Zilong Zheng

Contact Information

(+86)159-5011-5760
zlzheng.cs@gmail.com
https://zilongzheng.github.io
https://scholar.google.com/citations?user=9sDx70IAAAAJ

Education

University of California, Los Angeles **Los Angeles, CA**
Advisor: Prof. Song-Chun Zhu
Ph.D. in Computer Science Jan. 2018 - June. 2021
M.S. in Computer Science Sept. 2016 - Dec. 2017

University of Minnesota, Twin Cities **Minneapolis, MN**
B.A. in Computer Science Jan. 2015 - June. 2016
- GPA: 3.78/4.0, **Distinction**

University of Electronic Science and Technology of China **Chengdu, China**
B.E. in Microelectronic Technology Sept. 2012 - Jan. 2015
- GPA: 3.91, Rank: **1/164**

Research Interests

Faithful Reasoning: Developing reasoning language models that adhere to logic, symbolic rules, factual knowledge, and commonsense.
Trustworthy Agents: Building safe, self-explainable, and socially aware language agents aligned with human intent, ethics, social norms, and values.
Efficient Long-context Modeling: Enabling efficient, scalable, and precise understanding of long contexts across text, video, and multi-modal data.
Generative Modeling: Modeling high-dimensional data via energy-based models and diffusion processes for controllable generation.

Professional Experience

Beijing Institute for General Artificial Intelligence (BIGAI) June.2021 - Present
Executive Director of Research Department
Team Lead of Natural Language and Conversational AI (NLCo) Lab

- General management of computing resource (A100 clusters) procurement, cluster deployment, hosting, and Slurm scheduling for the entire research department; responsible for student management and researcher/engineer hiring...
- Team lead of TongAgents: building faithful reasoning models, language model acceleration, interactive conversational modeling, agentic system
- Project Lead of TongVerse system, in collaboration with Leju Robot Team https://gitee.com/leju-robot/leju_kuavo_tongverse-lite
- Principal Investigator (PI) of "Intelligent Legal System" (in collaboration with Prof. Muhan Zhang from Peking University), 1,210K Yuan
- Principal Investigator (PI) of "Human-Machine Value Alignment in Embodied Environments", National Natural Science Foundation of China, General Program (62376031), 490K Yuan
- Project Leader of "Vision-Language Multimodal Parsing", National Key Research and Development Program of China (2021ZD0150200), 3M Yuan
- Co-PI (co-led with Prof. Yang Liu from Peking University) of Multimodal Perception Lab, State Key Laboratory of General Artificial Intelligence

Awards

- Associate Research Fellow (自然科学序列副研究员), Beijing

- 北京市劳动模范

Center for Vision, Cognition, Learning and Autonomy (VCLA) Jan. 2017 - June. 2021
Graduate Student Researcher
- Supervisor: Prof. Song-Chun Zhu
- Thesis: Multimodal Conversation Modeling via Neural Perception, Structure Learning, and Communication

Baidu Research, Cognitive Computing Laboratory June 2020 - Sept. 2020
Research Scientist Intern (Supervisor: Prof. Ping Li)

Google, Mountain View, CA June. 2017 - Sept. 2017
Software Engineer Intern
- Google Assistant Backend: Implementing acoustic controlling feature using C++

Oracle, Chengdu, China Oct. 2014 - Dec. 2014
Software Engineer Intern
- Database Analysis

Publications

(*: equal contributions, †: corresponding author)

Journal

[4] L. Yuan*, X. Gao*, **Z. Zheng***, M. Edmonds†, Y. N. Wu, F. Rossano, H. Lu†, Y. Zhu† and S.-C. Zhu†. In situ bidirectional human-robot value alignment, *Science Robotics*, July 2022. **Featured as headline news on Science and Science Robotics, July 2022. Ref. XinhuaNet, ScienceNet, TechXplore Report.**

[3] Z. Wang, S. Cai, A. Liu, Y. Jin, J. Hou, B. Zhang. H. Lin, Z. He, **Z. Zheng**, Y. Yang, X. Ma† and Y. Liang†. JARVIS-1: Open-world Multi-task Agents with Memory-Augmented Multimodal Language Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2024.

[2] J. Xie*, **Z. Zheng***, X. Fang, S.-C. Zhu and Y. N. Wu. Cooperative Training of Fast Thinking Initializer and Slow Thinking Solver for Multi-modal Conditional Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2021.

[1] J. Xie*, **Z. Zheng***, R. Gao, W. Wang, S.-C. Zhu and Y.N. Wu. Generative VoxelNet: Learning Energy-Based Models for 3D Shape Synthesis and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020.

Technical Reports & Preprints

[3] A. Zhao, Y. Wu, Y. Yue, T. Wu, Q. Xu, Y. Yue, M. Lin, S. Wang, Q. Wu, **Z. Zheng**† and G. Huang†. Absolute Zero: Reinforced Self-play Reasoning with Zero Data. 2025. **Github Stars: 1.6K+, X Views: 495K+, No. 1 Paper of the Day on Huggingface, AlphaXiv Trending No. 1**

[2] Y. Liu, J. Li and **Z. Zheng**†. RuleReasoner: Reinforced Rule-based Reasoning via Domain-aware Dynamic Sampling. 2025.

[1] H. Li, C. Li, T. Wu, X. Zhu, Y. Wang, Z. Yu, E. H. Jiang, S.-C. Zhu, Z. Jia, Y. N. Wu† and **Z. Zheng**†. Seek in the Dark: Reasoning via Test-Time Instance-Level Policy Gradient in Latent Space. 2025.

Conference

[49] Y. Wang*, Y. Song*, C. Xie, Y. Liu and **Z. Zheng**†. VideoLLaMB: Long-context Video Understanding with Recurrent Memory Bridges. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2025.

[48] H. Li, Z. Jiao, X. Liu†, H. Liu†, and **Z. Zheng**†. In-situ Value-aligned Human-Robot Interactions with Physical Constraints. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2025. **[Oral]**

[47] J. Zhu, Z. Du, H. Xu, F. Lan, **Z. Zheng**, B. Ma, S. Wang, and T. Zhang. Navi2Gaze: Leveraging Foundation Models for Navigation and Target Gazing. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2025. **[Oral]**

[46] Y. Peng*, Y. Ma*, M. Wang, Y. Wang, Y. Wang, C. Zhang, Y. Zhu†, and **Z. Zheng**†. Probing and Inducing Combinational Creativity in Vision-Language Models. In *Proceedings of the 47th Annual Meeting of the Cognitive Science Society (CogSci)*, 2025. **[Oral]**

[45] Z. Lin*, H. Wu*, S. Wang, K. Tu†, **Z. Zheng**†, and Z. Jia†. Look Both Ways and No Sink: Converting LLMs into Text Encoders without Training. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

[44] J. Li, X. Dong, Y. Liu, Z. Yang, Q. Wang, X. Wang, S.-C. Zhu, Z. Jia†, and **Z. Zheng**†. ReflectEvo: Improving Meta Introspection of Small LLMs by Learning Self-Reflection. In *Findings of the Association for Computational Linguistics: ACL*, 2025.

[43] Y. Kang, J. Wang, Y. Li, M. Wang, W. Tu, Q. Wang, H. Li, T. Wu, X. Feng, F. Zhong, and **Z. Zheng**†. Are the Values of LLMs Structurally Aligned with Humans? A Causal Perspective. In *Findings of the Association for Computational Linguistics: ACL*, 2025.

[42] X. Zheng*, H. Lin*, K. He, Z. Wang, **Z. Zheng**† and Y. Liang†. MCU: An Evaluation Framework for Open-Ended Game Agents. In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025. **[Spotlight]**

[41] T. Wu*, J. Shen*, Z. Jia, Y. Wang and **Z. Zheng**†. TokenSwift: Lossless Acceleration of Ultra Long Sequence Generation. In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025.

[40] X. Zhu, D. Cheng, H. Li, K. Zhang, E. Hua, X. Lv, N. Ding, Z. Lin†, **Z. Zheng**† and B. Zhou†. How to Synthesize Text Data without Model Collapse? In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025.

[39] Y. Wang, Y. Wang, B. Chen, T. Wu, D. Zhao and **Z. Zheng**†. OmniMMI: A Comprehensive Multi-modal Interaction Benchmark in Streaming Video Contexts. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, 2025.

[38] Z. Zhang, F. Bai, Q. Chen, C. Ma, M. Wang, H. Sun, **Z. Zheng**† and Y. Yang†. Amulet: ReAlignment During Test Time for Personalized Preference Adaptation of LLMs. In *Proceedings of the Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[37] S. Qi†, B. Yang, K. Jiang, X. Wang, J. Li, Y. ZHong, Y. Yang and **Z. Zheng**†. In-Context Editing: Learning Knowledge from Self-Induced Distributions. In *Proceedings of the Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[36] Y. Du*, K. Jiang*, Z. Gao, C. Shi, **Z. Zheng**†, S. Qi and Q. Li†. MMKE-Bench: A Multimodal Editing Benchmark for Diverse Visual Knowledge. In *Proceedings of the Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[35] A. Zhao, Q. Xu, M. Lin, S. Wang, Y.-J. Liu, **Z. Zheng**† and G. Huang†. DiveR-CT: Diversity-enhanced Red Teaming Large Language Model Assistants with Relaxing Constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2025. **[Oral]**

[34] T. Wu, Y. Zhao and **Z. Zheng**†. An Efficient Recipe for Long Context Extension via Middle-Focused Positional Encoding. In *the 38th Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

[33] X. Tang, J. Li, Y. Liang, M. Zhang and **Z. Zheng**†. Mars: Situated Inductive Reasoning in an Open-World Environment. In *the 38th Conference on Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks*, 2024.

[32] Y. Wang, Y. Wang, P. Wu, J. Liang, D. Zhao, Y. Liu and **Z. Zheng**†. Efficient Temporal Extrapolation of Multimodal Large Language Models with Temporal Grounding Bridge for Long Video Understanding. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[31] Z. Lin, Q. Wang, Z. Jia† and **Z. Zheng**†. Varying Sentence Representations via Condition-Specified Routers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[30] S. Qiu, M. Liu, H. Li, S.-C. Zhu and **Z. Zheng**†. MindDial: Enhancing Conversational Agents with Theory-of-Mind for Common Ground Alignment and Negotiation. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 2024. **[Oral]**

[29] Y. Wang, A. Yuille, Z. Li† and **Z. Zheng**†. ExoViP: Step-by-step Verification and Exploration with Exoskeleton Modules for Compositional Visual Reasoning. In *The First Conference on Language Modeling (CoLM)*, 2024.

[28] J. Li, M. Wang, **Z. Zheng**† and M. Zhang†. LooGLE: Can Long-Context Language Models Understand Long Contexts? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

[27] Z. Jia, J. Li, S. Zhang and **Z. Zheng**†. Combining Supervised Learning and Reinforcement Learning for Multi-Label Classification Tasks with Partial Labels. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

[26] S. Wang, C. Liu, **Z. Zheng**†, S. Qi, S. Chen, Q. Yang, A. Zhao, S. Wang, S. Song and G. Huang†. Boosting LLM Agents with Recursive Contemplation for Effective Deception Handling. In *Findings of the Association for Computational Linguistics: ACL*, 2024.

[25] Z. Jia*, M. Wang*, B. Tong, S.-C. Zhu and **Z. Zheng**†. LangSuit·E: Controlling, Planning, and Interacting with Large Language Models in Embodied Text Environments. In *Findings of the Association for Computational Linguistics: ACL*, 2024.

[24] R. Gong, Q. Huang, X. Ma, H. Vo, Z. Durante, Y. Noda, **Z. Zheng**, D. Terzopoulos, F.-F. Li and J. Gao. MindAgent: Emergent Gaming Interaction. In *Findings of the Association for Computational Linguistics: NAACL*, 2024.

[23] J. Cui*, Z. Gong*, B. Jia*, S. Huang, **Z. Zheng**†, J. Ma† and Y. Zhu†. ProBio: A Protocol-guided Multimodal Dataset for Molecular Biology Lab. In *the 37th Conference on Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks*, 2023.

[22] H. Li, S.-C. Zhu and **Z. Zheng**†. DiPlomat: A Dialogue Dataset for Situated Pragmatic Reasoning. In *the 37th Conference on Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks*, 2023.

[21] X. Ma*, S. Yong*, **Z. Zheng**†, Q. Li, Y. Liang, S.-C. Zhu and S.Huang†. SQA3D: Situated Question Answering in 3D Scenes. In *the Eleventh International Conference on Learning Representations (ICLR)*, 2023. **Official challenge benchmark for workshop on 3D Scene Understanding for Vision, Graphics, and Robotics, CVPR 2025.**

[20] J. Li*, Z. Jia* and **Z. Zheng**†. Semi-automatic Data Enhancement for Document-Level Relation Extraction with Distant Supervision from Large Language Models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

[19] Y. Wang, **Z. Zheng**†, X. Zhao, J. Li, Y. Wang and D. Zhao†. VSTAR: A Video-grounded Dialogue Dataset for Situated Semantic Understanding with Scene and Topic Transitions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

[18] Z. Jia, Z. Yan, W. Han, **Z. Zheng**† and K. Tu †. Modeling Instance Interactions for Joint Information Extraction with Neural High-Order Conditional Random Field. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

[17] Y. Wang, J. Wang, D. Zhao† and **Z. Zheng**†. Shuō Wén Jiě Zì: Rethinking Dictionaries and Glyphs for Chinese Language Pre-training. *Findings of the association for computational linguistics: ACL*, 2023.

[16] L. Zhang, Z. Jia, W. Han, **Z. Zheng** and K. Tu. SHARP: Search-Based Adversarial Attack for Structured Prediction. In *Findings of the Association for Computational Linguistics: NAACL*, 2022.

[15] Y. Li, **Z. Zheng**, W. Han and L. Zou. VGStore: A Multimodal Extension to SPARQL for Querying RDF Scene Graph. In *ISWC Poster & Demo Track*, 2022.

[14] C. Lou*, W. Han, Y. Lin and **Z. Zheng***. Unsupervised Vision-Language Parsing: Seamlessly Bridging Visual Scene Graphs with Language Structures via Dependency Relationships. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[13] B. Wan, W. Han, **Z. Zheng** and T. Tuytelaars. Unsupervised Vision-Language Grammar Induction with Shared Structure Modeling. In *The Tenth International Conference on Learning Representations (ICLR)*, 2022. **[Oral]**

[12] J. Zhang, J. Xie, **Z. Zheng** and N. Barnes. Energy-Based Generative Cooperative Saliency Prediction. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 2022. **[Oral]**

[11] **Z. Zheng**, J. Xie and P. Li. Patchwise Generative ConvNet: Training Energy-Based Models from a Single Natural Image for Internal Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. **[Oral]**

[10] L. Fan*, S. Qiu*, **Z. Zheng**, T. Gao, S.-C. Zhu and Y. Zhu. Learning Triadic Belief Dynamics in Nonverbal Communication from Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. **[Oral]**

[9] J. Xie, Y. Xu, **Z. Zheng**, S.-C. Zhu and Y.N. Wu. Generative PointNet: Deep Energy-Based Learning on Unordered Point Sets for 3D Generation, Reconstruction and Classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[8] J. Xie*, **Z. Zheng***, X. Fang, S.-C. Zhu and Y.N. Wu. Learning Cycle-Consistent Cooperative Networks via Alternating MCMC Teaching for Unsupervised Cross-Domain Translation. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

[7] J. Xie, **Z. Zheng** and P. Li. Learning Energy-Based Model with Variational Auto-Encoder as Amortized Sampler. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

[6] **Z. Zheng**, S. Qiu, L. Fan, Y. Zhu and S.-C. Zhu. GRICE: A Grammar-based Dataset for Recovering Implicature and Conversational rEasoning. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP*, 2021.

[5] J. Xie*, R. Gao*, **Z. Zheng** and S.-C. Zhu. Motion-Based Generator Model: Unsupervised Disentanglement of Appearance, Trackable and Intrackable Motions in Dynamic Patterns. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020. **[Oral]**

[4] T. Yuan, H. Liu, L. Fan, **Z. Zheng**, T. Gao, Y. Zhu and S.-C. Zhu. Joint Inference of States, Robot Knowledge, and Human (False-)Beliefs. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[3] **Z. Zheng***, W. Wang*, S. Qi* and S.-C. Zhu. Reasoning Visual Dialogs with Structural and Partial Observations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. **[Oral]**

[2] J. Xie*, R. Gao*, **Z. Zheng**, S.-C. Zhu and Y.N. Wu. Learning Dynamic Generator Model by Alternating Back-Propagation Through Time. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, 2019. **[Spotlight]**

[1] J. Xie*, **Z. Zheng***, R. Gao, W. Wang, S.-C. Zhu and Y.N. Wu. Learning Descriptor Networks for 3D Shape Synthesis and Analysis. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. **[Oral]**

Professional Activities

Conference/Journal Area Chairs/Reviewers

- ❑ Senior Area Chair, Empirical Methods in Natural Language Processing (EMNLP), 2025
- ❑ Senior Area Chair, AACL: Asia-Pacific Chapter of the ACL (AACL), 2025
- ❑ Area Chair, Association for Computational Linguistics (ACL), 2025
- ❑ Journal Reviewer, International Journal on Computer Vision (IJCV), NeuroComputing, Pattern Recognition, Engineering
- ❑ Regular Reviewer, ICML, ICLR, NeurIPS, CVPR, ICCV, ECCV, ...

Workshop Organizers

- ❑ Primary Host, The 1st Joint Workshop on Large Language Models and Structure Modeling (XLLM), ACL 2025
- ❑ Host, General Embodied Intelligence in Real and Virtual Open Worlds, China Embodied AI Conference (CEAI), 2024, 2025
- ❑ Host, Open-World Agents: Synnergizing Reasoning and Decision-Making in Open-World Environments (OWA-2024), NeurIPS 2024
- ❑ Primary Host, Unimodal and Multimodal Induction of Linguistic Structures (UM-IoS), EMNLP 2021