Qingyi Si (侣庆一)

(+86)131-6168-5288 | 1997.12.01 | siqingyi@iie.ac.cn | https://phoebussi.github.io/ | https://github.com/PhoebusSi | Beijing, China | My research interests include OOD generalization, VQA, vision & language, and LLMs.

Education

University of Chinese Academy of Sciences (UCAS), Institute of Information Engineering — *Beijing*

PhD in Cyberspace Security and **MS** in Computer Applied Technology, advised by Prof. Zheng Lin and Weiping Wang. — Sep. 2019 - Jun. 2024

Beijing Language and Culture University (BLCU) — *Beijing*

BS in Computer Science and Technology — Sep. 2015 - Jun. 2019

Publications (Co-*) First author of 10 papers

[1] **Qingyi Si**, Yuchen Mo, Zheng Lin, et al. "Combo of Thinking and Observing for Outside-Knowledge VQA". ***ACL'23***

[2] **Qingyi Si**, Fandong Meng, et al. "Language Prior Is Not the Only Shortcut: A Benchmark for Shortcut Learning in VQA". ***EMNLP'22***
*ps: This paper received **high praise** from **Damien Teney**.*

[3] **Qingyi Si**, Fandong Meng, et al. "Towards Robust VQA: Making the Most of Biased Samples via Contrastive Learning". ***EMNLP'22***

[4] **Qingyi Si**, Zheng Lin, et al. "Check It Again: Progressive Visual Question Answering via Visual Entailment". ***ACL'21***

[5] **Qingyi Si**, Yuanxin Liu, et al. "Learning Class-Transductive Intent Representations for Zero-shot Intent Detection". ***IJCAI'21***

[6] Zhe Wen*, **Qingyi Si***, Zheng Lin, et al. "Schema Item Matters in Knowledge Base Question Answering". *IJCNN'23*

Under Review Papers

[7] **Qingyi Si**, Yuanxin Liu et al. "Compressing And Debiasing Vision-Language Pre-Trained Models for VQA". ***EMNLP'23***

[8] **Qingyi Si**, Tong Wang et al. "An Empirical Study of Instruction-tuning Large Language Models in Chinese". ***EMNLP'23***

[9] Huishan Ji*, **Qingyi Si***, Zheng Lin, et al. "Towards Many-to-one Visual Question Answering". ***NeurIPS'23***

[10] Jie Xu*, **Qingyi Si***, Hanbo Zhang*, et al. "TiO: A Unified Interactive Visual-Language Disambiguation Transformer". ***NeurIPS'23***

Co-author Papers

[11] Duo Zheng, Fandong Meng, **Qingyi Si**, et al. "Visual Dialog for Spotting the Differences between Pairs of Similar Images". ***MM'22***

[12] Ran Li, **Qingyi Si**, et al. "A Multi-channel Neural Network for Imbalanced Emotion Recognition". *ICTAI'19*

[13] Chunhua Liu, Yan Zhao, **Qingyi Si**, et al. "Multi-perspective fusion network for commonsense reading comprehension".*CCL'18*

Research Projects

Alpaca-CoT (***https://github.com/PhoebusSi/alpaca-CoT/***) — ***1.8k Star***

An Instruction-tuning Platform with Unified Interface of Instruction Collection, Parameter-efficient Methods, and Large Language Models. *(I was invited by Machine Heart (jiqizhixin) to give a talk about this project.)* — ***2.9k Downloads last month***

We unified the interfaces of instruction-tuning data, multiple LLMs and parameter-efficient methods (e.g., lora, p-tuning) together for easy use. Besides, we are the first to extend CoT data into LLaMA to improve its reasoning ability. We constructed the largest collection of instruction tuning datasets currently available. On this basis, we conduct a thorough empirical study of instruction tuning in Chinese.

A Multimodal Large Language Model — ***In Progress***

We empower LLMs with multi-modality and visual grounding capabilities.

Awards and Honors

[1] **National Scholarship for Doctoral students (Top 2%, RMB ¥ 30,000).** *Ministry of Education of P.R. China. 2021.*

[2] **Merit Student, University of Chinese Academy of Sciences** *UCAS. 2020, 2021.*

[3] **Excellent Undergraduate Graduation Project (Thesis).** *Education Commission of Beijing. 2019.*

[4] **Regular Institutions of Higher Education Outstanding Graduate.** *Education Commission of Beijing. 2019.*

[5] **National Scholarship for Undergraduates (Top 2%, RMB ¥ 8,000).** *Ministry of Education of P.R. China. 2017, 2018.*

Working Experience

Microsoft Research Asia (MSRA) — *Beijing*

Research Intern in NLC group — November.2021 - May.2022

WeChat, Tencent — *Beijing*

Research Intern in the Pattern Recognition Center (PRC), WeChat AI — March.2021 - November.2021

Research Review

Language & Vision

OOD robustness of VQA

[1] ***ACL'21***: **Overcoming the OOD problem in VQA.** Specifically, we propose a select-and-rerank (SAR) progressive framework formalizing VQA as Visual Entailment, which can make full use of the interactive information of image, question and candidate answers.

[2] ***EMNLP'22***: **Overcoming the OOD problem while maintaining in-distribution performance.** Specifically, instead of undermining the importance of the biased samples like the previous works, our method makes the most of them based on a contrstive learning method.

[3] ***EMNLP'22***: **Building a more reliable and comprehensive testbed for OOD robustness.** We benchmark a series of state-of-the-art models, and find that all existing debiasing methods fails to generalize to our proposed benchmark.

[4] ***EMNLP'23***: **Searching sparse and robust subnetworks for OOD VQA.** We systematically study the design of a debiasing and pruning pipeline. The obtained sparse and robust subnetworks clearly outperform the SoTAs with fewer parameters.

Open-domain VQA

[1] ***ACL'23***: **Bring in more comprehensive types of knowledge.** Specifically, we mimic human behavior, "thinking while observing", i.e., benefiting from the vast knowledge in natural-language space while making the most of the visual features for better image understanding.

[2] ***NeurIPS'23***: **Towards one-to-many VQA.** Specifically, we collect 12 VQA tasks to empirically explore how to balance the three ability of real open-domain VQA models: knowledge capacity, visual attribute recognition and scene comprehension.

Large Language Models

Instruction-tuning

[1] ***Alpaca-CoT***: **An Instruction-tuning Platform** with Unified Interface of Instruction Collection, Parameter-efficient Methods, and LLMs.

[2] ***EMNLP'23***: **Empirical study on instruction tuning in Chinese.** Specifically, we systematically explore the impact of LLM bases, parameter-efficient methods, instruction data types, which are the three most important elements for instruction-tuning. Besides, we also conduct experiment to study the impact of other factors, e.g., chain-of-thought data and human-value alignment.

Multimodal LLMs (MLLM)

[1] ***NeurIPS'23***: **An MLLM unifying visual dialog and visual grounding.** We propose an interactive visual-language Transformer (MLLM) that can hold a natural and informative dialog with human users to disambiguate their expression and identify the target object, which can also generalize to real-world products, e.g., household robot.

[2] ***In process***: **Empowering LLMs with multimodal understanding, as well as visual and language generation capabilities simultaneously.**

Self-Assessment

- Outgoing, energetic and a nice personality.
- Creative, self-motivated and a passion for problem-solving.
- Strong team work, but also ability to work independently.