HUANXUAN LIAO (廖桓萱)

EDUCATION

Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing

2023 - Now

PhD in Computer Science. In NLPR, Supervisor: Shizhu He and Kang Liu

North China Electric Power University (NCEPU), Beijing

2019 - 2023

Bachelor of Engineering in Intelligence Science and Technology (Innovation Class). Ranking 1/13

№ RESEARCH INTERESTS

- · Long Context Modeling:
 - Survey: Awesome-LLM-Long-Context-Modeling (GitHub Stars > 1.7K)
 - Context Compression: HyCo2 (ArXiv 2025)
 - KV Compression: SparK (ArXiv 2025)
- Efficient Reasoning and Knowledge Activation:
 - Instruction Learning: TAGI (NeurIPS 2024)
 - Knowledge Awakening: AAG (COLING 2025)
 - Knowledge Internalization: SKIntern (COLING 2025)
 - Knowledge Collaboration: NesyCD (AAAI 2025)
 - Knowledge Transfer: DATA (ArXiv 2025)
- · Efficient Multimodal Models and Reasoning

■ PUBLICATIONS (SELECTED) [FULL LIST]

(* stands for equal contribution; Listed in reverse chronological order.)

- <u>Huanxuan Liao</u>, Yixing Xu, Shizhu He, Guanchen Li, Xuanwu Yin, Dong Li, Emad Barsoum, Jun Zhao, Kang Liu. SparK: Query-Aware Unstructured Sparsity with Recoverable KV Cache Channel Pruning. (Preprint, 2025)
- <u>Huanxuan Liao</u>, Shizhu He, Yupu Hao, Jun Zhao, Kang Liu. Beyond Hard and Soft: Hybrid Context Compression for Balancing Local and Global Information Retention. (Preprint, 2025)
- <u>Huanxuan Liao</u>, Shizhu He, Yupu Hao, Jun Zhao, Kang Liu. DATA: Decomposed Attention-based Task Adaptation for Rehearsal-Free Continual Learning.
- <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Yuanzhe Zhang, Kang Liu, Jun Zhao. Neural-Symbolic Collaborative Distillation: Advancing Small Language Models for Complex Reasoning Tasks. (AAAI 2025, CCF-A)
- <u>Huanxuan Liao</u>, Shizhu He, Yupu Hao, Xiang Li, Yuanzhe Zhang, Jun Zhao, Kang Liu. SKIntern: Internalizing Symbolic Knowledge for Distilling Better CoT Capabilities into Small Language Models. (COLING 2025, CCF-B)
- <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Yuanzhe Zhang, Kang Liu, Shengping Liu, Jun Zhao. Awakening Augmented Generation: Learning to Awaken Internal Knowledge of Large Language Models for Question Answering. (COLING 2025, CCF-B)
- <u>Huanxuan Liao</u>, Shizhu He, Yao Xu, Yuanzhe Zhang, Yanchao Hao, Shengping Liu, Kang Liu, Jun Zhao. From Instance Training to Instruction Learning: Task Adapters Generation from Instructions. (NeurIPS 2024, CCF-A)

MINTERNSHIP/PROJECT EXPERIENCE

ByteDance Beijing

Jun. 2025 - Now

Intern Research Intern (Efficient VLM)

Ant Group Beijing

Feb. 2025 - May 2025

Intern Research Intern (Long Context)

Context Compression Based on Global and Local Strategy

• Implemented dynamic routing soft token global compression based on Qformer and MLP, and local compression based on the classification layer

Survey of Long Context Modeling

• Completed the writing of the survey based on my own long context modeling repository

KV Compression

• Perform dynamic unstructured channel pruning while restoring the pruned channels during attention computation

Scholarships & Awards Climbing Second Prize Scholarship Dec. 2024 Beijing Outstanding Graduate Awards, Beijing Ministry of Education Jun. 2023 China National Scholarship (top scholarship in China; 0.2% domestically), Ministry of Education Dec. 2022 China National Scholarship (top scholarship in China; 0.2% domestically), Ministry of Education Dec. 2021 SiFang Society Scholarship, NCEPU Dec. 2020 Competition National Third Prize, Information Security Competition Aug. 2022 National Excellence, College Student Innovation and Entrepreneurship Project Dec. 2021 Beijing Third Prize, Internet + Aug. 2021

♦ PROFESSIONAL SERVICES

Conference Reviewing: ACL ARR 2024, 2025. NeurIPS 2025. Community: Awesome-LLM-Long-Context-Modeling 1.7k+ Stars.