

XIAOSONG YUAN | CV

- » Languages: Chinese, English (IELTS 7.0 overall, 2022-Oct)
- » Interests: Deep Learning, NLP, LLMs
- » ML Tools: PyTorch, HuggingFace Transformers



» » » Summary

I am a 5th-year Ph.D. candidate at Jilin University, Changchun, China. My prior research topic concentrated on causal relation extraction in open-source texts and controllable text generation, now I'm working on improving LLM reasoning capabilities with SFT and inference optimization, and agent systems.

» » » Education

- Sep 2021 - Present **Ph.D. candidate (successive master-doctor program)**
College of Computer Science and Technology, Jilin University, Changchun, China
- Sep 2019 - Jun 2021 **Master candidate (successive master-doctor program)**
College of Computer Science and Technology, Jilin University, Changchun, China
- Sep 2014 - Jun 2018 **Bachelor's Degree**
College of Software, Jilin University, Changchun, China

» » » Work Experience

- Dec 2023 - Dec 2025 **Research Intern at the City Brain Lab, DAMO Academy, Alibaba Group**
advised by Chen Shen and Prof. Jieping Ye
- Jan 2026 - Feb 2026 **Research Intern at the AQ-Team, Ant Group**
- Feb 2026 - Now **Research Intern at Tongyi Lab, Alibaba Group**
advised by Xuexiang Xie, Yaliang Li, and Bolin Ding

» » » Publications

1st author publications

- Instance-adaptive Zero-shot Chain-of-Thought Prompting. (NeurIPS 2024, poster)**
 - 1) We analyze LLMs through the lens of information flow to detect the mechanism of zero-shot CoT;
 - 2) We propose an instance-adaptive prompting strategy to improve reasoning task performance.
- Differentially Fine-tuning LLMs Towards Better Diverse Reasoning Abilities. (ICLR 2026, poster)**
 - 1) We explore the parameter variations between fine-tuned and base LLMs during inference;
 - 2) We fine-tune the parameters of LLMs differentially to avoid reasoning conflicts and maintain benefits.
- SeaRAG: Reducing Hallucination in RAG via Statement-Entity Adaptive Ranking (WWW 2026, oral)**
 - 1) We utilize real-time statement- and entity-level probing to detect hallucinations via model uncertainty.
 - 2) We introduce entropy-ranking to select documents, reducing retrieval frequency while improving accuracy.
- LSFusion: Ladder Side Attribute Composition for Multi-Aspect Controllable Text Generation. (IEEE Transactions on Consumer Electronics)**
 - 1) We propose a parameter-efficient ladder-side fusion method LSFusion tailored for multi-aspect CTG;
 - 2) To mitigate attribute forgetting, we learn and fuse multi-aspect attribute knowledge in two stages.
- Causal Strength Bank: A New Benchmark for Causal Strength Classification. (PAKDD 2023, oral)**
 - 1) We present a new causal strength classification task, which is beneficial to causal relation extraction;
 - 2) We develop a new benchmark dataset-Causal Strength Bank and experiment with classic models.
- TC-GAT: Graph Attention Network for Temporal Causality Discovery. (IJCNN 2023, oral)**

- 1) We construct a causality dataset containing multiple temporal relations and build a causal graph;
- 2) We propose TC-GAT to obtain the weight of temporal relations from the Causal KG.

Other author publications

1. Hallucination Begins Where Saliency Drops. (ICLR 2026, oral)

- 1) We introduce LVLMs-Saliency to reveal that hallucinations emerge when prior tokens exhibit low saliency;
- 2) We propose a dual framework to filter low-saliency candidate tokens and reinforce contextual coherence.

2. Context Tokens are Anchors: Understanding the Repetition Curse in Diffusion MLLMs from an Information Flow Perspective. (ICLR 2026, poster)

- 1) We reveal dMLLMs caching triggers a "repeat curse" by chaos information flow and entropy inconvergence;
- 2) We enhance attention to context tokens and penalize uncertain predictions with entropy-guided voting.

3. Improving Complex Reasoning with Dynamic Prompt Corruption: A Soft Prompt Optimization Approach. (ICLR 2025, poster)

- 1) We discover that soft prompts positively impact certain instances while negatively affecting others;
- 2) We propose to optimize the soft prompts by dynamically adjusting their influence on reasoning tasks.

4. From Redundancy to Relevance: Enhancing Explainability in Multimodal Large Language Models. (NAACL 2025, oral)

- 1) We analyze LVLMs through information flow to detect the mechanism of complex reasoning tasks;
- 2) We propose truncation strategies to reveal a consistent pattern of information flow convergence.

5. Shallow Focus, Deep Fixes: Enhancing Shallow Layers Vision Attention Sinks to Alleviate Hallucination in LVLMs. (EMNLP 2025, oral)

- 1) We reveal that most hallucinations are linked to the attention sink of image tokens attention matrix;
- 2) We propose a training-free Enhancing Vision Attention Sinks to guide the attention sink convergence.

Academic & Industrial Projects

Intelligent causality learning for open domain texts (NSFC)	Jan 2020 - 2023
Deep learning-based causality learning and large knowledge base construction (Science and Technology Department of Jilin Province)	Sep 2019 - 2022
X1 Project (Apsara Lab, Alibaba Cloud Computing)	Oct 2024 - 2025
QwenPaw (Tongyi Lab, Alibaba Cloud Computing)	Feb 2026 -
Now	
ReMe (Tongyi Lab, Alibaba Cloud Computing)	Feb 2026 - Now

Research Activities

NLPCC 2020, NeurIPS 2020, PAKDD 2023, IJCNN 2023, NeurIPS 2024, NLPCC 2024, EMNLP 2025

Academic service

Conference:

NeurIPS 2024-2025, ICLR 2025-2026, ICML 2025-2026, AAAI 2026, EMNLP 2023, ACM MM 2025-2026, ACM SIGIR 2026, ICASSP 2024-2026, AISTATS 2025-2026, IJCNN 2025-2026

Journal:

TMLR, IEEE TNNLS, IEEE TII, PR, KBS, IEEE TBD, Neurocomputing

Honors & Awards

Academic Performance Scholarship, Jilin University, 2021.

Academic Performance Scholarship, Jilin University, 2022.

Academic Performance Scholarship, Jilin University, 2023.

