Jiaying Lu

Ph.D. Candidate
Department of Computer Science
Laney Graduate School, Emory University

jiaying.lu AT emory.edu
(+1) (404) 307-8738
lujiaying.github.io

ACADEMIC BACKGROUNDS

Ph.D.   Computer Science & Informatics, Emory University, USA, 2019–2024 (*Expected*)
Dissertation: Learning Structured Knowledge from Real-World Data without Excessive Annotations
Advisor: Dr. Carl Yang

M.Eng.   Electronics & Communication Engineering, Beijing University of Posts and Telecommunications, China, 2014–2017
Exchange student at Aalto University, Finland, Sep 2015–Jan 2016

B.Sc.   Information Engineering, Beijing University of Posts and Telecommunications, China, 2010–2014

GRANTS AND AWARDS

2022   Registration Award. *31st ACM International Conference on Information and Knowledge Management*. Atlanta, GA, USA.

2022   Student Travel Award. *28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Washington, D.C., USA.

2022   Registration Award. *Georgia Statistics Day 2022*. Athens, GA, USA.

2021–24   Emory Professional Development Support Funds. *The James T. Laney School of Graduate Studies*. Atlanta, GA, USA.

PUBLICATIONS

Journal Articles

2023   **Jiaying Lu**, Xiangjue Dong, and Carl Yang. Weakly supervised concept map generation through task-guided graph translation. *IEEE Transactions on Knowledge and Data Engineering*, 35(10): 10871–10883, March 2023a. doi: 10.1109/TKDE.2023.3252588

2023   Wenjing Ma, **Jiaying Lu**, and Hao Wu. Cellcano: supervised cell type identification for single cell atac-seq data. *Nature Communications*, 14(1):1864, April 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-37439-3 (2023 ASA SSGG **Distinguished Student Paper Award**)

Conference Proceedings

2024   **Jiaying Lu**, Jinmeng Rao, Kezhen Chen, Xiaoyuan Guo, Yawen Zhang, Baochen Sun, Carl Yang, and Jie Yang. Evaluation and enhancement of semantic grounding in large vision-language models. In *AAAI 2024 Workshop on Responsible Language Models*, ReLM'24, February 2024a (**Spotlight**)

2024   Zhexiong Liu, Jing Zhang, **Jiaying Lu**, Wenjing Ma, and Joyce Ho. Logicprpbank: A corpus for logical implication and equivalence. In *AAAI 2024 Workshop on AI for Education*, AAAI-AI4ED'24, February 2024b

2023   **Jiaying Lu**, Yongchen Qian, Shifan Zhao, Yuanzhe Xi, and Carl Yang. Mug: A multimodal classification benchmark on game data with tabular, textual, and visual fields. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Findings-EMNLP'23, December 2023b. doi: 10.18653/v1/2023.findings-emnlp.354

2023   Hejie Cui, **Jiaying Lu**, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Chen Ling, Joyce Ho, et al. A survey on knowledge graphs for healthcare: Resources, applications, and promises. In *ICML 2023 Workshop on Interpretable Machine Learning in Healthcare*, IMLH@ICML'23, July 2023

2023   **Jiaying Lu**, Jiaming Shen, Bo Xiong, Wengjing Ma, Staab Steffen, and Carl Yang. Hiprompt: Few-shot biomedical knowledge fusion via hierarchy-oriented prompting. In *46th International ACM SIGIR Conference on Research and Development in Information Retrieval - Short Paper*, SIGIR'23, April 2023c. doi: 10.1145/3539618.3591997

2023   Xiangjue Dong, **Jiaying Lu**, Jianling Wang, and James Caverlee. Closed-book question generation via contrastive learning. In *17th Conference of the European Chapter of the Association for Computational Linguistics*, EACL'23, May 2023. doi: 10.18653/v1/2023.eacl-main.230

2022   **Jiaying Lu** and Carl Yang. Open-world taxonomy and knowledge graph co-learning. In *4th Conference on Automated Knowledge Base Construction*, AKBC'22, November 2022

2022   **Jiaying Lu**, Xin Ye, Yi Ren, and Yezhou Yang. Good, better, best: Textual distractors generation for multi-choice VQA via policy gradient. In *CVPR 2022 Workshop on Open-Domain Retrieval Under a Multi-Modal Setting*, O-DRUM@CVPR'22, June 2022. doi: 10.1109/CVPRW56347.2022.00539

2022   Hejie Cui, **Jiaying Lu**, Yao Ge, and Carl Yang. How can graph neural networks help document retrieval: A case study on cord19 with concept map generation. In *44th European Conference on Information Retrieval - Short Paper*, ECIR'22, April 2022. doi: 10.1007/978-3-030-99739-7_9

2021   **Jiaying Lu** and Jinho D. Choi. Evaluation of Unsupervised Entity and Event Salience Estimation. In *Proceedings of the 34rd International Florida Artificial Intelligence Research Society Conference*, FLAIRS'21, May 2021. doi: 10.32473/flairs.v34i1.128482

Preprint Reports

2024   **Jiaying Lu**, Shifan Zhao, Wenjing Ma, Hui Shao, Xiao Hu, Yuanzhe Xi, and Carl Yang. Uncertainty-aware pre-trained foundation models for patient risk prediction via gaussian process. *preprint*, Feburary 2024b

2024   Darren Liu, Cheng Ding, Delgersuren Bold, Monique Bouvier, **Jiaying Lu**, Benjamin Shickel, Craig S. Jabaley, Wenhui Zhang, Soojin Park, Michael J. Young, Mark S. Wainwright, Gilles Clermont, Parisa Rashidi, Eric S. Rosenthal, Laurie Dimisko, Ran Xiao, Joo Heung Yoon, Carl Yang, and Xiao Hu. Evaluation of general large language models in contextually assessing semantic concepts extracted from adult critical care electronic health record notes. *arXiv:2401.13588*, 2024a

2024   Guangji Bai, Zheng Chai, Chen Ling, Shiyu Wang, **Jiaying Lu**, Nan Zhang, Tingwei Shi, Ziyang Yu, Mengdan Zhu, Yifei Zhang, Carl Yang, Yue Cheng, and Liang Zhao. Beyond efficiency: A systematic survey of resource-efficient large language models. *arXiv:2401.00625*, 2024

2023   Chen Ling, Xujiang Zhao, **Jiaying Lu**, Chengyuan Deng, Can Zheng, Junxiang Wang, Tanmoy Chowdhury, Yun Li, Hejie Cui, Tianjiao Zhao, et al. Beyond one-model-fits-all: A survey of domain specialization for large language models. *arXiv:2305.18703*, 2023

Patents

2023   Jinmeng Rao, **Jiaying Lu**, Kezhen Chen, Jie Yang, Yawen Zhang, Xiaoyuan Guo. "A Simulation-Feedback Frame-work to Mitigate Hallucinations in Multimodal Generative Models". *Filed 2023. X-52882-00-US*.

INVITED TALKS

2024   "Mutual Enhancement of Structured Knowledge and Large Language Model for Health Informatics". **Research Seminar.** Nell Hodgson Woodruff School of Nursing. Atlanta, USA. Jan 9, 2024.

2023   "Retrieval Augmented Generation for Biomedical Entity Alignment via Large Language Models". **Doctoral Symposium.** 2023 Conference on Health, Inference, and Learning. Boston, USA, Jun 24, 2023.

2022   "Learning to Construct Structured Knowledge". **Data Science Forum.** School of Data Science, The Chinese University of Hong Kong. Palo Alto, USA. Dec 17, 2022.

TEACHING EXPERIENCES

2021   Teaching Assistant. *CS570 Data Mining*. Emory University.

2020   Teaching Assistant. *CS253 Data Structure and Algorithms*. Emory University.

2020   Teaching Assistant. *CS170 Introduction to Computer Science I*. Emory University.

PROFESSIONAL EXPERIENCES

2023   **AI Residency**. Mountain View, USA. May–Sep 2023.
         Mineral Research Team, **Google X**.
         Supervisors: Dr. Jinmeng Rao, Dr. Kezhen Chen.

- Data-centric approach for mitigating hallucination in multimodal large language models.

2022   **Applied Scientist Intern**. Sunnyvale, USA. May–Aug 2022.
         Auto Machine Learning Team, **Amazon.com**.
         Supervisors: Dr. Andrew Borthwick, Nick Erickson.

- Multi-modal ensemble models for large-scale product classification problems.

2018–19 **Visiting Scholar**. Phoenix, USA. Aug 2018–Jan 2019.
         Active Perception Group, **Arizona State University**.
         Advisor: Dr. Yezhou Yang.

- Research Interests: Visual Question Answering, Multimodal Learning.

2018   **NLP Engineer**. Beijing, China. Jan–Jul 2018.
         Search Engine Group. **Xiaomi**.

- Word segmentation, named entity recognition, unsupervised synonym pairs mining.

2016–17 **Machine Learning Engineer Full-Time & Intern**. Beijing, China. Oct 2016–Dec 2017.
         Search Advertising Group. **Baidu**.

- Ads content generation, Query-Ads matching.

2016   **Machine Learning Engineer Intern**. Beijing, China. Jul–Sep 2016.
         Growth Hacking Team. **Meituan**.

- Gradient boosted regression trees for new customer coupon distribution prediction.

2016   **Machine Learning Engineer Intern**. Beijing, China. Mar–Jun 2016.
         Online Food Delivery Group. **Baidu**.

- Logistic regression model for customer CTR prediction.

2014–15 **Software Engineer Intern**. Beijing, China. Sep 2014–Feb 2015.
         Beijing YunJiang Technology Co., Ltd.

- Web & mobile application backend development.

SERVICES

Academic Event Organizations

2022   Web Developer Student Lead. *28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (**KDD**).

Journal Peer Reviews

2023–   The Journal of Agricultural, Biological and Environmental Statistics (**JABES**).

2022–   The IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**).

2021   The IEEE Transactions on Knowledge and Data Engineering (**TKDE**).

2021   The IEEE Transactions on Big Data (**TBD**).

Conference Peer Reviews

2024   The Association for Computational Linguistics Rolling Review (**ACL Rolling Review**).

2023   The ACM Special Interest Group on Information Retrieval (**SIGIR**).

2021-   The Conference on Empirical Methods in Natural Language Processing (**EMNLP**).

2023   The IEEE International Conference on Data Mining (**ICDM**).

2023   The IEEE International Conference on Bioinformatics and Biomedicine (**BIBM**).

2023   The Conference on Machine Learning for Health (**ML4H**)

2022-   The International World Wide Web Conference (**WWW**).

2022   The ACM Conference on Knowledge Discovery and Data Mining (**KDD**).

2022   The AAAI Conference on Artificial Intelligence (**AAAI**).

2021   The ACM International Conference on Information and Knowledge Management (**CIKM**).

2021   The International Joint Conference on Natural Language Processing (**IJCNLP**).

2020   The International Joint Conferences on Artificial Intelligence (**IJCAI**).

2019-23 The IEEE International Conference on Robotics and Automation (**ICRA**).

Symposium & Workshop Peer Reviews

2024   AAAI Spring Symposium on Clinical Foundation Models.

2023   NeurIPS Workshop on Socially Responsible Language Modelling Research (**SoLaR**).

2023   ICLR Workshop on Trustworthy Machine Learning for Healthcare (**TML4H**).

2023   ACL SIGEDU Workshop on Innovative Use of NLP for Building Educational Applications (**BEA**).

2021-   ICML Workshop on Interpretable Machine Learning in Healthcare (**IMLH**).

2021-   ICCV Workshop on Computer Vision for Automated Medical Diagnosis (**CVAMD**).