Yongkang Li

Mail: y.li7@uva.nl **Phone:**(+31) 0635630756 **Home Page**
Research Interests: Robust AI, Information Retrieval, NLP, Data Mining, Recommender System

Education Backgrounds

- **University of Amsterdam, Informatics Institute**

  Lab: *Information Retrieval Lab* **PhD**, Oct 2023–June 2027(Expected), Amsterdam, The Netherlands

  Supervisors: Prof. dr. Evangelos Kanoulas and dr. Panagiotis Eustratiadis, focusing on robust IR.

- **Southern University of Science and Technology, Department of Computer Science and Engineering**

  Lab: *SUSTech-UTokyo Joint Research Center on Super Smart City* **Master**, 2020–2023, Shenzhen, China

  Supervisors: Prof. Xuan Song and Prof. Zipei Fan at the University of Tokyo

- **Beijing University of Posts and Telecommunications, School of Information and Communication Engineering**

  Lab: *Pattern Recognition and Intelligent Systems (PRIS) Laboratory* **Bachelor**, 2016–2020, Beijing, China

Publications

Papers - First Author or Co-first Author

- Panagiotis Eustratiadis*, **Yongkang Li***, Jingfen Qiao*, Vaishali Pal, Yougang Lyu and Evangelos Kanoulas, *Retrieving Expertise in the Age of Agentic Information Access*. [J]// Under Review in **SIGIR-AP'2025**: The 3rd International ACM SIGIR Conference on Information Retrieval in the Asia Pacific. Date Submitted: 23-July-2025.
- Dominykas Seputis*, **Yongkang Li***, Karsten Langerak and Serghei Mihailov, *Rethinking the Privacy of Text Embeddings: A Reproducibility Study of "Text Embeddings Reveal (Almost) As Much As Text"*. [C]// Accepted by **RecSys'2025**: The 19th ACM Conference on Recommender Systems, reproducibility paper. Date Accepted: 04-July-2025.
- **Yongkang Li**, Panagiotis Eustratiadis, Simon Lupart, Evangelos Kanoulas, *Unsupervised Corpus Poisoning Attacks in Continuous Space for Dense Retrieval*. [C]// Published by **SIGIR'2025**: The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, full paper. [Paper Link] [GitHub]
- **Yongkang Li**, Panagiotis Eustratiadis, Evangelos Kanoulas, *Reproducing HotFlip for Corpus Poisoning Attacks in Dense Retrieval*. [C]// Published by **ECIR'2025**: Proceedings of the 47th European Conference on Information Retrieval. [Paper Link] [GitHub] **Best Reproducibility Paper Award**
- **Yongkang Li**, Zipei Fan, Xuan Song, *Heterogeneous Hyperbolic Hypergraph Neural Network for Friend Recommendation in Location-based Social Network*. [J]// Published by **TKDD Journal**: ACM Transactions on Knowledge Discovery from Data, 2024.[Paper Link] [GitHub]
- **Yongkang Li**, Zipei Fan, Jixiao Zhang, Dengheng Shi, Tianqi Xu, Du Yin, Jinliang Deng, Xuan Song, *Heterogeneous Hypergraph Neural Network for Friend Recommendation with Human Mobility*. [C]// Published by **CIKM'2022**: Proceedings of the 31st ACM International Conference on Information and Knowledge Management, 2022: 4209-4213 [Paper Link] [GitHub]
- **Yongkang Li**, Zipei Fan, Du Yin, Renhe Jiang, Jinliang Deng, Xuan Song, *HMGCL: Heterogeneous Multigraph Contrastive Learning for LBSN Friend Recommendation*. [J]// Published by **World Wide Web Journal** , 2022 [Paper Link] [GitHub]

Papers - Co-Author

- Yuxi Lin, **Yongkang Li**, Jie Xing, Zipei Fan, *Multifaceted Scenario-Aware Hypergraph Learning for Next POI Recommendation*. [J]// Under Review in **AAAI'2026**: The 40th Annual AAAI Conference on Artificial Intelligence. Date Submitted: 31-July-2025.
- Yuyue Zhao, **Yongkang Li**, Jin Huang, Xiang Wang, Maarten de Rijke, *Unseen Threats: Media Bias-Aware Textual Attacks on News Recommender Systems*. [J]// Under Review in **TOIS**: Transactions on Information Systems. Date Submitted: 25-July-2025.
- Jixiao Zhang, **Yongkang Li**, Ruotong Zou, Jingyuan Zhang, Renhe Jiang, Zipei Fan, Xuan Song, *Hyper-relational knowledge graph neural network for next POI recommendation*. [J]// Published by **World Wide Web Journal** , 2024 (JCR Q2, CCF Rank:B, IF=3.0) [Paper Link] [GitHub]

– Du Yin, Renhe Jiang, Jiewen Deng, **Yongkang Li**, Yi Xie, Zhongyi Wang, Yifan Zhou, Xuan Song, Jedi S Shang. *MTMGNN: Multi-time multi-graph neural network for metro passenger flow prediction*. [J]// Published by **Geoinformatica Journal**, 2022: 1573-7624. (JCR Q3, CCF Rank:B, IF=2.7) [Paper Link] [GitHub]

Main Research Projects

– **Corpus Poisoning Attacks in Dense Retrieval**
Our research explores corpus poisoning attacks in dense retrieval systems. In ***ECIR'2025***, we reproduced and extended HotFlip, a gradient-based word substitution attack originally designed for language models. We significantly improved its efficiency, reducing adversarial document generation time from 4 hours to 15 minutes, and demonstrated its effectiveness across multiple dense retrievers. We further analyzed its limitations in black-box and query-agnostic settings. Building on this, in ***SIGIR'2025*** we proposed an unsupervised optimization method that operates directly in the embedding space, generating adversarial documents without query knowledge. This method achieved a dramatic speed-up (5 minutes per document vs. 4 hours) and improved retriever robustness through adversarial training with minimal performance loss. Experiments on the BEIR benchmark, under both white-box and black-box settings, confirmed the vulnerability of a range of retrievers and the practical value of our approach.

– **Heterogeneous Graph Learning for Location-based Social Network Friend Recommendation**
This research, part of my master's studies, focuses on friend recommendation in location-based social networks (LBSN) using user trajectory data. We explored the impact of social network relationships on human mobility, particularly long-distance travel. Our approach used heterogeneous multi-graphs and hypergraphs, with novel neural network models for learning and recommendation. The solution achieved state-of-the-art results on multiple real-world datasets and has been published in **TKDD**, **CIKM**, and **WWWJ**.

Work & Intern Experiences

Temu Inc. - Research and Development Department Shanghai, China
Machine Learning Algorithm Engineer Jul 2023–Sep 2023

– **Conversational Intent Recognition**
On e-commerce platforms, customers typically have specific intents when communicating with customer service, such as requesting a return, a refund, or inquiring about shipping status. The goal of this project is to assist customer service in quickly identifying the user's intent during conversations, enabling better and faster service. We integrate deep learning methods with traditional feature-matching techniques, combining customer purchase history, current order status, and cleaned dialogue data. This approach has enabled us to develop an automated classification and statistical system that assists customer service agents and supports data analysis for other internal business departments.

Pinduoduo Inc. - Research and Development Department Shanghai, China
Machine Learning Internship Jul 2022–Sep 2022

– **Knowledge Graph Embedding**
Investigate the development history and typical technologies of knowledge graph embedding models (such as TransE, TransR, DisMult, ComplEx, KG-Bert, RotatE, and PairRE), implement the corresponding codes and do academic talk with colleagues across the department.

– **Aspect-based Sentiment Analysis**
Based on the e-commerce review data, I implemented the aspect-based sentiment analysis module via PaddleNLP, which extracts a quadruple like (attribute, opinion, sentiment, attribute category) from consumer review data. And this module has been applied to the company business.

Awards & Skills

English:
– TOEFL-107, Reading 29, Listening 29, Speaking 23, Writing 26

Professional Coding:
– Python, C++, Linux, Git and good coding skills.
– Familiar with PyTorch, DGL, PyG, Pandas, NumPy, Scikit-learn, TensorFlow, etc.

Self-evaluation

I am very curious about new things and thrive on taking challenges under pressure. I also enjoy teamwork and take great responsibility for every project.