. Rita Singh

CMU

- Developed an audio-visual speaker verification system for the FAME 2026 multilingual face-voice association challenge using the MAV-Celeb v3 English-German dataset, [3].
- Built on frozen DINOv2 (vision) and WavLM (speech) encoders with lightweight MLP adapters and a gated attention fusion module, enabling efficient end-to-end training with cross-entropy and contrastive losses.
- Tuned the training pipeline (data augmentation, feature pre-extraction, large-batch contrastive training) to reduce overfitting and improve the robustness under heard vs. unheard language conditions.
- Achieved an overall Equal Error Rate of 33.2% and ranked 4th in the FAME 2026 Challenge leaderboard.

Optimizing Secure Code Generation via LLMs and Constrained Decoding Methods

14-795 AI Applications in Information Security (2025 Spring)

Supervisor: Prof. David Varodayan

CMU

- Applied constrained beam search to enforce security rules at decode time, improving LLM code generation robustness without fine-tuning.
- Benchmarked six LLMs (Llama, Mistral, DeepSeek, StarCoder, Nxcoder) across decoding strategies (constrained beam, nucleus, vanilla beam) to identify best model-algorithm pairs.
- Designed custom constraints aligned to CWE patterns and evaluated with CodeQL static analysis; increased secure code pass rate by 15% (sec-pass@1: 60.85% → 69.90%).

Methods and Applications: Fine-Tuning LLMs for JSON-Structured Function Calls

11-667 Large Language Models (2025 Spring) Supervisor: Prof. Daphne Ippolito & Prof. Chenyan Xiong CMU

- Fine-tuned LLMs to generate structured JSON for autonomous API/function calls.
- Implemented LoRA fine-tuning on StarCoder2-7b, Mistral-7B and Llama-3-8B using Hermes Function Calling dataset (4,170 samples); achieved 97.1% accuracy.
- Benchmarked LoRA fine-tuning vs. in-context learning across multiple models (Llama-3.1, Mistral-7B) to diagnose failure modes in complex reasoning (productivity, puzzle-solving) and guide improvements.

Fantasy Football App

CS 409 The Art of Web Programming (2023 Fall)

Supervisor: Prof. Ranjitha Kumar

UIUC

- Built a full-stack web app (React, Node.js, MongoDB) for exploring football player stats.
- Implemented a Player List view with filtering, a Detailed Stats view, and roster management functionality.
- Added a secure user authentication system with session management and role-based access control.
- Designed a responsive, mobile-first UI for a seamless cross-device experience.

Vision-Assisted UAV General Seed Launcher System

CS 444 Deep Learning for Computer Vision (2023 Spring)

Supervisor: Prof. Svetlana Lazebnik

UIUC

- Built a vision-assisted UAV seed launcher; integrated a YOLO detector for real-time target detection, obstacle avoidance, and drop-zone selection.
- Designed a modular, 3D-printed seed launcher with a gun-style, bullet-loaded capsule firing mechanism; achieved 85%+ ground-target accuracy and compatibility with commercial UAVs.
- Designed and 3D-printed biodegradable seed capsules preloaded with seed, water, soil, and nutrients to boost post-drop viability of seeds, achieving 50%+ germination in field trials.

SKILLS

Technical Skills:

- Programming Languages: C, C++, Python, Java, CSS, HTML, JavaScript, x86-64 Assembly
- AI/ML Skills: Deep Learning, Computer Vision, Natural Language Processing, Signal Processing, LLM Fine-tuning & Deployment, Multimodal Machine Learning, Representation Learning, Vector Quantization, Self-supervised Learning, RAG, Graph Neural Networks, Trustworthy AI, Reinforcement Learning
- Frameworks and Tools in ML: PyTorch, TensorFlow, Spark, Apache Hadoop, Hugging Face, FlashAttention, scikit-learn, OpenCV, ONNX Runtime, TensorRT, Docker/Kubernetes, Git, MLflow/W&B
- Database & Platforms: PostgreSQL, MySQL, MongoDB, Neo4j, BigQuery, Spark, Hadoop, Kafka
- Cloud: GCP (Vertex AI, Cloud Run, IAM); AWS (EC2, S3, ECR/IAM)
- Web: Flask, Django, Node.js/Express, React, Nginx, Unicorn/Uvicorn
- Game Development: Unreal Engine 5, Unity

Language Skills: English (Proficient), Mandarin (Native)

Extracurricular Activities: CMU Table Tennis Club