Zeyuan Yang

(+1) 413-510-8350 YangZeyuan2020@gmail.com miicheyang.github.io

EDUCATION

University of Massachusetts, Amherst, USA . . . . . *Aug 2024 - Present*
Ph.D. student in Computer Science
Advisors: Chuang Gan, Hao Zhang
Research Interests: Embodied Agents, Multi-modal Foundation Models, Post-Training

Tsinghua University, Beijing, China . . . . . *Sep 2021 - July 2024*
M.S. in Computer Science
Advisor: Yang Liu
Research Interests: Natural Language Processing, Continual Learning

Tsinghua University, Beijing, China . . . . . *Sep 2017 - July 2021*
B.S. in Information Management, minor in Data Science
Major GPA: 3.63/4.0 (Rank 4/29)
Thesis Advisor: Qing Ye

EMPLOYMENT

Amazon's Frontier AI & Robotics . . . . . *May 2026 - Now*
Research Intern, Supervisors: Yunsheng Tian, Chen Feng, Rocky Duan San Francisco, USA
Navigation Foundation Agents

Samsung Research America . . . . . *May 2025 - Aug 2025*
Research Intern, Supervisors: Luowei Zhou, Deen Dayal Mohan Mountain View, USA
Multi-modal Retrieval and Foundation Models

Zhiyuan Innovation Technology Co., Ltd. . . . . . *Jan 2024 - Aug 2024*
Research Intern, Agibot Lab, Supervisor: Hao Dong Beijing, China
Humanoid Robotics

ONGOING PROJECTS

Agentic Navigation Harness for Any Model Any Task

Building **a model- and task-agnostic harness** for **tool-using multimodal agents** in **embodied navigation**. On this shared tool base, we systematically study which tools each model and task actually needs, and how to intuitively select **the optimal tool subset for certain scenarios.** *Ready to submit to CVPR 2027.*

Robotics Navigation *Multimodal Agents* *Tool-using*

A Multimodal Cascade Duplex System For Real-Time Interaction

Developing **a real-time multimodal duplex system** that **responds to audio and visual signals** during ongoing interaction. It substantially outperforms open-source and close-source full-duplex, turn-based, and agentic duplex systems across VideoFDB and Big Bench Audio. **Demo Available** (password: `paloalto`).

Multimodal Duplex *Real-Time Interaction* *Audio & Speech*

Embodied Duplex Benchmark for Real-time Emotional Support

Developing a benchmark focused on **duplex, real-time interaction** across **30 home-use scenarios** over **50**

characters. The environment continues to change while an agent reasons, requiring **rapid responses**, **active perception**, and **interaction with both the user and surroundings** to complete tasks. In initial test, nearly all frontier APIs achieve **under 20% success**, compared with **92% for humans**. *Plan to release on late Nov.*

Multimodal Duplex *Benchmark* *Embodied Agents*

Self-Improvement Creative Video Generation Agent

Current models can produce **high-fidelity** videos but often lack **expressiveness**. Even frontier VLMs fail to produce reliable critics. We build a **large-scale dataset** and annotation pipeline for creative video critique, and our fine-tuned Qwen3.8-27B matches human judgments better than GPT-6 Astra. Currently, we are developing an **VLM-as-a-judge self-improvement pipeline** to improve expressiveness while preserving core video content.

Video Generation *Self-Improvement*

Social World Model: Open-Ended Social Media Response Prediction

Developing a **Social World Model** to **predict and generate plausible, open-ended responses to social media videos**. We build a **1,000-video benchmark** spanning **multiple platforms with real audience responses and evaluation metrics**. We introduce **a perspective-first framework** that covers substantially more real audience reactions than existing methods. *Ready to submit to CVPR 2027.*

World Model *Open-Ended Generation* *Video Understanding*

PUBLICATIONS (Google Scholar, * denotes equal contribution.)

[13] Perception or Reasoning? Diagnosing and Tackling Reinforcement Learning Bottlenecks in VLMs
Zeyuan Yang*, Zhuohao Chen*, Jiaben Chen, Yao Zhang, Maohao Shen, Chuang Gan
In submission, 2026

[12] Multimodal Thinking with Renderable Programs
Sunli Chen, Ding Zhong, Ziqiao Ma, Jiaxin Liu, **Zeyuan Yang**, Hao Zhang, Joyce Chai, Chuang Gan
In submission, 2026

[11] UniCanvas: A Diffusion-based Unified Model with Text-in-Image Joint Generation
Zeyuan Yang*, Hao-Wei Chen*, Yuncong Yang, Xueyang Yu, Haoyu Zhen, Ziqiao Ma, Maohao Shen, Chuang Gan
Best Paper at *CVPR A2A-MML Workshop*, 2026
In submission, 2026

[10] SyncWorld: Visual Calibration Enables World Models as Zero-Shot Simulators
Yuncong Yang, Zhengtao Han, Furkan Ozyurt, **Zeyuan Yang**, Han Yang, Junyi Cao, Haoyu Zhen, Yilun Du, Chuang Gan
NeurIPS Oral, 2026

[9] Machine Mental Imagery: Empower Multimodal Reasoning with Latent Visual Tokens
Zeyuan Yang*, Xueyang Yu*, Delin Chen, Maohao Shen, Chuang Gan
Best Paper at *ICCV KnowledgeMR Workshop*, 2025
CVPR, 2026

[8] VCA: Video Curious Agent for Long Video Understanding
Zeyuan Yang*, Delin Chen*, Xueyang Yu, Maohao Shen, Chuang Gan
ICCV, 2025

[7] UBSoft: A Simulation Platform for Robotic Skill Learning in Unbounded Soft Environments
Chunru Lin, Jugang Fan, Yian Wang, **Zeyuan Yang**, Zhehuan Chen, Lixing Fang, Tsun-Hsuan Wang, Zhou

Xian, Chuang Gan
CoRL, 2024

[6] Rethinking Long Context Generation from the Continual Learning Perspective
Zeyuan Yang, Fangzhou Xiong, Peng Li and Yang Liu
COLING, 2025

[5] Towards Unified Alignment Between Agents, Humans, and Environment
Zonghan Yang, An Liu, Zijun Liu, Kaiming Liu, Fangzhou Xiong, Yile Wang, **Zeyuan Yang**, Qingyuan Hu, Xinrui Chen, Zhenhe Zhang, Fuwen Luo, Zhicheng Guo, Peng Li, Yang Liu
ICML, 2024

[4] Reflective and Imaginative Language Agent for Zero-Shot Semantic Audio-Visual Navigation
Zeyuan Yang*, Jiageng Liu*, Peihao Chen, Anoop Cherian, Tim Marks, Jonathan Le Roux, Chuang Gan
CVPR, 2024

[3] Failures Pave the Way: Enhancing Large Language Models through Tuning-free Rule Accumulation
Zeyuan Yang, Peng Li, Yang Liu
EMNLP, 2023

[2] Restricted Orthogonal Gradient Projection for Continual Learning
Zeyuan Yang, Zonghan Yang, Yichen Liu, Peng Li, Yang Liu
AI Open, 2023

[1] Dynamic Multi-Branch Layers for On-Device Neural Machine Translation
Zhixing Tan, **Zeyuan Yang**, Meng Zhang, Qun Liu, Maosong Sun, Yang Liu
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022

[0] Exploring deep learning for air pollutant emission estimation
Lin Huang, Song Liu, **Zeyuan Yang**, Jia Xing, Jia Zhang, Jiang Bian, Siwei Li, Shovan Kumar Sahu, Shuxiao Wang, Tie-Yan Liu
Geoscientific Model Development , 2021

TEACHING EXPERIENCE

Information Systems, University of Massachusetts, Amherst . . . . . Spring 2026
Teaching Assistant, Instructor: Trek Palmer

Foundations of Programming, University of Massachusetts, Amherst . . . . . Spring 2025
Teaching Assistant, Instructor: Cole Adam Reilly

Objected Oriented Programming, University of Massachusetts, Amherst . . . . . Fall 2024
Teaching Assistant, Instructor: Jaime Davila