Wenhao Chai

wenhao.chai@princeton.edu
homepage
google scholar
github.com/rese1f
x.com/wenhaocha1
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 learning; and (2) decoding: end-to-end multimodal generation beyond diffusion 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

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

Employment

Student Researcher Summer 2026	Google DeepMind Working on Multimodal Stuffs Mentor: Dr. Mu Cai
Research Intern Summer 2024	Pika Labs Working on Video Understanding Mentor: Prof. Christopher D. Manning
Research Assistant 2023-2024	University of Washington Information Processing Lab PI: Prof. Jenq-Neng Hwang
Research Intern Spring/Summer 2023	Microsoft Research Asia Working on Video Generation and Understanding Mentor: Dr. Xun Guo

Selected Publications

* equal contribution. † project lead.

C21 Yida Yin, Harish Krishnakumar, Chung Peng Lee, Boya Zeng, **Wenhao Chai**, Shengbang Tong, Wenhu Chen, Hu Xu, Xingyu Fu, Gabriel Herbert Sarch, Aleksandra Korolova, and Zhuang Liu. WorldBench: A Challenging and Visually Diverse Multimodal Benchmark. In *Computer Vision and Pattern Recognition* (*CVPR*). 2026.

C20 Jusheng Zhang, Xiaoyang Guo, Kaitong Cai, Qinhan Lv, Yijia Fan, **Wenhao Chai**, Jian Wang, and Keze Wang. Hybridtoken-vlm: Hybrid Token Compression for Vision-Language Models. In *Computer Vision and Pattern Recognition* (*CVPR*). 2026.

C19 Enxin Song, **Wenhao Chai**, Shusheng Yang, Ethan Armand, Xiaojun Shan, Haiyang Xu, Jianwen Xie, and Zhuowen Tu. VideoNSA: Native Sparse Attention Scales Video Understanding. In *International Conference on Learning Representations* (*ICLR*). 2026.

C18 Shang Zhou, Zihan Zheng, Kaiyuan Liu, Zeyu Shen, Zerui Cheng, Zexing Chen, Hansen He, Jianzhu Yao, Huanzhi Mao, Qiuyang Mang, Tianfu Fu, Beichen Li, Dongruixuan Li, **Wenhao Chai**†, Zhuang Liu, Aleksandra Korolova, Peter Henderson, Natasha Jaques, Pramod Viswanath, Saining Xie, and Jingbo Shang. AutoCode: LLMs as Problem Setters for Competitive Programming. In *International Conference on Learning Representations* (*ICLR*). 2026.

C17 Terminal-Bench Team. Terminal-Bench: Benchmarking Agents on Hard, Realistic Tasks in Command Line Interfaces. In *International Conference on Learning Representations* (*ICLR*). 2026.

C16 Qingyu Shi, Jinbin Bai, Zhuoran Zhao, **Wenhao Chai**, Kaidong Yu, Jianzong Wu, Shuangyong Song, Yunhai Tong, Xiangtai Li, Xuelong Li, and Shuicheng Yan. Beyond Text-to-Image: Liberating Generation with a Unified Discrete Diffusion Model. In *International Conference on Learning Representations* (*ICLR*). 2026.

C15 Xiangyu Zhao, Junming Lin, Tianhao Liang, Yifan Zhou, **Wenhao Chai**, Yuzhe Gu, Weiyun Wang, Kai Chen, Gen Luo, Wenwei Zhang, Junchi Yan, Hua Yang, Haodong Duan, and Xue Yang. MM-HELIX: Boosting Multimodal Long-Chain Reflective Reasoning with Holistic Platform and Adaptive Hybrid Policy Optimization. In *International Conference on Learning Representations* (*ICLR*). 2026.

C14 **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.

C13 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.

C12 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. LiveCodeBench Pro: How Do Olympiad Medalists Judge LLMs in Competitive Programming?. In *Neural Information Processing Systems* (*NeurIPS*) *DB Track*. 2025.

C11 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.

C10 Jusheng Zhang, Yijia Fan, Wenjun Lin, Ruiqi Chen, Haoyi Jiang, **Wenhao Chai**, Jian Wang, and Keze Wang. GAM-Agent: Game-Theoretic and Uncertainty-Aware Collaboration for Complex Visual Reasoning. In *Neural Information Processing Systems* (*NeurIPS*). 2025.

J2 Cheng-Yen Yang, Hsiang-Wei Huang, **Wenhao Chai**, Zhongyu Jiang, and Jenq-Neng Hwang. SAMURAI: Adapting Segment Anything Model for Zero-shot Visual Tracking with Motion-aware Memory. In *IEEE Transactions on Image Processing* (*TIP*). 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*). *Oral*. 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
Sky9 Fellowship	Oct 2025
Abaka AI and 2077AI	Sept 2025
Bitdeer AI	Aug 2025
Step-by-Step Construction of Agent Systems in Minecraft	
CAMEl-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 1.4K Stars	Aug 2025-
(Contributor) Support LiveCodeBench Pro for agents evaluation.	Sept 2025
LMMs-Eval 3.4K Stars	Oct 2024-
(Contributor) The Evaluation Suite of Large Multimodal Models.	Dec 2024
SAMURAI 7.0K Stars	Nov 2024-
Visual object tracking tool based on segment anything (SAM).	Dec 2024
MovieChat 650 Stars	Jun 2023-
Large multi-modal models evaluation for long video understanding.	Jan 2025
StableVideo 1.4K Stars	Mar 2023-
Diffusion-based video editing demo and interactive tool.	Aug 2023

Research Mentoring

LiveCodeBench Pro Team Members are at Princeton, Berkeley, NYU, UCSD, and UW. Topic: Large Language Models for Competitive Programming LiveCodeBench Pro: How Do Olympiad Medalists Judge LLMs? AutoCode: LLMs as Problem Setters for Competitive Programming FrontierCS: Evolving Challenges for Evolving Intelligence	Feb 2025- Present NeurIPS 2025 ICLR 2026 arXiv 2025
Amish Sethi B.S. at University of Pennsylvania ⇒ Ph.D. at Harvard Topic: Visual Generative Models Do Diffusion Models Learn to Generalize Basic Visual Skills?	Apr 2025- Nov 2025- In Submission
Kunjun Li B.S. at National University of Singapore ⇒ Ph.D. at UC Berkeley Topic: Efficient Models On the Role of Network Pruning in Diffusion Transformers	Apr 2025- Nov 2025 In Submission
Weili Xu B.S. at University of Illinois Urbana-Champaign & Zhejiang University Topic: Efficient Long Context Modeling AuroraLong: Bringing RNNs Back to Efficient Open-Ended Video Understanding	Oct 2024- Jun 2025 ICCV 2025
Hsiang-Wei Huang Ph.D. at University of Washington Topic: Spatial Understanding Zero-shot 3D Question Answering via Voxel-based Dynamic Token Compression ToSA: Token Merging with Spatial Awareness Reasoning Matters for 3D Visual Grounding	Jul 2024- Jun 2025 CVPR 2025 IROS 2025 CVPRW 2025
Ruizhe Chen Ph.D. at Zhejiang University Topic: (Multimodal) Large Language Models PAD: Personalized Alignment at Decoding-Time DiffPO: Diffusion-styled Preference Optimization for Inference Time Alignment Can "Think with Videos" Boost Video Understanding?	Jul 2024- Present ICLR 2025 ACL 2025 In Submission
Enxin Song M.S. at Zhejiang University Topic: Video Understanding MovieChat: From Dense Token to Sparse Memory for Long Video Understanding MovieChat+: Question-aware Sparse Memory for Long Video Question Answering VideoMMLU: A Massive Multi-Discipline Lecture Understanding Benchmark VideoNSA: Native Sparse Attention Scales Video Understanding	Jul 2023- Present CVPR 2024 TPAMI 2025 ICCVW 2025 ICLR 2026
Zhonghan Zhao Ph.D. at Zhejiang University ⇒ Intern at Shanghai AI Lab Topic: Embodied Agent See and Think: Embodied Agent in Virtual Environment Hierarchical Auto-Organizing System for Open-Ended Multi-Agent Navigation Steve Series: Step-by-step Construction of Agent Systems in Minecraft Do We Really Need a Complex Agent System? Distill Agent into a Single Model	Jul 2023- Jul 2024 ECCV 2024 ICLRW 2024 CVPRW 2024 arXiv 2024
Meiqi Sun M.S. at Zhejiang University ⇒ Alibaba Group Topic: Pose Estimation UniAP: Towards Universal Animal Perception in Vision via Few-shot Learning	Jul 2023- Feb 2024 AAAI 2024

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 Program Committee

Neural Information Processing Systems (NeurIPS)	2024-2025
International Conference in Learning Representations (ICLR)	2025-2026
International Conference in Machine Learning (ICML)	2024-2026
Computer Vision and Pattern Recognition (CVPR)	2024-2026
International Conference on Computer Vision (ICCV)	2025
European Conference on Computer Vision (ECCV)	2024-2026
Annual Meeting of the Association for Computational Linguistics (ACL) in ARR	2025-2026
Empirical Methods in Natural Language Processing (EMNLP) in ARR	2025
Conference on Language Modeling (COLM)	2025-2026
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

Workshop Program Committee

AI with Recursive Self-Improvement	ICLR 2026
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

Journal Program Committee

International Journal of Computer Vision (IJCV)	2023-2024
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	2023-2024

Community Service

Interview Special Host at Z Potentials.	2026-Present
Constructive Feedback for alphaXiv and Paper Copilot.	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