Junru Lu

NLP / LLM / MLLM

@ junru31415926@gmail.com lujunru.github.io Full Paper List



Education

Ph.D. in Computer Science — Nov. 2019 – June 2024
University of Warwick (Advisor: Prof. Yulan He) — Coventry, UK

M.S. in Applied Urban Science & Informatics — Sept. 2018 – Sept. 2019
New York University — New York City, USA

Double Bachelor of Economics — Sept. 2015 – July 2018
Peking University — Beijing, China

B.Eng. in Information Management and Information System — Sept. 2014 – July 2018
University of International Relations (Advisor: Prof. Binyang Li) — Beijing, China

Career

Foundation Model Researcher, YouTu Lab — May 2023 – present
Tencent — Shanghai, China

- **Data**: Built a 20TB+ text pipeline and evidence-checked QA synthesis workflows.
- **Arch**: Led native multimodal research and designed sparse-attention methods, guiding their implementation.
- **Infra**: Implemented attention kernels for efficient long-context training.
- **Train**: Worked on MoE pre-training up to 440B parameters, and developed post-training methods.
- **Eval**: Developed benchmarks for instruction-following conflicts and divergent long-tail knowledge.

Selected Publications

- **[From-scratch 2B LLM] Youtu-LLM: Unlocking the Native Agentic Potential for Lightweight Large Language Models**
  Junru Lu, Jiarui Qin, et al. (2025). arXiv preprint arXiv:2512.24618.
- **[From-scratch 4B MLLM] Youtu-VL: Unleashing Visual Potential via Unified Vision-Language Supervision**
  Zhixiang Wei, Yi Li, et al. (2026). arXiv preprint arXiv:2601.19798.
- **Toward Native Multimodal Modeling: A Roadmap**
  Siyu An, **Junru Lu**, et al. (2026). EMNLP 2026.
- **ContextPilot: Teaching Agents for Proactive Context Management via Fine-grained RL**
  Zhuoshi Pan, Qizhi Pei, et al. (2026). EMNLP 2026.
- **SSA: Sparse Sparse Attention by Aligning Full and Sparse Attention Outputs in Feature Space**
  Zhenyi Shen, **Junru Lu**, et al. (2025). ICML 2026.
- **Optiscene: Llm-driven indoor scene layout generation via scaled human-aligned data synthesis and multi-stage preference optimization**
  Yixuan Yang, Zhen Luo, et al. (2026). NeurIPS 2025.
- **Two heads are better than one: Dual-model verbal reflection at inference-time**
  Jiazheng Li, Yuxiang Zhou, et al. (2025). EMNLP 2025.
- **RoleMRC: A Fine-Grained Composite Benchmark for Role-Playing and Instruction-Following**
  Junru Lu, Jiazheng Li, Guodong Shen, et al. (2025). ACL 2025.
- **Tell Me What You Don't Know: Enhancing Refusal Capabilities of Role-Playing Agents via Representation Space Analysis and Editing**
  Wenhao Liu, Siyu An, et al. (2024). ACL 2025.
- **Eliminating Biased Length Reliance of Direct Preference Optimization via Down-Sampled KL Divergence**
  Junru Lu, Jiazheng Li, Siyu An, et al. (2024). EMNLP 2024.
- **FIPO: Free-form Instruction-oriented Prompt Optimization with Preference Dataset and Modular Fine-tuning Schema**
  Junru Lu, Siyu An, Min Zhang, et al. (2025). COLING 2025.
- **Advancing Video Anomaly Detection: A Bi-Directional Hybrid Framework for Enhanced Single- and Multi-Task Approaches**
  Guodong Shen, Yuqi Ouyang, et al. (2024). Transactions on Image Processing (TIP 2024).
- **NapSS: Paragraph-level Medical Text Simplification via Narrative Prompting and Sentence-matching Summarization**
  Junru Lu, Jiazheng Li, Pergola Gabriele, et al. (2022). EACL 2023.
- **Event-Centric Question Answering via Contrastive Learning and Invertible Event Transformation**
  Junru Lu, Xingwei Tan, et al. (2022). EMNLP 2022.
- **Chime: Cross-passage Hierarchical Memory Network for Generative Review Question Answering**
  Junru Lu, Pergola Gabriele, et al. (2020). COLING 2020.