

HAONAN LI

Email: nathan.8270.n@gmail.com

Postdoctoral Fellow at MBZUAI (Natural Language Processing)

Cofounder and CTO of LibrAI

EDUCATION BACKGROUND

The University of Melbourne Doctor of Philosophy (Natural Language Processing)	<i>Jul.2018 - Aug.2022</i>
Shanghai Jiao Tong University Bachelor of Science (Computer Science)	<i>Sep.2013 - Jun.2017</i>
Affiliated High School of Shanxi University Senior Middle School Diploma	<i>Sep.2010 - Jul.2013</i>

PUBLICATIONS

Yuxia Wang, **Haonan Li**, Xudong Han, Preslav Nakov, Timothy Baldwin, **Do-Not-Answer: A Dataset for Evaluating Safeguards in LLMs**, *under review*

Haonan Li, Yixuan Zhang, Fajri Koto, Yifei Yang, Hai Zhao, Yeyun Gong, Nan Duan, Timothy Baldwin, **CMMLU: Measuring massive multitask language understanding in Chinese**, *under review*

Haonan Li, Fajri Koto, Minghao Wu, Alham Fikri Aji, Timothy Baldwin, **Bactrian-X: A Multilingual Replicable Instruction-Following Model with Low-Rank Adaptation**, *under review*

Neha Sengupta, Sunil Kumar Sahu, Bokang Jia, Satheesh Katipomu, **Haonan Li**, Fajri Koto, Osama Mohammed Afzal, Samta Kamboj, Onkar Pandit, Rahul Pal, Lalit Pradhan, Zain Muhammad Mujahid, Massa Baali, Alham Fikri Aji, Zhengzhong Liu, Andy Hock, Andrew Feldman, Jonathan Lee, Andrew Jackson, Preslav Nakov, Timothy Baldwin, Eric Xing, **Jais and Jais-chat: Arabic-Centric Foundation and Instruction-Tuned Open Generative Large Language Models**, *technique report*

Fajri Koto, Nurul Aisyah, **Haonan Li**, Timothy Baldwin, **Large Language Models Only Pass Primary School Exams in Indonesia: A Comprehensive Test on IndoMMLU**, *EMNLP-23*

Yixuan Zhang, **Haonan Li**, **Can Large Language Model Comprehend Ancient Chinese? A Preliminary Test on ACLUE**, *ALP-23 Workshop*

Haonan Li, Maria Vasardani, Martin Tomko, Timothy Baldwin, **Lamb: Location-Aware Modular Bi-encoder for Tourism Question Answering**, *AAACL-23*.

Shuai Fan, Chen Lin, **Haonan Li**, Zhenghao Lin, Hang Zhang, Yeyun Gong, Jian Guo, Nan Duan, **Sentiment-Aware Word and Sentence Level Pre-training for Sentiment Analysis**, *EMNLP-22*

Haonan Li, Yameng Huang, Yeyun Gong, Jian Jiao, Ruofei Zhang, Timothy Baldwin, Nan Duan, **CULG: Commercial Universal Language Generation**, *NAACL-22 Industry Track*

Haonan Li, Maria Vasardani, Martin Tomko, Timothy Baldwin, **MultiSpanQA: A Dataset for Multi-Span Question Answering**, *NAACL-22*.

Werner Kuhn, Ehsan Hamzei, Martin Tomko, Stephan Winter, **Haonan Li**, **The Semantics of Place-related Questions**, *JOSIS-21*

Haonan Li, Ehsan Hamzei, Ivan Majic, Hua Hua, Jochen Renz, Martin Tomko, Maria Vasardani, Stephen Winter, Timothy Baldwin, **Neural Factoid Geospatial Question Answering**, *JOSIS-21*

Haonan Li, Yeyun Gong, Jian Jiao, Ruofei Zhang, Timothy Baldwin, Nan Duan, **KFCNet: Knowledge Filtering and Contrastive Learning Network for Generative Commonsense Reasoning**, *EMNLP-21*

Haonan Li, Maria Vasardani, Martin Tomko, Timothy Baldwin, **Target Word Masking for Location Metonymy Resolution**, *COLING-20*

Ehsan Hamzei, **Haonan Li**, Maria Vasardani, Timothy Baldwin, Stephan Winter, Martin Tomko, **Place Questions and Human-Generated Answers: A Data Analysis Approach**, *AGILE-19*

Haonan Li, Minghan Wang, Timothy Baldwin, Martin Tomko, Maria Vasardani, **UniMelb at SemEval-2019 Task 12: Multi-model combination for toponym resolution**, *SemEval-19*

Yuting Han, Zheng Zhou, **Haonan Li**, Guoyin Wang, Wei Deng, Zhixing Li, **Classifying Relation via Piecewise Convolutional Neural Networks with Transfer Learning**, *ICMMI-19*

Haonan Li, Zhisong Zhang, Yuqi Ju, Hai Zhao, **Neural Character-level Dependency Parsing for Chinese**, *AAAI-18*

EXPERIENCE

- Postdoctoral Fellow, MBZUAI, UAE *Feb.2023 - now*
 - Work on LLM development and evaluation. Responsible for Jais (the worlds highest quality Arabic LLMs) evaluation.
- Research Intern, Microsoft Research Asia, China *Jan.2021 - June.2021*
 - Focus on a constrained text generation challenge for generative commonsense reasoning (CommonGen), achieve top-1 on the public leaderboard, proposed method also improves keyword generation on a real-world ads dataset (crawled from Microsoft), paper published at EMNLP-21.
 - Working on language model pre-training for commercial language generation, conduct extensive experiments on real-world ads generation dataset, paper published at NAACL-22, Industry Track.
- Academic Tutor, The University of Melbourne, Australia *Mar.2020 - Jul.2020*
 - Tutor of two subjects: *Natural Language Processing* and *Introduction to Programming*.
- Algorithm Intern, NLP Team, Horizon Robotics, China *Oct.2017 - Jan.2018*
 - Participated (as a core developer) in the development of a rule-based semantic parser.
- Intern, Data Center Group, Intel Corporation, China *Dec.2015 - Apr.2016*
 - Charge for managing and updating the code, as well as help developed on a remote debug tool.

AWARDS AND ACHIEVEMENTS

Stars of Tomorrow Internship Program Award, Microsoft Research Asia, 2021

First place on CommonGen public leaderboard, A Constrained Text Generation Challenge for Generative Commonsense Reasoning, June 2021 – May 2022

Second Place in Competition of SemEval-2019 Task 12, Toponym Resolution in Scientific Papers, 2019

Melbourne Research Scholarship, The University of Melbourne, 2018

Second Class of Awarding in the National Mathematics League Match for Undergraduate, 2015

Second Class of Awarding in the National Mathematics League Match for Undergraduate, 2014

Excellent Awarding in the Hackathon organized by Microsoft Corporation, China, 2015

Honorable Mentions in Mathematical Contest in Modeling (MCM), 2013

First Class of Awarding in the National Mathematics League Match for Senior Middle School Student, China, 2011 and 2012

Second Class of Awarding in the National Mathematics League Match for Senior Middle School Student, China, 2011