

# Wenhao Chai

---

✉ [wenhao.chai@princeton.edu](mailto:wenhao.chai@princeton.edu)  
🏠 [homepage](#)  
🔍 [google scholar](#)  
🐙 [github.com/rese1f](https://github.com/rese1f)  
🐦 [x.com/wenhaochai](https://x.com/wenhaochai)  
🌐 [linkedin.com/in/wenhao-chai](https://linkedin.com/in/wenhao-chai)

(206) 349-8459  
35 Olden St.  
Princeton, NJ  
United States, 08544  
Dept. of Computer Science  
Princeton University

## Research Overview

I work toward long-context multimodal AI systems that can both understand and generate rich, interleaved streams of text, video, images, audio, code, and actions. My research tackles (1) encoding: efficient long-context perception, context-weight conflicts, and continual adaptation at deployment; and (2) decoding: end-to-end multimodal generation beyond diffusion-only visual pipelines, including better visual tokenization, training, and RL for multimodal outputs. I also design benchmarks and evaluation methods for long-context and video generation to make progress in this direction measurable and reliable.

## Education

<b>Ph.D.</b> CS 2025-2030 (exp.)	Princeton University Advisor: Zhuang Liu Department of Computer Science
<b>M.S.</b> EE 2023-2025	University of Washington Advisor: Jenq-Neng Hwang Thesis: Large Multimodal Models for Video Captioning
<b>Visiting</b> Spring/Summer 2022	University of Illinois Urbana-Champaign National Center for Supercomputing Applications
<b>B.S.</b> 2019-2023	Zhejiang University Advisor: Gaoang Wang
<b>High School</b> 2016-2019	Hangzhou No. 2 High School Hangzhou, Zhejiang

## Employment

<b>Research Intern</b> Summer 2024	Pika Labs Working on Video Understanding Mentor: Christopher D. Manning
<b>Research Assistant</b> 2023-2024	University of Washington Information Processing Lab PI: Jenq-Neng Hwang
<b>Research Intern</b> Spring/Summer 2023	Microsoft Research Asia Working on Video Generation Mentor: Xun Guo

## Selected Publications

\* equal contribution. † project lead.

- C13 Wenhao Chai, Enxin Song, Yilun Du, Chenlin Meng, Vashisht Madhavan, Omer Bar-Tal, Jenq-Neng Hwang, Saining Xie, and Christopher D. Manning. **AuroraCap: Efficient, Performant Video Detailed Captioning and a New Benchmark**. In *International Conference on Learning Representations (ICLR)*. 2025.
- C12 Ruizhe Chen\*, Xiaotian Zhang\*, Meng Luo\*, Wenhao Chai\*, and Zuozhu Liu. **PAD: Personalized Alignment at Decoding-time**. In *International Conference on Learning Representations (ICLR)*. 2025.
- C11 Zihan Zheng\*†, Zerui Cheng\*, Zeyu Shen\*, Shang Zhou\*, Kaiyuan Liu\*, Hansen He\*, Dongruixuan Li, Stanley Wei, Hangyi Hao, Jianzhu Yao, Peiyao Sheng, Zixuan Wang, Wenhao Chai†, Aleksandra Korolova, Peter Henderson, Sanjeev Arora, Pramod Viswanath, Jingbo Shang, and Saining Xie. **Live-CodeBench Pro: How Do Olympiad Medalists Judge LLMs in Competitive Programming?**. In *Neural Information Processing Systems (NeurIPS) DB Track*. 2025.
- C10 Xiangyu Zhao, Peiyuan Zhang, Kexian Tang, Xiaorong Zhu, Hao Li, Wenhao Chai, Zicheng Zhang, Renqiu Xia, Guangtao Zhai, Junchi Yan, Hua Yang, Xue Yang, and Haodong Duan. **Envisioning beyond the Pixels: Benchmarking Reasoning-informed Visual Editing**. In *Neural Information Processing Systems (NeurIPS) DB Track*. **Oral**. 2025.
- J1 Enxin Song\*, Wenhao Chai\*†, Tian Ye, Jenq-Neng Hwang, Xi Li, and Gaoang Wang. **MovieChat+: Question-aware Sparse Memory for Long Video Question Answering**. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. 2025.
- C9 Hsiang-Wei Huang, Fu-Chen Chen, Wenhao Chai, Che-Chun Su, Lu Xia, Sanghun Jung, Cheng-Yen Yang, Jenq-Neng Hwang, Min Sun, and Cheng-Hao Kuo. **Zero-shot 3D Question Answering via Voxel-based Dynamic Token Compression**. In *Computer Vision and Pattern Recognition (CVPR)*. 2025.
- C8 Jialuo Li, Wenhao Chai, Xingyu Fu, Haiyang Xu, and Saining Xie. **Science-T2I: Addressing Scientific Illusions in Image Synthesis**. In *Computer Vision and Pattern Recognition (CVPR)*. 2025.
- C7 Weili Xu, Enxin Song, Wenhao Chai†, Tian Ye, and Gaoang Wang. **AuroraLong: Bringing RNNs Back to Efficient Open-Ended Video Understanding**. In *International Conference on Computer Vision (ICCV)*. 2025.
- C6 Ruizhe Chen, Wenhao Chai, Zhifei Yang, Xiaotian Zhang, Joey Tianyi Zhou, Tony Quek, Soujanya Poria, and Zuozhu Liu. **DiffPO: Diffusion-styled Preference Optimization for Efficient Inference-Time Alignment of Large Language Models**. In *Annual Meeting of the Association for Computational Linguistics (ACL)*. **Main**. 2025.
- C5 Hsiang-Wei Huang, Wenhao Chai, Kuang-Ming Chen, Cheng-Yen Yang, Jenq-Neng Hwang. **ToSA: Token Merging with Spatial Awareness**. In *International Conference on Intelligent Robots and Systems (IROS)*. 2025.
- C4 Zhonghan Zhao\*, Wenhao Chai\*†, Xuan Wang\*, Boyi Li, Shengyu Hao, Shidong Cao, Tian Ye, Jenq-Neng Hwang, and Gaoang Wang. **See and Think: Embodied Agent in Virtual Environment**. In *European Conference on Computer Vision (ECCV)*. 2024.
- C3 Enxin Song\*, Wenhao Chai\*†, Guanhong Wang, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haozhe Chi, Xun Guo, Tian Ye, Yanting Zhang, Yan Lu, Jenq-Neng Hwang, and Gaoang Wang. **MovieChat: From Dense Token to Sparse Memory for Long Video Understanding**. In *Computer Vision and Pattern Recognition (CVPR)*. 2024.
- C2 Wenhao Chai, Zhongyu Jiang, Jenq-Neng Hwang, and Gaoang Wang. **Global Adaptation Meets Local Generalization: Unsupervised Domain Adaptation for 3D Human Pose Estimation**. In *International Conference on Computer Vision (ICCV)*. 2023.
- C1 Wenhao Chai, Xun Guo, Gaoang Wang, and Yan Lu. **StableVideo: Text-driven Consistency-aware Diffusion Video Editing**. In *International Conference on Computer Vision (ICCV)*. 2023.

## Awards and Honors

Outstanding Paper at ICCV 2025 Knowledge-Intensive Multimodal Reasoning Workshop	Oct 2025
Lambda AI Cloud Credits Grant Sponsorship	Sept 2025
Best Paper Award at IEEE MIPR 2025	Aug 2025
Featured in MIT Technology Review as lead author of LiveCodeBench Pro	Jun 2025
University of Washington Distinguished Thesis Award Nominee by ECE Department	May 2025

## Invited Talks

Better and Longer Video Understanding Slides	Oct 2025
Sky9 Fellowship	Sept 2025
Abaka AI and 2077AI	Aug 2025
Bitdeer AI	
Step-by-Step Construction of Agent Systems in Minecraft CAMEI-AI AgentX Seminar Host: Guohao Li	Apr 2024
Towards Universal Animal Perception in Vision Workshop on Imageomics at AAAI 2024, Vancouver, CA	Feb 2024

## Selected Open-Source Projects

Terminal Bench 1,000 Stars (Contributor) Support LiveCodeBench Pro for agents evaluation.	Aug 2025- Sept 2025
LMMs-Eval 3,100 Stars (Contributor) The Evaluation Suite of Large Multimodal Models.	Oct 2024- Dec 2024
SAMURAI 7,000 Stars (Core Contributor) Visual object tracking tool based on segment anything (SAM).	Nov 2024- Dec 2024
MovieChat 650 Stars (Main Contributor) Large multi-modal models evaluation for long video understanding.	Jun 2023- Jan 2025
StableVideo 1,400 Stars (Main Contributor) Diffusion-based video editing demo and interactive tool.	Mar 2023- Aug 2023

## Professional Service

### Workshop Organization

Multimodal Video Agent Workshop on Computer Vision and Pattern Recognition (CVPR)	Nashville, TN June 2025
Long-form Video Understanding Towards Multimodal AI Assistant and Copilot Workshop on Computer Vision and Pattern Recognition (CVPR)	Seattle, WA June 2024

### Conference and Journal Refereeing

Neural Information Processing Systems (NeurIPS)	2024-2025
International Conference in Learning Representations (ICLR)	2025-2026
International Conference in Machine Learning (ICML)	2024-2025
Computer Vision and Pattern Recognition (CVPR)	2024-2026
International Conference on Computer Vision (ICCV)	2025
European Conference on Computer Vision (ECCV)	2024
Annual Meeting of the Association for Computational Linguistics (ACL) by ARR	2025-2026
Empirical Methods in Natural Language Processing (EMNLP) by ARR	2025
Conference on Language Modeling (COLM)	2025
Association for the Advancement of Artificial Intelligence (AAAI)	2025-2026
International Conference on Artificial Intelligence and Statistics (AISTATS)	2025-2026
Winter Conference on Applications of Computer Vision (WACV)	2025
ACM International Conference on Multimedia (ACM MM)	2024
Pattern Recognition and Computer Vision (PRCV)	2023
International Journal of Computer Vision (IJCV)	
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	

### Workshop Refereeing

Efficient Reasoning	NeurIPS 2025
Frontiers in Probabilistic Inference: Learning meets Sampling	NeurIPS 2025
Structured Probabilistic Inference & Generative Modeling	NeurIPS 2025
Bridging Language, Agent, and World Models	NeurIPS 2025
SPACE in Vision, Language, and Embodied AI	NeurIPS 2025
Scaling Environments for Agents	NeurIPS 2025
Efficient Systems for Foundation Models	ICML 2025
Frontiers in Probabilistic Inference: Learning meets Sampling	ICLR 2025
Bidirectional Human-AI Alignment	ICLR 2025
Foundation Models in the Wild	ICLR 2025
Multi-Agent AI in the Real World	AAAI 2025

## Community Service

Constructive Feedback for <a href="#">alphaXiv</a> and <a href="#">Paper Copilot</a>	2024-Present
Host Discord Server (over 300 people) for arXiv Daily Paper Sharing.	2023-Present
Research Experience Sharing Session for Undergraduates at Zhejiang University	2023
Co-Director of the Publicity Dept. of the Student Union at Zhejiang University	2019-2022