

Ming Li

Homepage: mingliiii.github.io | [Github](#) | [Google Scholar](#) | [LinkedIn](#) | Email: minglii@umd.edu

Education

University of Maryland, Ph.D. in Computer Science	Aug 2023 – Present
Texas A&M University, M.S. in Computer Science	Aug 2021 – May 2023
Xi'an Jiaotong University, B.S. Computer Science	Aug 2016 – June 2020

Research Experience

University of Maryland	Aug 2023 – Present
<ul style="list-style-type: none">Supervisor: Prof. Tianyi ZhouFocus: Topics related to LLM/VLM Post-training and Interpretability.More specifically, (i) Data Selection: <i>Cherry LLM (IFD)</i>, <i>Superfiltering</i>; (ii) Data Synthesis: <i>Mosaic-IT</i>, <i>Reflection-Tuning</i>, <i>Selective Reflection-Tuning</i>; (iii) Controllability: <i>DEBATunE</i>, <i>RuleR</i>; (iv) Interpretability: <i>Layer_Gradient</i>, <i>Gradient_Unified</i>; (v) Reasoning: <i>MiP-Overthinking</i>, <i>FoReaL-Decoding</i>. (vi) Evaluation: <i>CaughtCheating</i>, <i>ColorBench</i>, <i>TRIG</i>; (vii) Agent: <i>ATLaS</i>.	
Texas A&M University	Aug 2021 – May 2023
<ul style="list-style-type: none">Supervisor: Prof. Ruihong HuangFocus: Natural Language Processing, Natural Discourse Parsing.	
Shenzhen Institute of Advanced Technology, Chinese Academy of Science	Jun 2019 – Jun 2021
<ul style="list-style-type: none">Supervisor: Prof. Yu QiaoFocus: Computer Vision, Scene Text Recognition and Text Detection.	

Internship Experience

Amazon - Applied Scientist Internship	May 2025 – Present
<ul style="list-style-type: none">Reinforcement learning on multi-turn conversation optimization for LLMs	
Microsoft - Research Internship	Feb 2025 – May 2025
<ul style="list-style-type: none">Interpretability and efficient collaborative decoding systems for LLMs	
Adobe - Research Scientist/Engineer Internship	May 2024 – Nov 2024
<ul style="list-style-type: none">Visual grounding and instruction fine-tuning for Multimodal LLMs	
Ping An Technology - Research Internship	May 2023 – Aug 2023
<ul style="list-style-type: none">Automatic data selection for instruction tuning on LLMs	

Selected Publications

- [1] Chenrui Fan*, **Ming Li***, Lichao Sun, Tianyi Zhou. Missing Premise exacerbates Overthinking: Are Reasoning Models losing Critical Thinking Skill? [COLM'25](#).
- [2] **Ming Li**, Yanhong Li, Tianyi Zhou. What Happened in LLMs Layers when Trained for Fast vs. Slow Thinking: A Gradient Perspective. [ACL'25 Oral](#).
- [3] **Ming Li**, Pei Chen, Chenguang Wang, Hongyu Zhao, Yijun Liang, Yupeng Hou, Fuxiao Liu, Tianyi Zhou. Mosaic-IT: Free Compositional Data Augmentation Improves Instruction Tuning. [ACL'25](#).
- [4] Zhixun Chen*, **Ming Li***, Yuxuan Huang, Yali Du, Meng Fang, Tianyi Zhou. ATLaS: Agent Tuning via Learning Critical Steps. [ACL'25](#).
- [5] **Ming Li**, Han Chen, Chenguang Wang, Dang Nguyen, Dianqi Li, Tianyi Zhou. RuleR: Improving LLM Controllability by Rule-based Data Recycling. [NAACL'25](#).
- [6] Hongyu Zhao, Ming Li, Lichao Sun, Tianyi Zhou. BenTo: Benchmark Task Reduction with In-Context Transferability. [ICLR'25](#).
- [7] **Ming Li**, Yong Zhang, Shwai He, Zhitao Li, Hongyu Zhao, Jianzong Wang, Ning Cheng, Tianyi Zhou. Superfiltering: Weak-to-Strong Data Filtering for Fast Instruction-Tuning. [ACL'24](#).
- [8] **Ming Li**, Lichang Chen, Jiuhai Chen, Shwai He, Jiuxiang Gu, Tianyi Zhou. Selective Reflection-Tuning: Student-Selected Data Recycling for LLM Instruction-Tuning. [ACL'24](#).
- [9] **Ming Li**, Jiuhai Chen, Lichang Chen, Tianyi Zhou. Can LLMs Speak For Diverse People? Tuning LLMs via Debate to Generate Controllable Controversial Statements. [ACL'24](#).
- [10] **Ming Li**, Yong Zhang, Zhitao Li, Jiuhai Chen, Lichang Chen, Ning Cheng, Jianzong Wang, Tianyi Zhou, Jing Xiao. From Quantity to Quality: Boosting LLM Performance with Self-Guided Data Selection for Instruction Tuning. [NAACL'24](#).
- [11] **Ming Li**, Bin Fu, Han Chen, Junjun He, Yu Qiao. Dual relation network for scene text recognition. [IEEE Transactions on Multimedia](#).
- [12] **Ming Li**, Bin Fu, Zhengfu Zhang, Yu Qiao. Character-aware sampling and rectification for scene text recognition. [IEEE Transactions on Multimedia](#).

Additional Information

- **Open Source Projects**: *Cherry_LLM* (380 stars), *Reflection_Tuning* (360 stars), *Superfiltering* (160 stars)
- **Invited Talks & Tutorials**: “Data Synthesis for Data Mining”, ACM CIKM 2025 Tutorial (accepted, Oct 2025)
- **Awards**: UMD Jacob K. Goldhaber Award, UMD Graduate School Dean’s Fellowship
- **Area Chair**: EMNLP’25 **Reviewer**: ACL, EMNLP, NAACL, ICLR, ICML, NeurIPS