Longju Bai

Ann Arbor, MI | longju@umich.edu | (+1)734-881-5520 | github.com/LongjuBai

RESEARCH INTERESTS

I am broadly interested in promoting research in **NLP and Multimodal AI** from the following aspects:

- *Responsible AI*: Improving AI fairness, inclusiveness, and cross-cultural capabilities; enhancing AI for social good applications; reducing bias and harmful outputs.
- *AI Transparency*: Improving AI interpretability and controllability through mechanistic analysis and representation engineering; improving AI alignment, robustness, and reasoning ability.
- *AI Agent*: Designing AI agents for task-solving and human-simulation; enhancing interactive capabilities among models, humans, and external environments.

EDUCATION

University of Michigan, Ann Arbor, M.S. in Electrical and Computer Engineering   Sept 2022 - May 2024
- GPA: 3.79/4.00
- **Related Course:** Large Language Models, Machine Learning Theory, Natural Language Processing, Causal Machine Learning, Computer Vision, Mobile Robotics, Self-Drving Cars

Beijing Forestry University, B.S. in Electrical Engineering and Automation   Sep 2018 - Jun 2022
- GPA: 87.55/100
- **Related Course:** Data Structure, C++ Program Design, Mathematical Modeling, Signal and System, Algorithm Design and Analysis, Java Development and Practice, Microcontroller

PUBLICATIONS

[Manuscript in Preparation] Beyond Training Bounds: Generalization in Multi-Hop Knowledge-Based Reasoning for Language Models   2024

Longju Bai, Tianci Xue, Boshi Wang, Zeyi Liao, Huan Sun

[NAACL Under review] The Power of Many: Multi-Agent Multimodal Models for Cultural Image Captioning   Oct 2024

Longju Bai, Angana Borah, Oana Ignat, Rada Mihalcea

[PDF Version]

Why AI Is WEIRD and Shouldn't Be This Way: Towards AI For Everyone, With Everyone, By Everyone   Oct 2024

Rada Mihalcea, Oana Ignat, *Longju Bai*, Angana Borah, Luis Chiruzzo, Zhijing Jin, Claude Kwizera, Joan Nwatu, Soujanya Poria, Thamar Solorio

[10.48550/arXiv.2410.16315]

Chumor 1.0: A Truly Funny and Challenging Chinese Humor Understanding Dataset from Ruo Zhi Ba   June 2024

Ruiqi He, Yushu He, *Longju Bai*, Jiarui Liu, Zhenjie Sun, Zenghao Tang, He Wang, Hanchen Xia, Naihao Deng

[10.48550/arXiv.2406.12754]

Annotations on a Budget: Leveraging Geo-Data Similarity to Balance Model Performance and Annotation Cost   May 2024

Oana Ignat, *Longju Bai*, Joan C. Nwatu, Rada Mihalcea

[https://aclanthology.org/2024.lrec-main.112]

RESEARCH EXPERIENCE

Monitoring and Controlling LLM Creativity Through Representation Analysis   Oct 2024 – Present
Supervisor: Prof. Rada Mihalcea, University of Michigan

- Track creativity-related model activity during inference by observing internal model activations.
- Enable controlled creative generation by adjusting activations within model layers during inference.
- Analyze the overlap between creativity and related concepts like truthfulness or hallucination

Expanding Vision-Language Model Diversity through Functional Groupings   Oct 2024 – Present
Supervisor: Prof. Rada Mihalcea, University of Michigan

- Conducted data augmentation on low-resource datasets using functions and object names
- Improving Foundational Vision-Language Model's diversity and cross-cultural ability by introducing functional embeddings

Transformer Generalization in Multi-Hop Knowledge-Based Reasoning   June 2024 – Oct 2024
Supervisor: Prof. Huan Sun, Ohio State University

- Analyzed the influence of different factors on the length generalization of multi-hop composition tasks
  - For data format, compared the implicit reasoning and CoT reasoning format;
  - For positional encoding, compared the influence of Absolute Positional Encoding, Relative Positional Encoding, Rotary Positional Encoding, and the Contextual Positional Encoding during pretraining;
  - For model architecture, compared the vanilla transformer and the transformer with shared memory.
- Broke down the CoT process into distinct subtasks (retrieval, copy and induction) for more thorough analysis.

W.E.I.R.D. AI: Towards AI for Everyone, with Everyone, by Everyone   Apr 2024 – Oct 2024
Supervisor: Prof. Rada Mihalcea, University of Michigan

- Contributed in literature survey on the inclusiveness and diversity of current AI in terms of *model performance*
  - The presence of stereotypes for inappropriate contexts (bias)
  - The vulnerability regarding Jailbreaking attacks and toxic contents
  - The poor performance of LLMs with dialects
  - The issue of value alignment of LLMs towards non-well-defined problems

Multimodal Multi-Agent system for Cross-Cultural Image Understanding   Apr 2024 – Oct 2024
Supervisor: Prof. Rada Mihalcea, Prof. Oana Ignat, University of Michigan

- Constructed a multi-modal multi-agent system for culture-aware image description;
- Proposed 'Cultural Words Number' as a new metric to evaluate the culture content in image captioning;
- Performed ablation study to analyze the influence of prompt strategy (Anthropological prompt and CoT prompt), conversation rounds, and model-fine tuning on the caption quality.

Chumor 1.0: A Dataset for Chinese Humor Understanding   Feb 2024 - June 2024
Independent Research, University of Michigan

- constructed Chumor, a truly funny and challenging Chinese humor understanding dataset, addressing the lack of non-English humor understanding dataset
- revealed that on Chumor, human explanations for jokes are significantly better than explanations from SOTA LLMs

Annotations on a Budget: Leveraging Geo-Data Similarity to Balance Model Performance and Annotation Cost   June 2023 – Sept 2023
Supervisor: Prof. Rada Mihalcea, Prof. Oana Ignat, University of Michigan

- Analyzed visual diversity across 52 countries and 94 topics using three visual representations: CLIP, BLIP-2, and ALIGN, to understand common topic distribution.
- Identified underrepresented countries and topics in the training data of vision-language models, suggesting areas for improved data collection.

- Demonstrated that cosine similarity based data augmentation method can be used as an active learning approach to improve the model performance of underrepresented countries.

INTERNSHIP EXPERIENCE

Ohio State University June 2024 - Sept 2024
Summer Research Intern
- Research on the step generalization ability of transformers on composition task
- Exploration on the LLM interpretability and factual knowledge editing

Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences Mar 2022 - June 2022
Research Assistant
- Electrical control based on STM32 and circuit diagram drawing for white light interferometer.
- Realized the data acquisition, output, and processing of the white light interferometer.

TEACHING AND PRESENTATION

Paper Presenter, NLP Reading Group, University of Michigan 2023-2024

Poster Presentation, Midwest Language and Speech Days, Ann Arbor, MI 2024

Poster Presentation, Michigan AI Symposium, Ann Arbor, MI 2023

Teaching Assistant, C++ Program Design, Beijing Forestry University 2022

AWARDS AND HONOURS

Outstanding Graduates, Beijing Forestry University 2022

Excellent Graduation Design Award, Beijing Forestry University 2022

Academic Excellence Scholarship, Beijing Forestry University 2021

Outstanding Student Scholarship, Beijing Forestry University 2021

Second Prize, China Undergraduate Mathematical Contest in Modeling 2020

SKILLS

Machine Learning: PyTorch, TensorFlow, Scikit-learn, Hugging Face, OpenCV, Pandas

Programming: C, C++, Python, Java, Julia, Matlab, Assembly language

Language: Chinese (Native), English (GRE: 332, Verbal: 162, Quantitative: 170, AW: 4.0), Japanese (Basic)

Others: ROS, Linux, Matlab (Simulink), SolidWorks, AutoCAD, SLAM, Keil

Leadership

Event Leader, NLP Reading Group, University of Michigan 2024

Association President, Beijing Forestry University Computer Association 2019-2021

VOLUNTEERING AND CO-CURRICULARS

Volunteer, Midwest Language and Speech Days, MI 2024

Finisher, Ann Arbor Marathon, MI 2023

Volunteer, Songtang Caring Hospital, China Aged Foundation 2021-2022

Volunteer, Autistic Children Care Center, Yuhua Health Service Center 2020-2021

Level 8 Certification of Bamboo Flute, Chinese National Orchestral Society 2020

Volunteer, Mental Health Center, Peking University Sixth Affiliated Hospital 2019-2020