

## Education

- Aug, 2023–  
Present **Ph.D., Computer Science**, *Department of Computer Science*, University of Illinois Urbana-Champaign, USA.  
Advisor: **James M. Rehg**
- Sep, 2020–  
Sep, 2022 **Master of Science, Computer Science**, *Courant Institute of Mathematical Sciences*, New York University, USA.  
Advisor: **Elena Sizikova, Megan Coffee**
- Sep, 2016–  
Jun, 2020 **Bachelor of Science, Chemistry and Data Science**, *School of Data Science*, Fudan University, China.

## Work Experience

### PediaMed AI

- June, 2025 –  
Present **Co-Founder & Tech Lead**.  
Manage and organize the AI Lab. Lead the research and development of GenAI tools for neurodevelopmental disorders assessment. (Lead 6 Tech Reports: ChildGait (ECCV 2026), HippoVLM (CVPR 2026), CogSense, ChildAgentEval, ChildPose, ChronoVision)

### Google AR

- Sep, 2025 –  
July, 2026 **Student Researcher**, Advisor: **Yinda Zhang, Cady Xu, Zhongyi Zhou**.  
Research for Multimodal Human Behavior Modeling for Agentic AI system and Augmented Reality (AR).

### NEC Laboratories America, Inc.

- May, 2024 –  
Aug, 2025 **Research Intern**, Advisor: **Manmohan Chandraker, Abhishek Aich**.  
Research for Reinforcement Learning post-training for video vision language models.

### SambaNova Systems, Inc.

- Oct, 2022 –  
July, 2023 **Software Engineer (Machine Learning)**, Manager: **Tejas Nama**, Computer Vision Team.  
Research and develop the latest computer vision algorithm for industry applications including Multimodal Vision Transformer, Stable Diffusion, LLMs into reconfigurable dataflow unit (RDU).

### Tencent Holdings Ltd.

- May, 2021 –  
Aug, 2021 **Research Intern**, Manager: **Chao Zheng**, Tencent T Lab.  
Developed a new algorithm for automatic lane markings and lane change points detection in HD maps. Our project won the **Innovative Applications of Artificial Intelligence Award** in the IAAI track.

### Dow Chemical Company

- Sep, 2020 –  
Jan, 2021 **Software Engineer Intern**.  
Managed Dow's database and file system via SAP server, GlobalScape, and Microsoft SQL Server.

## Publications

### In Conference Proceedings

- 2026 Xu Cao, Houze Yang, Vipin Gunda, Zhongyi Zhou, Tianyu Xu, Adarsh Kowdle, Inki Kim, and James M. Rehg. Gaze target estimation anywhere with concepts. *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2026.

- 2026 Xu Cao, Kaizhao Liang, Kuei-Da Liao, Tianren Gao, Zheng Chen, Jintai Chen, Zhiguang Ding, Jianguo Cao, James M Rehg, and Jimeng Sun. Video-based disease progression simulation. *Conference on Health, Inference, and Learning (CHIL)* (**Best Paper Award**), 2026.
- 2026 Xu Cao. Enhancing video vision language model with hippocampal sensing. *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2026.
- 2026 Yifan Shen, Boyi Li, Meihuan Huang, Yuanzhe Liu, Xu Cao, Jinyang Jin, Zhengyuan Li, Anglin Liu, Junho Kim, Jingyuan Zhu, Fangzhou Lan, Jianguo Cao, Jintai Chen, Ismini Lourentzou, and James M. Rehg. Decoding children's gait behavior. *European Conference on Computer Vision*, 2026.
- 2026 Yuanzhe Liu, Jingyuan Zhu, Yuchen Mo, Gen Li, Xu Cao, Jin Jin, Yifan Shen, Zhengyuan Li, Tianjiao Yu, Wenzhen Yuan, et al. Palm: Progress-aware policy learning via affordance reasoning for long-horizon robotic manipulation. *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2026.
- 2026 Wenqi Jia, Bolin Lai, Xu Cao, Miao Liu, Danfei Xu, and James M Rehg. Learning predictive visuomotor coordination. *Findings of IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2026.
- 2025 Xu Cao, Pranav Virupaksha, Sangmin Lee, Bolin Lai, Wenqi Jia, Jintai Chen, and Jim Rehg. Toward human deictic gesture target estimation. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- 2025 Xu Cao, Pranav Virupaksha, Wenqi Jia, Bolin Lai, Fiona Ryan, Sangmin Lee, and Jim Rehg. Socialgesture: Delving into multi-person gesture understanding. *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2025.
- 2025 Xu Cao, Yifan Shen, Bolin Lai, Wenqian Ye, Yunsheng Ma, Joerg Heintz, Jintai Chen, Meihuan Huang, Jianguo Cao, and James M Rehg. What is the visual cognition gap between humans and multimodal llms? *Conference on Language Modeling (COLM)*, 2025.
- 2025 Yifan Shen, Yuanzhe Liu, Jingyuan Zhu, Xu Cao, Xiaofeng Zhang, Yixiao He, Wenming Ye, James Matthew Rehg, and Ismini Lourentzou. Fine-grained preference optimization improves spatial reasoning in vlms. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- 2025 Yanwen Fang, Wenqi Jia, Xu Cao, Peng-Tao Jiang, Guodong Li, and Jintai Chen. Proxy-bridged game transformer for multi-person highly interactive extreme motion prediction. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.
- 2024 Xu Cao, Tong Zhou, Yunsheng Ma, Wenqian Ye, Can Cui, Kun Tang, Zhipeng Cao, Kaizhao Liang, Ziran Wang, James M Rehg, et al. Maplm: A real-world large-scale vision-language benchmark for map and traffic scene understanding. In *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2024.
- 2024 Xu Cao, Wenqian Ye, Kenny Moise, and Megan Coffee. Mpoxvlm: A vision-language model for diagnosing skin lesions from mpox virus infection. In *Machine Learning for Health (ML4H)*, 2024.
- 2024 Yunsheng Ma, Xu Cao, Wenqian Ye, Can Cui, Mei Kai, and Ziran Wang. Learning autonomous driving tasks via human feedbacks with large language models. In *Findings of the Association for Computational Linguistics (EMNLP)*, 2024.
- 2024 Yunsheng Ma, Juanwu Lu, Can Cui, Sicheng Zhao, Xu Cao, Wenqian Ye, and Ziran Wang. Macp: Efficient model adaptation for cooperative perception. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.
- 2024 Yunsheng Ma, Can Cui, Xu Cao, Wenqian Ye, Peiran Liu, Amr Abdelraouf, Rohit Gupta, Kyungtae Han, Aniket Bera, Jim Rehg, and Ziran Wang. Lampilot: An open benchmark dataset for autonomous driving with language model programs. In *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2024.

- 2024 Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, Yang Zhou, Kaizhao Liang, Jintai Chen, Juanwu Lu, Zichong Yang, Kuei-Da Liao, et al. A survey on multimodal large language models for autonomous driving. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.
- 2024 Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, and Ziran Wang. Drive as you speak: Enabling human-like interaction with large language models in autonomous vehicles. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.
- 2023 Wenqian Ye, Yunsheng Ma, Xu Cao, and Kun Tang. Mitigating transformer overconfidence via lipschitz regularization. In *Uncertainty in Artificial Intelligence (UAI)*, 2023.
- 2023 Xu Cao, Wenqian Ye, Elena Sizikova, Xue Bai, Megan Coffee, Hongwu Zeng, and Jianguo Cao. Vitasd: Robust vision transformer baselines for autism spectrum disorder facial diagnosis. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.
- 2023 Kun Tang, Xu Cao, Zhipeng Cao, Tong Zhou, Erlong Li, Ao Liu, Shengtao Zou, Chang Liu, Shuqi Mei, Elena Sizikova, and Chao Zheng. Thma: Tencent hd map ai system for creating hd map annotations. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (Innovative Applications of Artificial Intelligence Award)*, 2023.
- 2023 Yunsheng Ma, Wenqian Ye, Xu Cao, Amr Abdelraouf, Kyungtae Han, Rohit Gupta, and Ziran Wang. Cemformer: Learning to predict driver intentions from in-cabin and external cameras via spatial-temporal transformers. In *The 26th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2023.
- 2022 Xu Cao, Xiaoye Li, Liya Ma, Yi Huang, Xuan Feng, Zening Chen, Hongwu Zeng, and Jianguo Cao. Aggpose: Deep aggregation vision transformer for infant pose estimation. In *The 31st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5045–5051, 2022.
- 2022 Elena Sizikova, Xu Cao, Ashia Lewis, Kenny Moise, and Megan Coffee. Improving computed tomography (ct) reconstruction via 3d shape induction. In *Machine Learning for Health (ML4H) Extended Abstract*, 2022.
- 2022 Elena Sizikova, Joshua Vendrow, Xu Cao, Rachel Grotheer, Jamie Haddock, Lara Kassab, Alona Kryshchenko, Thomas Merkh, RWMA Madushani, Kenny Moise, et al. Automatic infectious disease classification analysis with concept discovery. In *Machine Learning for Health (ML4H) Extended Abstract*, 2022.
- 2021 Xu Cao and Yanghao Lin. Caggnet: Crossing aggregation network for medical image segmentation. In *2020 25th International conference on pattern recognition (ICPR)*, pages 1744–1750, 2021.
- 2020 Xu Cao, Fangzhou Lan, Chi-Man Liu, Tak-Wah Lam, and Ruibang Luo. Chromseg: two-stage framework for overlapping chromosome segmentation and reconstruction. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2335–2342, 2020.

#### Journal Articles

- 2026 Yidong Zhou, Jintai Chen, Jinglei Cheng, Xu Cao, Yuanyuan Zhang, Gopal Karemore, Marinka Zitnik, Frederic T. Chong, Junyu Liu, Tianfan Fu, and Zhiding Liang. Quantum-machine-assisted drug discovery. *npj Drug Discovery*, 2026.
- 2025 Yihao Liu, Xu Cao, Tingting Chen, Yankai Jiang, Junjie You, Minghua Wu, Xiaosong Wang, Mengling Feng, Yaochu Jin, and Jintai Chen. From screens to scenes: A survey of embodied ai in healthcare. *Information Fusion*. Elsevier, 2025.
- 2024 Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, and Ziran Wang. Receive, reason, and react: Drive as you say, with large language models in autonomous vehicles. *IEEE Intelligent Transportation Systems Magazine*. IEEE, 2024.
- 2024 Jintai Chen, Yaojun Hu, Mingchen Cai, Yingzhou Lu, Yue Wang, Xu Cao, Miao Lin, Hongxia Xu, Jian Wu, Cao Xiao, et al. Trialbench: Multi-modal artificial intelligence-ready clinical trial datasets. *Nature Scientific Data*, 2024.

- 2023 Chao Zheng, Xu Cao, Kun Tang, Zhipeng Cao, Elena Sizikova, Tong Zhou, Erlong Li, Ao Liu, Shengtao Zou, Xinrui Yan, and Shuqi Mei. High-definition map automatic annotation system based on active learning. *AI Magazine*. Wiley Online Library, 2023.
- 2023 Xu Cao and Jianguo Cao. Commentary: Machine learning for autism spectrum disorder diagnosis—challenges and opportunities—a commentary on schulte-rüther et al.(2022). *Journal of Child Psychology and Psychiatry*. Wiley Online Library, 2023.

## Fellowships & Awards & Grants

- 2026 **Best Paper Award** in *CHIL 2026*
- 2026 **Google PhD Fellowship Finalist**: One of 4 candidates nominated by UIUC for invitation-only fellowship submission
- 2025 **Google Cloud Gift Funding** \$1,000: support my research on gaze foundation models
- 2025 **Top Reviewer Award (Top 5%)** in NeurIPS 2025
- 2025 **Canadian Institute for Advanced Research (CIFAR) Fellowship** \$10,000
- 2024 **OpenAI Gift Funding** \$3,000: support my research on human behavior foundation models
- 2023 **Innovative Applications of Artificial Intelligence Award** in *AAAI 2023*
- 2023 **AAAI Student Scholarship Grant** in *AAAI 2023*
- 2022 **IJCAI-AIJ Travel Grant** in *IJCAI-ECAI 2022*
- 2017-2019 **Outstanding Academic Scholarship** at *Fudan University*
- 2019 **Meritorious Winner** in *The Mathematical Contest in Modeling (MCM)*
- 2016 **Outstanding Freshman Scholarship** at *Fudan University*
- 2015 **Silver Medal** in *24th China National Biology Olympiad (CNBO)*

## Academia Service

- Founder & Workshops.**
- Chair CVPR 2026 Workshop on Computer Vision for Children  
CVPR 2026 AI Children Challenge  
ICLR 2025 Workshop on AI for Children: Healthcare, Psychology, Education  
WACV 2024 Workshop on Large Language and Vision Models for Autonomous Driving (LLVM-AD)
- Organizer Conferences.**
- Research Roundtable Subchair of **Machine Learning for Health Symposium (ML4H) 2024**
- Workshops.**
- ICCV 2025 2nd Workshop on Distillation of Foundation Models for Autonomous Driving  
CVPR 2025 Workshop on Distillation of Foundation Models for Autonomous Driving  
ITSC 2025 4th Workshop on Large Language and Vision Models for Autonomous Driving (LLVM-AD)  
WACV 2024 3rd Workshop on Large Language and Vision Models for Autonomous Driving (LLVM-AD)  
ITSC 2024 2nd Workshop on Large Language and Vision Models for Autonomous Driving (LLVM-AD)
- Area Chair Conferences.**
- NeurIPS 2026
- Workshops.**
- CVPR 2026 Workshop on Multi-Modal Reasoning for Agentic Intelligence (MMRAgI)  
ICLR 2026 Workshop on Efficient Spatial Reasoning

## Teaching

- 2026 **[SCCM Microlearning Series] AI and Personalized Data in Pediatric Critical Care**, Society of Critical Care Medicine (SCCM).