Meng Cao

📞+1-438-722-6288   ✉ meng.cao@mail.mcgill.ca   ⌨ mcao516   🔗 mcao516.github.io

EDUCATION

McGill University / Mila - Québec AI Institute   Sept. 2019 – Apr. 2025
Ph.D. in Computer Science   GPA: 4.0/4.0
Advisor: Jackie Chi Kit Cheung

Northeastern University (China)   Sept. 2015 – Jun. 2019
Major in Software Engineering   GPA: 3.96/4.0
Ranking: 1/59

RESEARCH INTERESTS

Text Generation, Reinforcement Learning, LLMs

PUBLICATIONS

[1] Can LLMs Reason Abstractly Over Math Word Problems Without CoT? Disentangling Abstract Formulation From Arithmetic Computation
Ziling Cheng, **Meng Cao**, Leila Pishdad, Yanshuai Cao, Jackie CK Cheung
ArXiv

[2] SCAR: Shapley Credit Assignment for More Efficient RLHF
Meng Cao, Shuyuan Zhang, Xiao-Wen Chang, Doina Precup
ArXiv

[3] Stochastic Chameleons: Irrelevant Context Hallucinations Reveal Class-Based (Mis)Generalization
Ziling Cheng*, **Meng Cao***, Marc-Antoine Rondeau, Jackie Chi Kit Cheung
ACL 2025

[4] The Reasoning-Memorization Interplay in Language Models Is Mediated by a Single Direction
Yihuai Hong, **Meng Cao**, Dian Zhou, Lei Yu, Zhijing Jin
ACL 2025 Findings

[5] Enhancing Reinforcement Learning with Dense Rewards from Language Model Critic
Meng Cao, Lei Shu, Lei Yu, Yun Zhu, Nevan Wichers, Yinxiao Liu, Lei Meng
EMNLP 2024

[6] Mechanisms of Non-Factual Hallucination in Language Models
Lei Yu*, **Meng Cao***, Jackie CK Cheung, Yue Dong
Findings of EMNLP 2024

[7] Successor Features for Efficient Multi-Subject Controlled Text Generation
Meng Cao, Mehdi Fatemi, Jackie CK Cheung and Samira Shabanian
ICML 2024

[8] Systematic Rectification of Language Models via Dead-end Analysis
Meng Cao, Mehdi Fatemi, Jackie CK Cheung and Samira Shabanian
ICLR 2023

[9] Responsible AI Considerations in Text Summarization Research: A Review of Current Practices
Yu Lu Liu, **Meng Cao**, Su Lin Blodgett, Jackie Chi Kit Cheung, Alexandra Olteanu and Adam Trischler
Findings of EMNLP 2023

[10] Hallucinated but Factual! Inspecting the Factuality of Hallucinations in Abstractive Summarization
Meng Cao, Yue Dong and Jackie C. K. Cheung.
ACL 2022

[11] Learning with Rejection for Abstractive Text Summarization
Meng Cao, Yue Dong, Jingyi He and Jackie C. K. Cheung
EMNLP 2022

[12] Factual Error Correction for Abstractive Summarization Models
Meng Cao, Yue Dong, Jiapeng Wu and Jackie C. K. Cheung
EMNLP 2020

[13] TeMP: Temporal Message Passing for Temporal Knowledge Graph Completion
Jiapeng Wu, **Meng Cao**, Jackie Chi Kit Cheung, William L. Hamilton
EMNLP 2020

[14] Referring Expression Generation Using Entity Profiles
Meng Cao and Jackie Chi Kit Cheung
EMNLP-IJCNLP 2019

INDUSTRY EXPERIENCES

Google Student Researcher Jan. 2024 – May. 2024
Student Researcher Advisor: Lei Meng, Shu Lei
- Explore the self-critique setting, where the same LLM is used as both the generator and the critic
- Human evaluation of RL fine-tuned model & paper writing
- First author paper published at EMNLP 2024

Google Research – Mountain View Aug. 2023 – Nov. 2023
Research Intern Advisor: Lei Meng, Shu Lei
- Investigated the challenge of reward sparsity in RLHF
- Leverage the critique ability of LLMs for generating dense intrinsic rewards
- Demonstrated that combining extrinsic and intrinsic rewards substantially enhances sample efficiency

Microsoft Research – Montréal Apr. 2022 – Mar. 2023
Machine Learning Research Intern Advisor: Samira Shabanian, Mehdi Fatemi
- Work on detoxification of large language models like GPT-3
- Our approach, *rectification*, utilizes a separate but significantly smaller value function for LM detoxification
- Paper accepted by ICLR 2023

Huawei Noah's Ark Lab – Montréal May 2021 – Mar. 2022
Applied Machine Learning Research Intern Advisor: Mehdi Rezagholizadeh
- Knowledge distillation for text generation tasks
- Data augmentation for knowledge distillation using cluster labels

Borealis AI – Toronto May 2020 – Aug. 2020
Research Intern, Alan Team Advisor: Yanshuai Cao
- Build a distributed data-parallel Text-to-SQL system
- Applied meta-learning algorithms (MAML, Reptile, MetaReg) in cross-database semantic parsing task
- Achieved promising results in the zero-shot domain transfer setting

SCHOLARSHIPS

China National Scholarship 2015-2016, 2017-2018
The Chinese Government, among top-0.2% all undergraduates in China

The Bao Gang Education Scholarship Sept. 2018
China BAOWU Steel Group, selected rate: 4/20000 in NEU

The First-Class Scholarship in Northeastern University 2015-2016, 2016-2017, 2017-2018

Northeastern University, among top-3% undergraduates in NEU

AWARDS

"Top Ten Undergraduates" in Northeastern University May 2018

Northeastern University, among top-0.2% undergraduates in NEU

The Excellent Performance Award for IBM Undergraduate Innovation Lab Program Apr. 2017

IBM Research China, Beijing

The First Prize in the 8th National College Students Mathematics Competition Oct. 2016

Chinese Mathematical Society

The First Prize in the 9th National College Students Mathematics Competition Oct. 2017

Chinese Mathematical Society

The Meritorious Winner for the Mathematical Contest in Modeling (MCM/ICM), 2018 Feb. 2018

COMAP, the Consortium for Mathematics and Its Applications (USA)

The Honorable Mention for the Mathematical Contest in Modeling (MCM/ICM), 2017 Feb. 2017

COMAP, the Consortium for Mathematics and Its Applications (USA)

SKILLS

Programming Languages: skilled in: Python (especially NumPy, PyTorch); familiar with: Tensorflow, JavaScript, SQL

Extensive knowledge of: Natural Language Processing, Text Generation, Reinforcement Learning