YINYA HUANG

Phone: (86)15521136607 ⋄ Mail: yinya.huang@hotmail.com ⋄ Google Scholar ⋄ Homepage

RESEARCH FOCUS

My research in general lies in evaluating [11, 10, 12, 6, 4] and improving [9, 8, 7, 5, 3, 2, 1] models' system 2 thinking, especially large language models' complex reasoning abilities (e.g., mathematical reasoning [11, 9, 12, 6], commonsense reasoning [2, 1], logical reasoning [5, 3, 4], and counterfactual thinking [10]) and to further facilitate LLM application for industry (e.g., industrial optimization problems, automated research assistant).

ACADEMIC POSITION

Postdoctoral Fellow *Sep. 2023 - Present*
Department of Computer Science, City University of Hong Kong

EDUCATION

Ph.D. in Computer Science *Sep. 2018 - Jun. 2023*
HCP Lab, Sun Yat-sen University
Advisor: Prof. Dr. Xiaodan Liang

Master in Logic and Computational Linguistics *Sep. 2015 - Jun. 2018*
Institute of Logic and Cognition, Sun Yat-sen University
Head: Prof. Dr. Shier Ju

Bachelor in Logic and Computational Linguistics *Sep. 2011 - Jun. 2015*
Institute of Logic and Cognition, Sun Yat-sen University

PUBLICATIONS

(*: Equal Contribution)

[12] Xiaohan Lin, Qingxing Cao, **Yinya Huang**, Zhicheng Yang, Zhengying Liu, Zhenguo Li and Xiaodan Liang, "ATG: Benchmarking Automated Theorem Generation for Generative Language Models", 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics **(NAACL 2024 Findings)**.

[11] **Yinya Huang**, Xiaohan Lin, Zhengying Liu, Qingxing Cao, Huajian Xin, Haiming Wang, Zhenguo Li, Linqi Song and Xiaodan Liang, "MUSTARD: Mastering Uniform Synthesis of Theorem and Proof Data", in the Twelfth International Conference on Learning Representations **(ICLR 2024)**.

[9] Haiming Wang*, Huajian Xin*, Chuanyang Zheng, Lin Li, Zhengying Liu, Qingxing Cao, **Yinya Huang**, Jing Xiong, Han Shi, Enze Xie, Jian Yin, Zhenguo Li and Xiaodan Liang, "LEGO-Prover: Neural Theorem Proving with Growing Libraries", in the Twelfth International Conference on Learning Representations (**ICLR 2024**).

[6] Jing Xiong, Jianhao Shen, Ye Yuan, Haiming Wang, Yichun Yin, Zhengying Liu, Lin Li, Zhijiang Guo, Qingxing Cao, **Yinya Huang**, Chuanyang Zheng, Xiaodan Liang, Ming Zhang and Qun Liu, "TRIGO: Benchmarking Formal Mathematical Proof Reduction for Generative Language Models". Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2023**).

[5] **Yinya Huang**, Lemao Liu, Kun Xu, Meng Fang, Liang Lin and Xiaodan Liang, "Discourse-Aware Graph Networks for Textual Logical Reasoning". IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI 2023**).

[4] **Yinya Huang**, Hongming Zhang, Ruixin Hong, Xiaodan Liang, Changshui Zhang and Dong Yu, "MetaLogic: Logical Reasoning Explanations with Fine-Grained Structure". Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2022**).

[3] **Yinya Huang**, Meng Fang, Yu Cao, Liwei Wang and Xiaodan Liang, "DAGN: Discourse-Aware Graph Network for Logical Reasoning". Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL-HLT 2021**).

[2] **Yinya Huang**, Meng Fang, Xunlin Zhan, Qingxing Cao, Xiaodan Liang and Liang Lin, "REM-Net: Recursive Erasure Memory Network for Commonsense Evidence Refinement". Thirty-Fifth AAAI Conference on Artificial Intelligence (**AAAI 2021**).

[1] Xunlin Zhan*, **Yinya Huang***, Xiao Dong, Qingxing Cao and Xiaodan Liang, "PathReasoner: Explainable Reasoning Paths for Commonsense Question Answering". Knowledge-Based System (**KBS 2021**).

PREPRINTS

(*: Equal Contribution)

[13] Sichun Luo, Yuxuan Yao, Bowei He, **Yinya Huang**, Aojun Zhou, Xinyi Zhang, Yuanzhang Xiao, Mingjie Zhan and Linqi Song, "Integrating Large Language Models into Recommendation via Mutual Augmentation and Adaptive Aggregation", in arXiv preprint arXiv:2401.13870.

[10] **Yinya Huang***, Ruixin Hong*, Hongming Zhang, Wei Shao, Zhicheng Yang, Dong Yu, Changshui Zhang, Xiaodan Liang and Linqi Song, "CLOMO: Counterfactual Logical Modification with Large Language Models", in arXiv preprint arXiv:2311.17438.

[8] Zhicheng Yang, Yiwei Wang, **Yinya Huang**, Jing Xiong, Xiaodan Liang and Jing Tang "Speak Like a Native: Prompting Large Language Models in a Native Style", in arXiv preprint arXiv:2311.13538.

[7] Sichun Luo, Bowei He, Haohan Zhao, **Yinya Huang**, Aojun Zhou, Zongpeng Li, Yuanzhang Xiao, Mingjie Zhan and Linqi Song, "RecRanker: Instruction Tuning Large Language Model as Ranker for Top-k Recommendation", in arXiv preprint arXiv:2312.16018.

EXPERIENCES

Research Intern in Tencent AI Lab, Seattle (remote) *Dec. 2021 - Aug. 2022*
Advisors: Dr. Hongming Zhang, Dr. Lemao Liu, Dr. Dong Yu
Topic: Natural language explanation. [4, 10]

- Introducing an evaluation method with a hierarchical and multi-label explanation structure. The evaluation of SoTA language models uncovers typical reasoning errors and reveals that certainty modeling should be improved. (Paper accepted by EMNLP 2022 as the first author.)
- Investigating large language models' counterfactual reasoning ability by introducing a generation task in which the language models modify texts under specified logical constraints. Proposing a decomposed self-evaluation metric where the large language models can evaluate the generation task by performing several simple discrimination tasks. The proposed benchmark is challenging to current models.

Research Intern in Tencent AI Lab, Seattle (remote) *May. 2020 - Dec. 2021*
Advisors: Prof. Liwei Wang, Prof. Meng Fang, Dr. Kun Xu, Dr. Dong Yu
Topic: Natural language reasoning. [3, 5]

- Uncovering logical structure from plain texts for effective machine logical reasoning. Constructing logic graphs with discourse relations and elementary discourse units and solving logical reasoning QA (e.g., ReClor and LogiQA). The model's performance ranks 1st on the ReClor leaderboard by Nov. 2020. (Paper accepted by NAACL 2021 as the first author.)
- Extending the logic graph learning to an adaptable structure with a meta-path-based edge propagation and a learnable edge-reasoning mechanism. The proposed method results in effective generalization to unseen logical texts. It also greatly relieves the over-smoothing problem in logic graph learning. (Paper accepted by TPAMI in 2023 as the first author.)

Teaching Assistant in Sun Yat-sen University *Sep. 2019 - Jan. 2020*
Instructor: Prof. Xiaodan Liang. Course Title: Introduction to Deep Learning.

Teaching Assistant in Sun Yat-sen University *Mar. 2018 - Jul. 2018*
Instructor: Prof. Liang Lin. Course Title: Deep Learning in Practice.

PROFESSIONAL SERVICE

Workshops:

- Lead organizer for AI for MATH Workshop and Challenge at ICML 2024.

Program Committee Member: ICLR (2024), NeurIPS (2023), ACL (2024, 2023), EMNLP (2023, 2022), NAACL (2024), COLING (2020), ACM MM (2024), IJCAI (2024)

Journal Reviewer:

- IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**)
- Cognitive Computation

SELECTED AWARDS

- **Outstanding Doctoral Dissertation Award**, Shenzhen Association for Artificial Intelligence, 2024 (*3 awardees per year*)
- **Honors Graduate**, Sun Yat-sen University, 2023
- **Excellent Student Award** of Rhino-Bird Elite Talent Development Program, Tencent, 2021
- **1st Class Scholarship**, Sun Yat-sen University, 2017
- **Outstanding Undergraduate Thesis**, Sun Yat-sen University, 2015

OPEN SOURCE PROJECTS

MUSTARD
https://github.com/Eleanor-H/MUSTARD
Implementation code of the MUSTARD framework and the MUSTARDSAUCE dataset. The MUSTARD framework synthesizes unlimited automated theorem-proving (ATP) data and math word problem (MWP) data when given a mathematical concept seed. The generated data includes informal statement (a theorem/math question in natural language), informal proof (the solution in natural language), formal statement (the theorem/math question in Lean 3), and formal proof (the solution in Lean 3). The MUSTARDSAUCE dataset has a total of 28,316 data points, with a high-quality valid subset with 5,866 data points.

LEGO-Prover
https://github.com/wiio12/LEGO-Prover
Code for LEGO-Prover, a framework that employs a growing skill library containing verified lemmas as skills to augment the capability of LLMs used in theorem proving. By constructing the proof modularly, LEGO-Prover enables LLMs to utilize existing skills retrieved from the library and to create new skills during the proving process. These skills are further evolved (by prompting an LLM) to enrich the library on another scale. Modular and reusable skills are constantly added to the library to enable tackling increasingly intricate mathematical problems. Moreover, the learned library further bridges the gap between human proofs and formal proofs by making it easier to impute missing steps.

TRIGO
https://github.com/menik1126/TRIGO
An ATP benchmark not only requires a model to reduce a trigonometric expression with step-by-step proofs but also evaluates a generative LM's reasoning ability on formulas and its capability to manipulate, group, and factor number terms. Furthermore, we develop an automatic generator based on Lean-Gym to create dataset splits of varying difficulties and distributions to thoroughly analyze the model's generalization ability.

MetaLogic
https://github.com/tencent-ailab/MetaLogic
a comprehensive benchmark to investigate models' logical reasoning capabilities in complex real-life scenarios. This comprehensive logical reasoning explanation form is based on the multi-hop chain of reasoning, the explanation form includes three main components: (1) The condition of rebuttal that the reasoning node can be challenged; (2) Logical formulae that uncover the internal texture of reasoning nodes; (3) Reasoning strength indicated by degrees of certainty.

DAGN
https://github.com/Eleanor-H/DAGN
Code for discourse-aware graph network (DAGN) that relies on the discourse structure of the texts. The model encodes discourse information as a graph with elementary discourse units (EDUs) and discourse relations and learns the discourse-aware features via a graph network for downstream QA tasks.

REM-Net
https://github.com/Eleanor-H/REM-Net
Code for REM-Net, a model equipped with a module to refine the evidence by recursively erasing the low-quality evidence that does not explain the question-answering. Besides, instead of retrieving evidence from existing knowledge bases, REM-Net leverages a pre-trained generative model to generate candidate evidence customized for the question.