Hao ZHANG

Computer Science · Principal Engineer · Ph.D · SG Permanent Resident

Noah's Ark Lab, Singapore Research Center, Huawei International. #13-01, The Metropolis Tower 1, 9 N Buona Vista Dr, 138588, Singapore

hzhang26@outlook.com | 26hzhang.github.io | 26hzhang | hzhang26 | @hzhang26

Summary

ZHANG Hao is a principal engineer with Noah's Ark Lab, Singapore Research Center (2012 Laboratories), Huawei International Pte. Ltd. Before, he was a senior research engineer with Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR). He obtained his Ph.D in Computer Science at the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU), advised by Prof. Aixin Sun. His research interest:

- **Natural Language Processing**: Retrieval-Augmented LLM, Language Model Training and Instruction Tuning, Prompt Engineering, Parameter-efficient Fine-tuning
- **Vision-Language Learning**: Natural Language Video Localization, Spatio-Temporal Sentence Grounding in Videos, Video Corpus Moment Retrieval, Vision-Language Reasoning
- **Recommendation**: Interactive & Conversational Recommendation, LLM for Recommendation, Recommendation as NLP

Education

Nanyang Technological University (NTU) *50 Nanyang Avenue, SG 639798*

Ph.D. in Computer Science *Aug 2019 - Jul 2022*

School of Computer Science and Engineering (SCSE), supervisor: Prof. **Aixin SUN**

M.Sc. in Communications Engineering *Aug 2015 - Jul 2016*

School of Electrical and Electronic Engineering (EEE)

Dalian University of Technology (DUT) *Dalian, P.R. China, 116024*

B.Eng. in Communications Engineering *Sep 2011 - Jul 2015*

School of Information and Communication Engineering (SICE)

Experience

Singapore Research Center (2012 Laboratories), Huawei International *9 N Buona Vista Dr, SG 138588*

Principal Engineer @ Noah's Ark Lab *Jul 2022 - Present*

- **Retrieval-Augmented LLM**: Jointly responsible for the design of RAG framework based on Huawei's PanGu LLM, mainly focusing on the pre-search control (search planning), post-search process (content critique) and response generation (internal & external knowledge fusion, rejection sampling, reasoning, etc.). This model has been deployed and used in various products of Huawei and third-party companies for the knowledge QA of open and vertical domains. Some advanced features, such as "re-search", "preview refinement", are developed to improve the performance and user experience. This model has been used in Xiaoyi Suggestion (Mate 60, Mate X5, etc.), Petal Search, W3 Search, as well as various Huawei internal business/products and products of third-party companies.
- **LLM4Rec**: (1) Reformulating recommendation as a NLP task under the conversational recommendation scenario and develop a model that directly using LLM as retriever, re-ranker and response generator; (2) LLM-enhanced recommendation, responsible for the cross-modal knowledge alignment between textual modality and tabular modality.
- **LLM Agent**: Explore and develop LLM agent based on PanGu LLM (from scratch) and open-sourced LLMs, like ChatGLM, LLaMA, Mistral, etc. The LLM agent is designed to execute versatile tasks by calling APIs and tools, especially focusing on searching and recommendation scenarios. The overall framework is developed and is continuously being optimized to incorporate more tools and improve performance.
- **LLM exploration**: (1) Exploring the retrieval timing (kind of knowledge boundary) for Retrieval-augmented LLM via carefully designed honesty and confidence probes using representation engineering techniques; (2) Exploring the impacts of external distractors to parametric knowledge of LLM: prompting LLM's internal knowledge, noise injection, LLM evaluation, and conflicts alleviation; (3) Exploring the fusion of external knowledge and internal knowledge of LLM, especially the external knowledge contains useless or error information, via insturction tuning and prompt engineering.
- **Miscellaneous**: Involved in different projects and product lines to contribute ideas and codes, for instance, LLM continue training, instruction tuning, parameter-efficient fine-tuning, NL2SQL, query generation, data distillation, etc.

Agency for Science, Technology and Research (A*STAR) *1 Fusionopolis Way, SG 138632*

Senior Research Engineer & Principal Investigator @ Centre for Frontier AI Research (CFAR) *Jan 2022 - May 2022*

- **Towards Efficient Spatio-Temporal Video Grounding via Self-supervision**. Explore to develop an efficient and effective spatio-temporal video grounding framework via self-supervised learning, which could be used to facilitate large-scale video surveillance, vehicle navigation, video stream editing, etc.

Research Engineer @ Computing & Intelligence (C&I), Institute of High Performance Computing (IHPC) *Aug 2020 - Dec 2021*

- **Safe & Robust AI, WP1: Poisoning & Backdoor in Model Training**. Deep learning requires large mounts of data for training. The absence of human supervision over the data collection process exposes organizations to security vulnerabilities. This project aims to explore data poisoning and backdoor attacks, and develop effective approaches to defence against poisoning & backdoor attacks.
- **Human-Robot Collaborate AI for Advanced Manufacturing Engineering (AME), WP3: Human-like Concept and Task Learning**.

Research Engineer @ Artificial Intelligence Initiative (A*AI), Institute of High Performance Computing (IHPC) *Jun 2018 - Jul 2020*

- **Human-Robot Collaborate AI for Advanced Manufacturing Engineering (AME), WP3: Human-like Concept and Task Learning**. Exploit planning, reinforcement learning and implicit learning techniques for concept grounding and human-interactive task learning to achieve efficient and easy-to-use Automatic Robotic Programming and Human-Robot Collaboration.

Nanyang Technological University (NTU) *50 Nanyang Avenue, SG 639798*

Research Associate @ Temasek Laboratories (TL) *Jul 2016 - May 2018*

- ***Collaborated with Social & Cognitive Computing (SCC) Department, IHPC, A*STAR***
- **PrimeNet: Human-inspired Framework for Commonsense Knowledge Representation and Reasoning**. Set out a framework for a commonsense knowledge base that allows for efficient processing, in order to meet the demands of commonsense reasoning and, hence, support intelligent machine performance in real-world tasks.
- **MARACANA: Behavioural Understanding and Narrative Descriptions from Videos**. Develop technology components that can analyze and narrate real world events captured in video, and provide user with a rapid understanding of the happenings in an environment.

Selected Publications

More publications at **Google Scholar** (* indicates equal contributions)

◇ Mathieu Ravaut, **Hao Zhang**, Lu Xu, Aixin Sun, Yong Liu. "*Parameter-Efficient Conversational Recommender System as a Language Processing Task*". The 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2024.

◇ Jing Wang, Aixin Sun, **Hao Zhang**, Xiaoli Li. "*MS-DETR: Natural Language Video Localization with Sampling Moment-Moment Interaction*". The 61th Annual Meeting of the Association for Computational Linguistics (ACL), 2023.

◇ **Hao Zhang**, Aixin Sun, Wei Jing, Joey Tianyi Zhou. "*Temporal Sentence Grounding in Videos: A Survey and Future Directions*". The IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2023.

◇ Sicheng Yu, Jing Jiang, **Hao Zhang**, Yulei Niu, Qianru Sun, Lidong Bing. "*Interventional Training for Out-Of-Distribution Natural Language Understanding*". The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.

◇ Sicheng Yu, Qianru Sun, **Hao Zhang**, Jing Jiang. "*Translate-Train Embracing Translationese Artifacts*". The 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022.

◇ En Yen Puang, **Hao Zhang**, Hongyuan Zhu, Wei Jing. "*Hierarchical Point Cloud Encoding and Decoding with Lightweight Self-Attention based Model*". The IEEE Robotics and Automation Letters (RA-L), 2022.

◇ Guoshun Nan, Rui Qiao, Yao Xiao, Jun Liu, Sicong Leng, **Hao Zhang** and Wei Lu. "*Interventional Video Grounding with Dual Contrastive Learning*". The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

◇ **Hao Zhang**, Aixin Sun, Wei Jing, Liangli Zhen, Joey Tianyi Zhou and Rick Siow Mong Goh. "*Parallel Attention Network with Sequence Matching for Video Grounding*". The Findings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021.

◇ Sicheng Yu, **Hao Zhang**, Yulei Niu, Qianru Sun and Jing Jiang. "*COSY: COunterfactual SYntax for Cross-Lingual Understanding*". The 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021.

◇ **Hao Zhang**, Aixin Sun, Wei Jing, Guoshun Nan, Liangli Zhen, Joey Tianyi Zhou and Rick Siow Mong Goh. "*Video Corpus Moment Retrieval with Contrastive Learning*". The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2021.

◇ **Hao Zhang**, Aixin Sun, Wei Jing, Liangli Zhen, Joey Tianyi Zhou and Rick Siow Mong Goh. "*Natural Language Video Localization: A Revisit in Span-based Question Answering Framework*". The IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.

◇ Fuzhao Xue, Aixin Sun, **Hao Zhang** and Eng Siong Chng. "*GDPNet: Refining Latent Multi-View Graph for Relation Extraction*". The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), 2021.

◇ **Hao Zhang**, Aixin Sun, Wei Jing and Joey Tianyi Zhou. "*Span-based Localizing Network for Natural Language Video Localization*". The 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020.

◇ Ming Yan, **Hao Zhang**, Di Jin and Joey Tianyi Zhou. "*Multi-source Meta Transfer for Low Resource Multiple Choice Question Answering*". The 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020.

◇ Tianying Wang*, Wei Qi Toh*, **Hao Zhang***, Xiuchao Sui, Shaohua Li, Yong Liu and Wei Jing. "*RoboCoDraw: Robotic Avatar Drawing with GAN-based Style Transfer and Time-efficient Path Optimization*". The Thirty-Forth AAAI Conference on Artificial Intelligence (AAAI), 2020.

◇ Joey Tianyi Zhou*, **Hao Zhang***, Di Jing and Xi Peng. "*Dual Adversarial Transfer for Sequence Labeling*". The IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.

◇ Joey Tianyi Zhou*, **Hao Zhang***, Di Jing, Hongyuan Zhu, Rick Siow Mong Goh and Kenneth Kwok. "*Dual Adversarial Neural Transfer for Low-resource Named Entity Recognition*". The 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019.

◇ Tianying Wang, **Hao Zhang**, Wei Qi Toh, Hongyuan Zhu, Cheston Tan, Yan Wu, Yong Liu and Wei Jing. "*Efficient Robotic Task Generalization Using Deep Model Fusion Reinforcement Learning*". The IEEE International Conference on Robotics and Biomimetics (ROBIO), 2019.

◇ Joey Tianyi Zhou*, **Hao Zhang***, Meng Fang, Di Jing, Xi Peng, Yang Xiao and Zhiguo Cao. "*RoSeq: Robust Sequence Labeling*". IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2019.

◇ Joey Tianyi Zhou, Meng Fang, **Hao Zhang**, Chen Gong, Xi Peng, Zhiguo Cao and Rick Siow Mong Goh. "*Learning With Annotation of Various Degrees*". IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2019.

Theses

Towards Temporal Sentence Grounding in Videos

Supervisor: Prof. **Aixin SUN**

Doctor of Philosophy (Ph.D.), School of Computer Science and Engineering, Nanyang Technological University

Activities & Services

2023- **Reviewer**, ACL Rolling Review
2023-2024 **PC Member & Reviewer**, The International Joint Conferences on Artificial Intelligence (IJCAI)
2024 **PC Member & Reviewer**, The International World Wide Web Conferences (WWW)
2024 **PC Member**, The ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
2019-2023 **PC Member & Reviewer**, The Annual Meeting of the Association for Computational Linguistics (ACL)
2023 **Reviewer**, The Conference on Empirical Methods in Natural Language Processing (EMNLP)
2023 **Co-organizer**, Interactive Recommendation System Workshop, 16th ACM International WSDM Conference
2023 **Reviewer**, The Association for Computing Machinery (ACM)'s Annual Conference on Multimedia (ACM MM)
2022-2023 **PC Member & Reviewer**, The International Conference on Computational Linguistics (COLING)
2021-2022 **PC Member & Reviewer**, The ACM SIGIR Conference on Research and Development in Information Retrieval
2021 **PC Member**, The AAAI Conference on Artificial Intelligence (AAAI)

Honors & Awards

2020 **Best Paper Award**, The 13th EAI International Conference on Mobile Multimedia Communications *Singapore*
2018 **1st Runner Up**, Artificial Intelligence (AI) Hackathon in Agency for Science, Technology and Research *Singapore*
2015 **Outstanding Graduates**, Dalian University of Technology *P. R. China*
2013 **Second Prize**, Higher Education Press Cup College Student Mathematical Modeling Contest (National Level) *P. R. China*
2013 **First Prize**, Higher Education Press Cup College Student Mathematical Modeling Contest (Provincial Level) *P. R. China*
2013 **Certificated**, Undergraduate Innovation and Entrepreneurship Training Program (National Level) *P. R. China*