Junyi Li (李军毅)

junyi.li@umontreal.ca Google Scholar https://lijunyi.tech/

Education

2021 – present **Ph.D., Université de Montréal**
Supervisor: *Prof. Jian-Yun Nie (Canada Research Chair, Tier 1).*

2018 – 2021 **M.Sc. Computer Science, Renmin University of China**
Supervisor: *Prof. Xin Zhao.*

2014 – 2018 **B.Sc. Computer Science, Renmin University of China**
Supervisor: *Prof. Xin Zhao.*

Research Interests

My research interests lie broadly in natural language processing and large language models, with an emphasis on multi-agent systems. I am always excited about building trustworthy and evolving proactive agentic systems.

• *Planning, Reasoning, and Proactive Thinking with LLM Agents*: developing techniques that make LLM agents plan and reason more rigorously, enabling the agent to proactively verify and refine its own reasoning, and building deep research&search systems like OpenAI's DeepResearch;

• *Mitigating Hallucinations and Ensuring Trustworthiness*: rooting out hallucinations by attacking the problem from multiple angles: evaluating and detecting false outputs, understanding the sources of hallucination in LLMs, and actively preventing or correcting hallucinations during generation;

• *Evolving Agentic Behavior through Memory and Experience*: building agentic systems with persistent memory and adaptive learning from historical experience and interactions, adapting to new domains and user needs even after deployment.

Publications

† These authors contributed equally to this work.

Journal Articles

Xiaoxue, Cheng†, **Junyi**, **Li**†, Wayne Xin, Zhao, and Ji-Rong, Wen, "Think more, hallucinate less: Mitigating hallucinations via dual process of fast and slow thinking," *CoRR*, vol. abs/2501.01306, 2025. DOI: 10.48550/ARXIV.2501.01306. arXiv: 2501.01306.

Zican, Dong†, **Junyi**, **Li**†, Jinhao, Jiang, *et al.*, "Longred: Mitigating short-text degradation of long-context large language models via restoration distillation," *CoRR*, vol. abs/2502.07365, 2025. DOI: 10.48550/ARXIV.2502.07365. arXiv: 2502.07365.

Ruiyang, Ren†, Yuhao, Wang†, **Junyi**, **Li**†, *et al.*, "Holistically guided monte carlo tree search for intricate information seeking," *CoRR*, vol. abs/2502.04751, 2025. DOI: 10.48550/ARXIV.2502.04751. arXiv: 2502.04751.

Junyi, **Li** and Hwee Tou, Ng, "Think&cite: Improving attributed text generation with self-guided tree search and progress reward modeling," *CoRR*, vol. abs/2412.14860, 2024. DOI: 10.48550/ARXIV.2412.14860. arXiv: 2412.14860.

Junyi, **Li**, Tianyi, Tang, Wayne Xin, Zhao, Jian-Yun, Nie, and Ji-Rong, Wen, "Pre-trained language models for text generation: A survey," *ACM Comput. Surv.*, vol. 56, no. 9, Apr. 2024, ISSN: 0360-0300. DOI: 10.1145/3649449.

Wayne Xin, Zhao, Kun, Zhou†, **Junyi**, **Li**†, *et al.*, "A survey of large language models," *CoRR*, vol. abs/2303.18223, 2023. DOI: 10.48550/ARXIV.2303.18223. arXiv: 2303.18223.

Conference Proceedings

Jinhao, Jiang†, Jiayi, Chen†, **Junyi**, **Li**†, *et al.*, "RAG-star: Enhancing deliberative reasoning with retrieval augmented verification and refinement," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Albuquerque, New Mexico: Association for Computational Linguistics, Apr. 2025, pp. 7064–7074, ISBN: 979-8-89176-189-6. URL: https://aclanthology.org/2025.naacl-long.361/.

Jinhao, Jiang†, **Junyi**, **Li**†, Xin, Zhao, Yang, Song, Tao, Zhang, and Ji-Rong, Wen, "Mix-CPT: A domain adaptation framework via decoupling knowledge learning and format alignment," in *The Thirteenth International Conference on Learning Representations*, 2025. URL: https://openreview.net/forum?id=h1XoHOd19I.

Xiaolei, Wang, Xinyu, Tang, **Junyi**, **Li**, Xin, Zhao, and Ji-Rong, Wen, "Investigating the pre-training dynamics of in-context learning: Task recognition vs. task learning," in *The Thirteenth International Conference on Learning Representations*, 2025. URL: https://openreview.net/forum?id=htDczodFN5.

Xiaoxue, Cheng, **Junyi**, **Li**, Wayne Xin, Zhao, and Ji-Rong, Wen, "Chainlm: Empowering large language models with improved chain-of-thought prompting," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy*, ELRA and ICCL, 2024, pp. 2969–2983. URL: https://aclanthology.org/2024.lrec-main.265.

Xiaoxue, Cheng†, **Junyi**, **Li**†, Xin, Zhao, *et al.*, "Small agent can also rock! empowering small language models as hallucination detector," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, Association for Computational Linguistics, 2024, pp. 14 600–14 615. URL: https://aclanthology.org/2024.emnlp-main.809.

Zican, Dong†, **Junyi**, **Li**†, Xin, Men, *et al.*, "Exploring context window of large language models via decomposed positional vectors," in *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL: http://papers.nips.cc/paper_files/paper/2024/hash/1403ab1a427050538ec59c7f570aec8b-Abstract-Conference.html.

Zican, Dong, Tianyi, Tang, **Junyi**, **Li**, Wayne Xin, Zhao, and Ji-Rong, Wen, "BAMBOO: A comprehensive benchmark for evaluating long text modeling capacities of large language models," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy*, ELRA and ICCL, 2024, pp. 2086–2099. URL: https://aclanthology.org/2024.lrec-main.188.

Junyi, **Li**, Jie, Chen, Ruiyang, Ren, *et al.*, "The dawn after the dark: An empirical study on factuality hallucination in large language models," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, Association for Computational Linguistics, 2024, pp. 10 879–10 899. DOI: 10.18653/V1/2024.ACL-LONG.586.

Yuhao, Wang, Ruiyang, Ren, **Junyi**, **Li**, Xin, Zhao, Jing, Liu, and Ji-Rong, Wen, "REAR: A relevance-aware retrieval-augmented framework for open-domain question answering," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, Association for Computational Linguistics, 2024, pp. 5613–5626. URL: https://aclanthology.org/2024.emnlp-main.321.

Yifan, Du, **Junyi**, **Li**, Tianyi, Tang, Wayne Xin, Zhao, and Ji-Rong, Wen, "Zero-shot visual question answering with language model feedback," in *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, Association for Computational Linguistics, 2023, pp. 9268–9281. DOI: 10.18653/V1/2023.FINDINGS-ACL.590.

Junyi, **Li**, Xiaoxue, Cheng, Xin, Zhao, Jian-Yun, Nie, and Ji-Rong, Wen, "Halueval: A large-scale hallucination evaluation benchmark for large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, Association for Computational Linguistics, 2023, pp. 6449–6464. DOI: 10.18653/V1/2023.EMNLP-MAIN.397.

Junyi, **Li**, Tianyi, Tang, Wayne Xin, Zhao, Jingyuan, Wang, Jian-Yun, Nie, and Ji-Rong, Wen, "The web can be your oyster for improving language models," in *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, Association for Computational Linguistics, 2023, pp. 728–746. DOI: 10.18653/V1/2023.FINDINGS-ACL.46.

Tianyi, Tang, Yushuo, Chen, Yifan, Du, **Junyi**, **Li**, Wayne Xin, Zhao, and Ji-Rong, Wen, "Learning to imagine: Visually-augmented natural language generation," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, Association for Computational Linguistics, 2023, pp. 9468–9481. DOI: 10.18653/V1/2023.ACL-LONG.526.

Tianyi, Tang, **Junyi**, **Li**, Wayne Xin, Zhao, and Ji-Rong, Wen, "MVP: multi-task supervised pre-training for natural language generation," in *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, Association for Computational Linguistics, 2023, pp. 8758–8794. DOI: 10.18653/V1/2023.FINDINGS-ACL.558.

Yifan, Du, Zikang, Liu, **Junyi**, **Li**, and Wayne Xin, Zhao, "A survey of vision-language pre-trained models," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, ijcai.org, 2022, pp. 5436–5443. DOI: 10.24963/IJCAI.2022/762.

Junyi, **Li**, Tianyi, Tang, Zheng, Gong, *et al.*, "Eliteplm: An empirical study on general language ability evaluation of pretrained language models," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, Association for Computational Linguistics, 2022, pp. 3519–3539. DOI: 10.18653/V1/2022.NAACL-MAIN.258.

Junyi, **Li**, Tianyi, Tang, Jian-Yun, Nie, Ji-Rong, Wen, and Xin, Zhao, "Learning to transfer prompts for text generation," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, Association for Computational Linguistics, 2022, pp. 3506–3518. DOI: 10.18653/V1/2022.NAACL-MAIN.257.

Junyi, **Li**, Tianyi, Tang, Wayne Xin, Zhao, Jian-Yun, Nie, and Ji-Rong, Wen, "ELMER: A non-autoregressive pre-trained language model for efficient and effective text generation," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, Association for Computational Linguistics, 2022, pp. 1044–1058. DOI: 10.18653/V1/2022.EMNLP-MAIN.68.

Tianyi, Tang, **Junyi**, **Li**, Zhipeng, Chen, *et al.*, "Textbox 2.0: A text generation library with pre-trained language models," in *Proceedings of the The 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022 - System Demonstrations, Abu Dhabi, UAE, December 7-11, 2022*, Association for Computational Linguistics, 2022, pp. 435–444. DOI: 10.18653/V1/2022.EMNLP-DEMOS.42.

Tianyi, Tang, **Junyi**, **Li**, Wayne Xin, Zhao, and Ji-Rong, Wen, "Context-tuning: Learning contextualized prompts for natural language generation," in *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, International Committee on Computational Linguistics, 2022, pp. 6340–6354. URL: https://aclanthology.org/2022.coling-1.552.

Junyi, **Li**, Tianyi, Tang, Gaole, He, *et al.*, "Textbox: A unified, modularized, and extensible framework for text generation," in *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL 2021 - System Demonstrations, Online, August 1-6, 2021*, Association for Computational Linguistics, 2021, pp. 30–39. DOI: 10.18653/V1/2021.ACL-DEMO.4.

Junyi, **Li**, Tianyi, Tang, Wayne Xin, Zhao, Zhicheng, Wei, Nicholas Jing, Yuan, and Ji-Rong, Wen, "Few-shot knowledge graph-to-text generation with pretrained language models," in *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, ser. Findings of ACL, vol. ACL/IJCNLP 2021, Association for Computational Linguistics, 2021, pp. 1558–1568. DOI: 10.18653/V1/2021.FINDINGS-ACL.136.

Junyi, **Li**, Wayne Xin, Zhao, Zhicheng, Wei, Nicholas Jing, Yuan, and Ji-Rong, Wen, "Knowledge-based review generation by coherence enhanced text planning," in *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, ACM, 2021, pp. 183–192. DOI: 10.1145/3404835.3462865.

Gaole, He†, **Junyi**, **Li**†, Wayne Xin, Zhao, Peiju, Liu, and Ji-Rong, Wen, "Mining implicit entity preference from user-item interaction data for knowledge graph completion via adversarial learning," in *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, ACM / IW3C2, 2020, pp. 740–751. DOI: 10.1145/3366423.3380155.

Junyi, **Li**, Siqing, Li, Wayne Xin, Zhao, *et al.*, "Knowledge-enhanced personalized review generation with capsule graph neural network," in *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, ACM, 2020, pp. 735–744. DOI: 10.1145/3340531.3411893.

Junyi, **Li**, Wayne Xin, Zhao, Ji-Rong, Wen, and Yang, Song, "Generating long and informative reviews with aspect-aware coarse-to-fine decoding," in *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Association for Computational Linguistics, 2019, pp. 1969–1979. DOI: 10.18653/V1/P19-1190.

Books and Chapters

X. Zhao, **J. Li**, K. Zhou, T. Tang, and J.-R. Wen, Eds., *Large Language Model*. Beijing, China: Higher Education Press, 2024, ISBN: 978-704-0634-96-9.

Academic Service

Reviewer
- Journal: TALLIP, TKDD, ACM Computing Survey, Computational Intelligence.
- Conference: AAAI 2021-24, IJCAI 2021-24, EMNLP 2022-2024, COLING 2022-2023, ACL 2023-2024

Chair
- ACL ARR (Area Chair), CSSNLP 2020 (Track Chair).

Awards and Honors

Nov, 2021 — **National Scholarship** (2%), Awarded by Ministry of Education of China.

Apr, 2021 — **SIGIR Student Travel Grant**, Awarded by CIKM 2021.

Nov, 2019 — **National Scholarship** (2%), Awarded by Ministry of Education of China.