

Ruoxi Ning

Email: ruoxi.ning@uwaterloo.ca Website: ruoxining.github.io

GitHub: ruoxining Scholar: Ruoxi Ning

EDUCATION

University of Waterloo MMath in Computer Science, Computational Linguistics Advisor: Prof. Freda Shi	Waterloo, Canada 2024–2026
Zhejiang University Bachelor's double majors in English Linguistics and Computer Science Advisor: Prof. Bin Shao	Zhejiang, China 2019–2024

RESEARCH EXPERIENCE

WestlakeNLP, Westlake University Research Assistant; Advisor: Prof. Yue Zhang	Hangzhou, China 2021–2024
StatNLP, Singapore University of Technology and Design Research Assistant; Advisor: Prof. Wei Lu	Singapore 2023
Imperial College London Winter School	London, UK 2022

PUBLICATIONS

- On the Effect of Hyperparameters in Language Modeling for Computational Linguistics**
Ruoxi Ning; Advisor: Freda Shi
In ACL 2026 Main; University of Waterloo Thesis, 2026.
- Logical Reasoning in Large Language Models: A Survey**
Hanmeng Liu, Zhizhang Fu, Mengru Ding, Ruoxi Ning, Chaoli Zhang, Xiaozhang Liu, Yue Zhang
arXiv, 2025.
- From Behavioral Performance to Internal Competence: Interpreting Vision-Language Models with VLM-Lens**
Hala Sheta*, Eric Haoran Huang*, Shuyu Wu*, Iliia Alenabi, Jiajun Hong, Ryker Lin, Ruoxi Ning, Daniel Wei, Jialin Yang, Jiawei Zhou, Ziqiao Ma, Freda Shi
In EMNLP System Demonstrations, 2025. (*: Equal Contribution)
- NovelQA: Benchmarking Question Answering on Documents Exceeding 200K Tokens**
Cunxiang Wang*, Ruoxi Ning*, Boqi Pan, Tonghui Wu, Qipeng Guo, Cheng Deng, Guangsheng Bao, Xiangkun Hu, Zheng Zhang, Qian Wang, Yue Zhang
In ICLR, 2025. (*: Equal Contribution)
- TAIL: A Toolkit for Automatic and Realistic Long-Context Large Language Model Evaluation**
Gefei Gu, Yilun Zhao, Ruoxi Ning, Yanan Zheng, Arman Cohan
In EMNLP System Demonstrations, 2024.
- Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4**
Hanmeng Liu, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou, Yue Zhang
arXiv, 2023.
- A Quantitative Study of Register Differentiation and Change in Contemporary English: Latinate/Germanic Ratio as a Parameter**
Ruoxi Ning, Bin Shao
Zhejiang University Thesis, 2024. Outstanding Thesis.

8. Challenges to Open-Domain Constituency Parsing

Sen Yang, Leyang Cui, Ruoxi Ning, Di Wu, Yue Zhang
In Findings of ACL, 2022.

SELECTED PROJECTS

VLM-Lens Interpretability Toolkit

2024–2025

Contributor

PyTorch, Hugging Face Transformers, SQL

- Developed probes and test cases for vision-language models including Qwen-VL and InternLM.
- Visualized language gaps and modality gaps on color concepts in vision-language models.

NovelQA Benchmark

2023–2024

Lead

Python, LangChain, Azure

- Designed 2,305 questions over novels averaging 200K+ tokens for long-document QA.
- Benchmarked question answering, retrieval, and reasoning on long-context documents.
- Deployed NovelQA as an online evaluation platform while preserving data privacy.

GLoRE Benchmark Pipeline

2023–2024

Contributor

Python, OpenAI Evals

- Built an LLM logical-reasoning benchmark unifying LogiQA, ReClor, FOLIO, AR-LSAT, ProofWriter, and RuleTaker.
- Supported pipeline evaluation with OpenAI Evals and Hugging Face models.

TEACHING AND SERVICE

Teaching Assistant, University of Waterloo
CS 486/686 Introduction to AI.

Fall 2024–Winter 2026

Teaching Assistant, University of Waterloo
CS 245 Logic and Computation.

Spring 2025

Conference Reviewer

2024–2026

ARR 2024–2026; NeurIPS 2024; ICLR 2025–2026.

SELECTED HONORS & AWARDS

University of Waterloo Graduate Scholarship

2024–2026

Cohere Graduate Scholarship in Computer Science

2024

Outstanding Thesis, Zhejiang University

2024

SKILLS

Frameworks & Tools: PyTorch; NumPy; Hugging Face; scikit-learn; Weights & Biases; Docker; Slurm; React; Praat

Systems & Concepts: Agentic Retrieval-Augmented Generation (RAG); ML Systems; Evaluation; Benchmarking

Programming Languages: Python; C++; R; SQL; JavaScript; TypeScript; x86 Assembly

Natural Languages: Mandarin Chinese (Native); English (Professional); French (B1)