Erxin Yu

✉ erxin.yu@connect.polyu.hk   🌐 https://erxinyu.com/   📞 +86 15659126198

Education

2022 – Now	**Department of Computing, The HongKong Polytechnic University** Ph.D., Computer Science, Supervisor: Dr. Jing Li
2019 – 2022	**School of Artificial Intelligence, Jilin University**. M.S. Computer Science, Supervisor: Dr. Yi Chang
2015 – 2019	**School of Software, Jilin University**. B.S. Software Engineering.

Research Experience

Data Selection/Data Synthesis for LLM.
Research Area: I am investigating methods to diversify and generalize data, selecting data that enhances model performance. This work is ongoing.

LLM Safety.
I have conducted research on the safety of large language models in multi-turn dialogue coreference, creating a dataset of 1,400 questions covering 14 categories of coreference safety attacks. I have evaluated five widely-used open-source LLMs, and the paper will be presented at EMNLP 2024.

Popularity-Aligned LLM.
I have worked on aligning large models with language from social media, constructing datasets such as QuoteTweet, QuoteLink, and Weibo Comment, and using PPO methods for alignment. This work has been published in COLING 2024 and ACL 2024.

Research Publications

- **CoSafe: Evaluating Large Language Model Safety in Multi-Turn Dialogue Coreference.**

  Erxin Yu, Jing Li, Ming Liao, Siqi Wang, Zuchen Gao, Fei Mi, Lanqing Hong. *The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).* 2024

- **RePALM: Popular Quote Tweet Generation via Auto-Response Augmentation.**

  Erxin Yu, Jing Li, Chunpu Xu. *The 62nd Annual Meeting of the Association for Computational Linguistics (Findings of ACL 2024).* 2024

- **PopALM: Popularity-Aligned Language Models for Social Media Trendy Response Prediction.**

  Erxin Yu, Jing Li, Chunpu Xu. *The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (COLING 2024).* 2024

- **Semantic Textual Matching with Topic-enhanced Discrete Latent Variables.**

  Erxin Yu, Lan Du, Yuan Jin, Zhepei Wei, Yi Chang. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).* 2022

- **Towards Unified Representations of Knowledge Graph and Expert Rules for Machine Learning and Reasoning.**

  Zhepei Wei, Yue Wang, Jinnan Li, Zhining Liu, **Erxin Yu**, Yuan Tian, Xin Wang, Yi Chang. *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (AACL 2022).* 2022

- **ToHRE: A Top-Down Classification Strategy with Hierarchical Bag Representation for Distantly Supervised Relation Extraction.**

Erxin Yu, Wenjuan Han, Yuan Tian, Yi Chang. *in Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*   2020

- **Context and Type Enhanced Representation Learning for Relation Extraction.**

  Erxin Yu, Yantao Jia, Shang Wang, Fengfu Li, Yi Chang. *in 11th IEEE International Conference on Knowledge Graph (ICKG 2020), Nanjing, China, August 2020.*   2020.

- **A Two-Level Noise-Tolerant Model for Relation Extraction with Reinforcement Learning.**

  Erxin Yu, Yantao Jia, Yuan Tian, Yi Chang. *in 11th IEEE International Conference on Knowledge Graph (ICKG 2020)*   2020

Service and Award

- **Journal Reviewer**

  – IEEE Transactions on Knowledge and Data Engineering (TKDE)   2021, 2020, 2019
  – Elsevier Neurocomputing (Neurocomputing)   2021, 2020

- **Conference Reviewer**

  – Association for Computational Linguistics Rolling Review (ACL ARR)   2024, 2023
  – ACM International Conference on Multimedia (ACM MM)   2023
  – International Joint Conference on Artificial Intelligence (IJCAI)   2022
  – Conference on Information and Knowledge Management (CIKM)   2022
  – International Conference on Data Mining (ICDM)   2022
  – Association for the Advancement of Artificial Intelligence (AAAI)   2022, 2021

- **Award**

  – First Prize of "Internet +" University Student Innovation and Entrepreneurship Competition, Jilin University   2018
  – First Prize of University Student Innovation and Entrepreneurship Competition, Jilin university 2018
  – First Class Scholarship, Jilin University   2018
  – First Class Scholarship, Jilin University   2017
  – Second Prize of "Cambridge Cup" National Software and Information Technology, Jilin Province 2017
  – College Excellent Student, Jilin University   2017
  – Second Prize of National College Students Mathematical Modeling Contest   2017