

MINGZHE DU

(+65) 9658 2486 | dumingzhex@gmail.com | <https://mingzhe.space>
Innovation 4.0, 3 Research Link, Singapore, 117602

RESEARCH INTERESTS

Code Generation / Preference Alignment / Bias Evaluation and Mitigation

EDUCATION

Nanyang Technological University <i>PhD Candidate in Computer Science, College of Computing and Data Science</i>	Singapore Aug. 2022 – Present
The University of Melbourne <i>Master of Information Technology, Faculty of Engineering</i>	Melbourne, Australia Oct. 2017 – Dec. 2019
• Advisor: Prof. Anh Tuan Luu	
• Advisor: Prof. Richard O. Sinnott	
• Specialization: Distributed Computing (Honorable Dissertation)	
Central South University <i>Bachelor of Information Security, School of Computer Science and Engineering</i>	Changsha, China Sep. 2013 – Jun. 2017
• Excellent Scholarships (GPA: 3.7/4.0)	

PROFESSIONAL EXPERIENCE

National University of Singapore, Institute of Data Science	Aug. 2021 – Present
• Research Associate (Advisor: Prof. See-Kiong Ng)	
• Affiliated with Cisco-NUS Accelerated Digital Economy Corporate Laboratory.	
ByteDance, Tiktok Search	Dec. 2019 – Aug. 2021
• Software Architect (Full Time)	
• Architected a scalable, general-purpose search engine (<i>Search Eternal</i>) that now powers core search services for major ByteDance platforms, including <i>Douyin</i> , <i>TikTok</i> and <i>TouTiao</i> .	
Microsoft, Asia-Pacific Research and Development Group	Nov. 2018 – Mar. 2019
• Research Intern (Mentor: Wei Liu)	
• Designed and implemented a Transformer-based model for text-to-image synthesis, generating visually coherent images from textual descriptions.	
Alibaba, Ant Finance Core Clearing Team	Jul. 2016 – Oct. 2016
• Software Engineer Intern	
• Developed components for a core system in Alipay's payment clearing flow, enabling the aggregation of millions of orders at minute-level latency for data exchange with banking partners.	
Tencent, Social Network Group	Mar. 2016 – Jun. 2016
• Software Engineer Intern	
• Built and deployed a URL analysis platform to process and detect suspicious content in high-volume public message traffic, significantly improving the filtering of malicious information to enhance content security.	

ACADEMIC SERVICES

- **Conference Reviewer:** *NeurIPS* (24,25), *ICLR* (24,25,26), *ICML* (25,26), *AISTATS* (24,26), *AAAI'26*, *AISI'26*, *CIKM'25*, *ECAI'25*, *NAACL'24*, *EMNLP'24*, *ECAI'24*, *CLEF'23*.
- **Journal Reviewer:** *ACM Computing Survey*, *Frontiers in Artificial Intelligence*.
- **Volunteer:** *NeurIPS'25*, *ACL'25*, *WWW'24*, *EMNLP'24*, *AAAI'23*.
- **STEM Mentor:** *The University of Melbourne* (2019-Present).

SELECTED PUBLICATIONS

- [NeurIPS'25] [Afterburner: Reinforcement Learning Facilitates Self-Improving Code Efficiency Optimization.](#)
Mingzhe Du, Luu Anh Tuan, Yue Liu, Yuhao Qing, Dong Huang, Xinyi He, Qian Liu, Zejun Ma, See-kiong Ng
- [ACL'25] [CodeArena: A collective evaluation platform for LLM code generation.](#)
Mingzhe Du, Luu Anh Tuan, Bin Ji, Xiaobao Wu, Dong Huang, Terry Yue Zhuo, Qian Liu, See-Kiong Ng
- [NeurIPS'24] [Mercury: A Code Efficiency Benchmark for Code Large Language Models.](#)
Mingzhe Du, Luu Anh Tuan, Bin Ji, Qian Liu, See-Kiong Ng

PUBLICATIONS

- [Preprint] [Nexus: Execution-Grounded Multi-Agent Test Oracle Synthesis](#)
Dong Huang, **Mingzhe Du[†]**, Jie M. Zhang, Zheng Lin, Meng Luo, Qianru Zhang, See-kiong Ng.
- [Preprint] [SWE-Perf: Can Language Models Optimize Code Performance on Real-World Repositories?](#)
Xinyi He, Qian Liu, **Mingzhe Du**, Lin Yan, Zhijie Fan, Yiming Huang, Zejian Yuan, and Zejun Ma.
- [Preprint] [Benchmarking LLMs for Unit Test Generation from Real-World Functions.](#)
Dong Huang, Jie M. Zhang, Mark Harman, Qianru Zhang, **Mingzhe Du[†]**, and See-Kiong Ng.
- [EACL'26] [Pro-QuEST: Prompt-chaining Quiz Engine for testing Specialized Technical Product Knowledge](#)
Sujatha Das Gollapalli, Mouad Hakam, **Mingzhe Du**, See-Kiong Ng, and Mohammed Hamzeh.
- [AMiya'26] [Improving Arabic Dialectness in LLMs with Reinforcement Learning](#)
Sujatha Das Gollapalli, Mouad Hakam, **Mingzhe Du**, See-Kiong Ng, and Mohammed Hamzeh.
- [ICLR'26] [Beyond Prompt-Induced Lies: Investigating LLM Deception on Benign Prompts.](#) (Oral)
Zhaomin Wu, **Mingzhe Du**, See-Kiong Ng, and Bingsheng He.
- [EMNLP'25] [On Assigning Product and Software Codes to Service Requests with Large Language Models.](#)
Sujatha Das Gollapalli, Mouad Hakam, **Mingzhe Du**, See-Kiong Ng, and Mohammed Hamzeh.
- [NeurIPS'25] [EffiBench-X: A Multi-Language Benchmark for Measuring Efficiency of LLM-Generated Code.](#)
Yuhao Qing*, Boyu Zhu*, **Mingzhe Du***, Zhijiang Guo, Terry Yue Zhuo, Qianru Zhang, Jie M. Zhang, Heming Cui, Siu-Ming Yiu, Dong Huang, See-Kiong Ng, and Luu Anh Tuan.
- [NeurIPS'25] [Guardreasoner-VL: Safeguarding VLMs via Reinforced Reasoning.](#)
Yue Liu, Shengfang Zhai, **Mingzhe Du**, Yulin Chen, Tri Cao, Hongcheng Gao, Cheng Wang, Xinfeng Li, Kun Wang, Junfeng Fang, Jiaheng Zhang, and Bryan Hooi.
- [PRAL@ICML'25] [Afterburner: Reinforcement Learning Facilitates Self-Improving Code Efficiency Optimization.](#)
Mingzhe Du, Luu Anh Tuan, Yue Liu, Yuhao Qing, Dong Huang, Xinyi He, Qian Liu, Zejun Ma, and See-kiong Ng.
- [CFAgentic@ICML'25] [Collective Bias Mitigation via Model Routing and Collaboration.](#)
Mingzhe Du, Luu Anh Tuan , Xiaobao Wu, Yichong Huang, Yue Liu, Dong Huang, Huijun Liu, Bin Ji, Jie M. Zhang, and See-Kiong Ng.
- [ICSE'25] [Measuring the Influence of Incorrect Code on Test Generation.](#)
Dong Huang, Jie M. Zhang, Mark Harman, **Mingzhe Du[†]**, and Heming Cui.
- [ACL'25] [AntiLeakBench: Preventing Data Contamination by Automatically Constructing Benchmarks with Updated Real-World Knowledge.](#) (Oral and SAC Highlight Award)
Xiaobao Wu, Liangming Pan, Yuxi Xie, Ruiwen Zhou, Shuai Zhao, Yubo Ma, **Mingzhe Du**, Rui Mao, Anh Tuan Luu, and William Yang Wang.

[ICML'25] Position: Current Model Licensing Practices are Dragging Us into a Quagmire of Legal Noncompliance. (Oral)

Moming Duan*, **Mingzhe Du***, Rui Zhao, Mengying Wang, Yinghui Wu, Nigel Shadbolt, and Bingsheng He.

[AAAI'25] Towards Verifiable Text Generation with Generative Agent. (Oral)

Ji, Bin, Huijun Liu, **Mingzhe Du**, Shasha Li, Xiaodong Liu, Jun Ma, Jie Yu, and See-Kiong Ng.

[SAC'25] Curriculum Demonstration Selection for In-Context Learning.

Duc Anh Vu, Cong-Duy Nguyen, Xiaobao Wu, Nhat Hoang, **Mingzhe Du**, Thong Nguyen, and Anh Tuan Luu.

[JMIR'25] Unraveling Online Mental Health Through the Lens of Early Maladaptive Schemas: AI-Enabled Content Analysis of Online Mental Health Communities.

Beng Heng Ang, Sujatha Das Gollapalli, **Mingzhe Du**, and See-Kiong Ng.

[AAAI'24] From Static to Dynamic: Knowledge Metabolism for Large Language Models.

Mingzhe Du, Anh Tuan Luu, Bin Ji, and See-Kiong Ng.

[AAAI'24] Chain-of-Thought Improves Text Generation with Citations in Large Language Models.

Bin Ji, Huijun Liu, **Mingzhe Du**, and See-Kiong Ng.

[ECAI'24] Counseling Responses for Mental Health Forum Questions with Early Maladaptive Schema Prediction.

Sujatha Das Gollapalli, Beng Heng Ang, **Mingzhe Du**, and See-Kiong Ng.

[TLDK'23] Constituency-Informed and Constituency-Constrained Extractive Question Answering with Heterogeneous Graph Transformer.

Mingzhe Du, Mouad Hakam, See-Kiong Ng, and Stéphane Bressan.

[WWW'23] Identifying Checkworthy Cure Claims on Twitter.

Sujatha Das Gollapalli, **Mingzhe Du**, and See-Kiong Ng.

[AAAI'23] Generating Reflective Questions for Engaging Gallery Visitors in ArtMuse.

Sujatha Das Gollapalli, **Mingzhe Du**, and See-Kiong Ng.

[CLEF'22] NUS-IDS at CheckThat! 2022: Identifying Check-worthiness of Tweets using CheckthaT5.

Mingzhe Du, Sujatha Das Gollapalli, and See-Kiong Ng.

† Corresponding Author, * Equal Contribution.

AWARDS

- First Runner-Up Award at the **Singapore Healthcare AI Datathon 2022**: *Identifying ICU Patient with Central Lines at Risk of CLABSI*.
- First Prize at **Catalyst 2019**: *A Path Recommendation System based on Data Aggregation and Analysis*.
- First Prize at **Codebrew 2019**: *After Dark* (Finding Safe and Fast Routes for Pedestrians at Night).
- First Prize at the **Mathematical Contest in Modeling (MCM, 2014)**: *Study on the Wreckage Position of Malaysia Airlines Flight MH370*.
- Second Prize at the **8th National Information Security Competition (2015)**: *An Efficient Express Label System Based on Hierarchical Encryption*.
- Second Prize at the **Second National Internet of Things Application Innovation Competition (2015)**: *Intelligent Remote Control System Based on Raspberry Pi*.
- Technology Pioneer Award at the **Cyberspace Security Talent Development and Entrepreneurship Forum (2015)**.
- First Prize at the **National Internet of Things Competition (2015)**.

MISCELLANEOUS

- **LinkedIn:** <https://www.linkedin.com/in/dumingzhe> **WeChat:** spirit_song
- **GitHub:** <https://www.github.com/elfsong> **Language:** Mandarin (native) | English (professional)