YI YANG

CONTACT INFORMATION	Department of Information Systems, Business Statistics, and Operations Management, School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

imyiyang@ust.hk

EMPLOYMENT

Hong Kong University of Science and Technology

Lee Heng Fellow, School of Business and Management	July 2023-present
Associate Professor, School of Business and Management	July 2023-present
Assistant Professor, School of Business and Management,	July 2017-June 2023
Director, Center for Business and Social Analytics,	July 2024-present

EDUCATION

Northwestern University, Evanston, Illinois
Ph.D. in Computer Science — Aug 2015

University of Science and Technology of China, Anhui, China
B.S. in Computer Science — July 2009

RESEARCH INTERESTS

Method Area: Large Language Model, Machine Learning, Natural Language Processing, Deep Learning
Applied Area: Business and Finance

JOURNAL PUBLICATIONS

1. **Yi Yang**, Yixuan Tang, Yangyang Fan, and Kunpeng Zhang. Conversation Modeling: Analyzing Earnings Conference Calls for Financial Risk Prediction. *Information Systems Research*, Accepted

2. Yi He, **Yi Yang**, Defu Lian, and Kunpeng Zhang. Divide and Contrast: A Text-Based Method for Firm Market Risk Prediction. *INFORMS Journal on Computing*, 2025

3. **Yi Yang**, John P Lalor, Ahmed Abbasi, and Daniel Dajun Zeng. Hierarchical Deep Document Model. *IEEE Transactions on Knowledge and Data Engineering*, 2024

4. John P Lalor, Ahmed Abbasi, Kezia Oketch, **Yi Yang**, and Nicole Forsgren. Should Fairness be a Metric or a Model? A Model-based Framework for Assessing Bias in Machine Learning Pipelines. *ACM Transactions on Information Systems*, 2024

5. Dongcheng Zhang, Kunpeng Zhang, **Yi Yang**, and David A Schweidel. TM-OKC: An Unsupervised Topic Model for Text in Online Knowledge Communities. *MIS Quarterly*, 48(3), 2024

6. **Yi Yang** and Ramanath Subramanyam. Extracting Actionable Insights from Text Data: a Stable Topic Model Approach. *MIS Quarterly*, 47(3), 2023

7. Allen H Huang, Hui Wang, and **Yi Yang**. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2):806–841, 2023

8. **Yi Yang**, Yu Qin, Yangyang Fan, and Zhongju Zhang. UNLOCKING THE POWER OF VOICE FOR FINANCIAL RISK PREDICTION: A THEORY-DRIVEN DEEP LEARNING DESIGN APPROACH. *MIS Quarterly*, 47(1), 2023

9. **Yi Yang**, Kunpeng Zhang, and Yangyang Fan. sdtm: A supervised bayesian deep topic model for text analytics. *Information Systems Research*, 34(1):137–156, 2023

10. Zhuoyi Peng, **Yi Yang**, Liu Yang, and Kai Chen. Federated Meta Embedding Concept Stock Recommendation. *IEEE Transactions on Big Data*, 2022

11. Zhe Tang, **Yi Yang**, Wen Li, Defu Lian, and Lixin Duan. Deep Cross-Attention Network for Crowdfunding Success Prediction. *IEEE Transactions on Multimedia*, 2022

12. **Yi Yang**, Kunpeng Zhang, and PK Kannan. Identifying market structure: A deep network representation learning of social engagement. *Journal of Marketing*, 86(4):37–56, 2022

13. **Yi Yang**, Kunpeng Zhang, and Yangyang Fan. Analyzing Firm Reports for Volatility Prediction: A Knowledge-Driven Text-Embedding Approach. *INFORMS Journal on Computing*, 34(1):522–540, 2022

14. Fan Zhou, Kunpeng Zhang, Bangying Wu, **Yi Yang**, and Harry Jiannan Wang. Unifying online and offline preference for social link prediction. *INFORMS Journal on Computing*, 33(4):1400–1418, 2021

15. **Yi Yang**, Shimei Pan, Jie Lu, Mercan Topkara, and Yangqiu Song. The stability and usability of statistical topic models. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6(2):1–23, 2016

16. Tiantian Zhu, Hongyu Gao, **Yi Yang**, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, and Alok N Choudhary. Beating the artificial chaos: Fighting OSN spam using its own templates. *IEEE/ACM Transactions on Networking*, 24(6):3856–3869, 2016

17. Fei Huang, Arun Ahuja, Doug Downey, **Yi Yang**, Yuhong Guo, and Alexander Yates. Learning representations for weakly supervised natural language processing tasks. *Computational Linguistics*, 40(1):85–120, 2014

CONFERENCE PUBLICATIONS

1. Hongyi Tang, Zhihao Zhu, and **Yi Yang**. Identifying Pre-training Data in LLMs: A Neuron Activation-Based Detection Framework. *EMNLP 2025*

2. Yixuan Tang and **Yi Yang**. Finmteb: Finance massive text embedding benchmark. *EMNLP 2025*

3. Jiaxin Liu, Yixuan Tang, **Yi Yang**, and Kar Yan Tam. Evaluating and aligning human economic risk preferences in llms. *EMNLP 2025*

4. Jiaqi Li, Yixuan Tang, and **Yi Yang**. Know the unknown: An uncertainty-sensitive method for llm instruction tuning. *Findings of ACL 2025*

5. Yubai Wei, Jiale Han, and **Yi Yang**. Adapting General-Purpose Embedding Models to Private Datasets Using Keyword-based Retrieval. *Findings of ACL 2025*

6. Zhuoyi Peng, Yixuan Tang, and **Yi Yang**. Adversarial Mixup Unlearning. *ICLR 2025*

7. Zhihao Zhu, **Yi Yang**, and Defu Lian. TDDBench: A Benchmark for Training data detection. *ICLR 2025*

8. Yue Guo, Wentao Zhang, Xiaojun Zhang, Vincent W Zheng, and **Yi Yang**. Efficient Multi-Expert Tabular Language Model for Banking. *KDD 2025*

9. Hanyu Duan, Yixuan Tang, **Yi Yang**, Ahmed Abbasi, and Kar Yan Tam. Exploring the Relationship between In-Context Learning and Instruction Tuning. *Findings of EMNLP 2024*

10. Yixuan Tang and **Yi Yang**. MultiHop-RAG: Benchmarking Retrieval-Augmented Generation for Multi-Hop Queries. *COLM 2024*

11. Yue Guo and **Yi Yang**. EconNLI: Evaluating Large Language Models on Economics Reasoning. *Findings of ACL 2024*

12. Jiaxin Liu, **Yi Yang**, and Kar Yan Tam. Beyond Surface Similarity: Detecting Subtle Semantic Shifts in Financial Narratives. *Findings of NAACL 2024*

13. Zhuoyi Peng and **Yi Yang**. Connecting the Dots: Inferring Patent Phrase Similarity with Retrieved Phrase Graphs. *Findings of NAACL 2024*

14. Yue Guo, Chenxi Hu, and **Yi Yang**. Predict the Future from the Past? On the Temporal Data Distribution Shift in Financial Sentiment Classifications. *EMNLP 2023*

15. Yue Guo, Zian Xu, and **Yi Yang**. Is ChatGPT a Financial Expert? Evaluating Language Models on Financial Natural Language Processing. ***Findings of EMNLP 2023***

16. Yixuan Tang, **Yi Yang**, Allen Huang, Andy Tam, and Justin Tang. FinEntity: Entity-level Sentiment Classification for Financial Texts. ***EMNLP 2023***

17. Fan Zhou, Yuzhou Mao, Liu Yu, **Yi Yang**, and Ting Zhong. Causal-Debias: Unifying Debiasing in Pretrained Language Models and Fine-tuning via Causal Invariant Learning. ***ACL 2023***

18. Yue Guo, **Yi Yang**, and Ahmed Abbasi. Auto-Debias: Debiasing Masked Language Models with Automated Biased Prompts. ***ACL 2022***

19. Chengyu Chuang and **Yi Yang**. Buy Tesla, Sell Ford: Assessing Implicit Stock Market Preference in Pre-trained Language Models. ***ACL 2022***

20. John P Lalor, **Yi Yang**, Kendall Smith, Nicole Forsgren, and Ahmed Abbasi. Benchmarking intersectional biases in NLP. ***NAACL 2022***

21. Zhenyu He, Ce Li, Fan Zhou, and **Yi Yang**. Rumor detection on social media with event augmentations. ***SIGIR 2021***

22. Ahmed Abbasi, David Dobolyi, John P Lalor, Richard G Netemeyer, Kendall Smith, and **Yi Yang**. Constructing a psychometric testbed for fair natural language processing. ***EMNLP 2021***

23. Hanyu Duan, **Yi Yang**, and Kar Yan Tam. Learning Numeracy: A Simple Yet Effective Number Embedding Approach Using Knowledge Graph. ***Findings of EMNLP 2021***

24. Linyi Yang, Eoin Kenny, Tin Lok James Ng, **Yi Yang**, Barry Smyth, and Ruihai Dong. Generating Plausible Counterfactual Explanations for Deep Transformers in Financial Text Classification. ***COLING 2020***

25. Xinyi Wang and **Yi Yang**. Neural topic model with attention for supervised learning. ***AISTATS 2020***

26. Ting Zhong, Tianliang Wang, Fan Zhou, Goce Trajcevski, Kunpeng Zhang, and **Yi Yang**. Interpreting Twitter user geolocation. ***ACL 2020***

27. Fan Zhou, Shengming Zhang, and **Yi Yang**. Interpretable operational risk classification with semi-supervised variational autoencoder. ***ACL 2020***

28. Yu Qin and **Yi Yang**. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. ***ACL 2019***

29. Fan Zhou, Bangying Wu, **Yi Yang**, Goce Trajcevski, Kunpeng Zhang, and Ting Zhong. Vec2link: Unifying heterogeneous data for social link prediction. ***CIKM 2018***

30. Doug Downey, Chandra Bhagavatula, and **Yi Yang**. Efficient methods for inferring large sparse topic hierarchies. ***ACL 2015***

31. **Yi Yang**, Shimei Pan, Yangqiu Song, Jie Lu, Mercan Topkara, and JW Player. Improving Topic Model Stability for Effective Document Exploration. ***IJCAI 2016***

32. **Yi Yang**, Doug Downey, and Jordan Boyd-Graber. Efficient Methods for Incorporating Knowledge into Topic Models. ***EMNLP 2015***

33. **Yi Yang**, Shimei Pan, Yangqiu Song, Jie Lu, and Mercan Topkara. User-directed non-disruptive topic model update for effective exploration of dynamic content. ***IUI 2015***

34. **Yi Yang**, Alexander Yates, and Doug Downey. Overcoming the memory bottleneck in distributed training of latent variable models of text. ***NAACL 2013***

GRANTS	**Principal Investigator** (PI)

1. HKUST Strategic Fund for Futuristic Business Research, HKD $2,500,000, 2025-2027
   Title: *The Triad of the Future: AI, Humans, and Organizations*

2. Hong Kong Research Grants Council, General Research Fund (GRF), HKD $661,000, 2025-2027
   Title: *From Inquiry to Purchase: Optimizing Large Language Model-Based Conversational AI for Enhanced Customer Conversion*

3. HKUST-WeBank Joint Laboratory Project (WeBank), HKD $500,000, 2025-2027
   Title: *Key Technologies for Enhanced Case Retrieval in Anti-Money Laundering Systems*

4. HKUST-WeBank Joint Laboratory Project (WeBank), HKD $500,000, 2025-2026
   Title: *Key Technologies for Combating GenAI-based Financial Fraud*

5. Ping An Insurance Group, RMB $1,000,000, 2025-2026
   Title: *Generative AI in Asset Management*

6. HKUST-WeBank Joint Laboratory Project (WeBank), HKD $580,608, 2024-2026
   Title: *Multimodal Financial Large Language Model for Stock Market Prediction*

7. Hong Kong Research Grants Council, General Research Fund (GRF), HKD $701,000, 2024-2027
   Title: *How Generative AI Impacts Financial Analysis*

8. HKUST-WeBank Joint Laboratory Project (WeBank), HKD $600,000, 2023-2024
   Title: *Financial Large Language Model As-A-Service: Development and Applications*

9. HKUST-WeBank Joint Laboratory Project (WeBank), HKD $497,000, 2022-2023
   Title: *Pre-training for Tabular Business Data*

10. HKUST-Kaisa Joint Research Institute Seed Fund, HKD $670,000, 2021-2022
    Title: *Deep Learning in Financial NLP*

11. Hong Kong Research Grants Council, Early Career Scheme (ECS), HKD $458,000, 2019-2022
    Title: *Beyond Text: The Information Role of Voice in Financial Market*

Co-Principal Investigator (Co-PI)
1. Hong Kong Research Grants Council, Theme-based Research Scheme, HKD $5,624,000, 2024
   Title: Strengthening Hong Kong as a Global Financial Center by Enhancing the Overall Regtech Capacity of the Industry and Facilitating the Regulation and Infrastructure Design of CBDC and Digital Assets in Hong Kong

Co-Investigator (Co-I)
1. Hong Kong Research Grants Council, Theme-based Research Scheme, HKD $20,659,000, 2018-2022
   Title: Contributing to the Development of Hong Kong into a Global Fintech Hub

2. Hong Kong Research Grants Council, General Research Fund, HKD $605,216, 2022-2023
   Title: Product Recommendations in Email Campaigns for Passive Customers

WORK EXPERIENCE	**University of Illinois Urbana-Champaign**, Urbana, Illinois	
	Postdoc Research Associate, College of Business	Aug 2015-June 2017
	Amazon, Seattle, Washington	
	Software Engineer Intern, Amazon Web Services	Sep 2013-Dec 2013
	IBM T.J. Watson Research, Yorktown Heights, New York	
	Research Intern, Intelligent Multimedia Interaction Division	May 2013-Sep 2013

TEACHING   **Hong Kong University of Science and Technology** ,

Instructor,   July 2017-present

    ISOM3360: Data Mining for Business Analytics, (Undergradaute)
    ISOM3370: Big Data Technologies, (Undergradaute)
    ISOM5270: Big Data Analytics (MBA full-time and part-time)
    ISOM5270: Big Data Analytics (Master)

    ★ Finalist for *The Franklin Prize for Teaching Excellence* for Year 2024
    ★ Finalist for *The Franklin Prize for Teaching Excellence* for Year 2023
    ★ Finalist for *The Franklin Prize for Teaching Excellence* for Year 2021
    ★ Finalist for *The Franklin Prize for Teaching Excellence* for Year 2018

University of Illinois Urbana-Champaign, Urbana, Illinois

Instructor,   Aug 2015-June 2017

    BADM352: Database Design and Management

    ★ Awarded *Teachers Ranked as Excellent by Their Students* in Spring 2017 and Fall 2016.

STUDENTS
SUPERVISION   **Postgraduate students**:
  - Hanyu Duan, PhD Candidate, Co-supervisor (with Prof. Kar Yan Tam at ISOM), 2021-present
  - Zhuoyi Peng, PhD Candidate, Supervisor, 2023-present
  - Yixuan Tang, PhD Candidate, Supervisor, 2024-present
  - Austin Cheung, PhD Candidate, Supervisor, 2025-present
  - Jingzhou Jiang, PhD Candidate, Co-supervisor (with Prof. Kar Yan Tam at ISOM), 2025-present

Graduated students:
  - Yue Guo, PhD, Co-supervisor (with Prof. Yangqiu Song at CSE), Placement: Tencent
  - Jiaxin Liu, PhD, Co-supervisor (with Prof. Kar Yan Tam at ISOM), Placement: Barclays Capital
  - Yingpeng Zhu, PhD, ISOM, Supervisor, Placement: University of Macao

SERVICE   **University Service**
    PhD/MPhil Committee (IS), 2023-2025
    HKUST JUPAS Interview Taster class, 2023
    FT/PT MScISM Committee (Information Systems), 2022-2025
    Search and Appointments Committee (Information Systems), 2021-2022
    UG committee (Information Systems), 2020-2021
    Technology Review Committee (TRC), 2022-present

    Editorial Board
    Guest Associate Editor, **Information Systems Research**, 2025-present
    Editorial Review Board, **Information Systems Research**, 2024-present
    Associate Editor, **MIS Quarterly**, Special Issue on The Institutional Press in the Digital Age, 2024
    Associate Editor, **INFORMS Journal on Computing**, 2024-present
    Associate Editor, Information Technology and Management, 2020-present