Yang Liu

E-mail: yangliu@nlp.ist.i.kyoto-u.ac.jp/yangliu.nlp@gmail.com   Website: nlply.tech

Education

Kyoto University, Japan

Ph.D. in Department of Intelligence Science and Technology   *Oct. 2024 - Now*
(advised by Prof.Chenhui Chu)

Tianjin University, China

Master Degree of Computer Technology (advised by Prof.Yuexian Hou)   *Sept. 2019 - Dec. 2022*
Thesis title: *Mining Effective Features Using Quantum Entropy for Humor Recognition*

Tianjin Normal University, China

Bachelor Degree of Software Engineering   *Sept. 2012 - Jun. 2016*

Experiences

Independent Researcher   *Mar. 2023 - Sept. 2024*
Research on social biases in language models.

Antrou Co. Ltd.   *Sept. 2019 - Mar. 2023*
Full-Stack Software Engineer (part-time)
Technical Support.

Antrou Co. Ltd.   *Jul. 2017 - Sept. 2019*
Full-Stack Software Engineer
Geographical Information System base on Three.js and Security Management System for Petrochemical Industry.

Publications

Yang Liu, Chenhui Chu. *Understanding the Prompt Sensitivity.* under review, 2025.

Yang Liu, Chenhui Chu. *Do LLMs Align Human Values Regarding Social Biases? Judging and Explaining Social Biases with LLMs.* EMNLP, 2025.

Yang Liu, Masahiro Kaneko, Chenhui Chu. *On the Alignment of Large Language Models with Global Human Opinion.* under review, 2025.

Yang Liu, Chenhui Chu. *Generating Explanations of Stereotypical Biases with Large Language Model.* NLP, 2025.

Yang Liu. *Quantifying Stereotypes in Language.* EACL, 2024. (**single-authored work**)

Yang Liu. *Robust Evaluation Measures for Evaluating Social Biases in Masked Language Models.* AAAI, 2024. (**single-authored work**)

Yang Liu, Yuexian Hou. *Syntax-Aware Complex-Valued Neural Machine Translation.* ICANN, 2023.

Yang Liu, Yuexian Hou. *Mining Effective Features Using Quantum Entropy for Humor Recognition.* EACL (Short Fingdings), 2023.

Technical Skills

Programming: Python, PyTorch, Java, Javascript, CSS

Techniques: Natural Language Processing, Deep Learning, Machine Learning, Quantum Theory, WEB Application Development

Links

Personal Blog: https://www.nlply.tech
Github: https://github.com/nlply
Huggingface: https://huggingface.co/lauyon

Reviewer

ARR, AAAI