

YINHONG LIU

+44 7923592505 ◇ yl535@cam.ac.uk

Homepage: <http://www.yinhongliu.com>

EDUCATION EXPERIENCE

University of Cambridge, Language Technology Lab; PhD 2021.10 - present

- Supervisor: Prof. Nigel Collier
- Research area: Large Language Model (LLM) Evaluation and Alignment. Text and Multi-modal Evaluation Techniques. Controllable Text Generation. Data Synthesis for Alignment and Down-stream applications. Advanced Dialog Systems. Discourse Structure Analysis for Long-form Text.
- Scholarship: Cambridge Trust

University of Cambridge, Computer Lab; Mphil 2020 - 2021

- Subject: **Advanced Computer Science, in Natural Language Processing.**
- Grade: **Distinction.**
- Research area: Learning Functional Distributional Semantics from Multi-modal data.

University of Cambridge, Engineering Department; BA, MEng 2014 - 2018

- Subject: **Information Engineering and Computer Science.**
- Grade in Master of Engineering: **Merit degree** with honours.
- Grade in Bachelor of Arts: **1st Class.**
- Final year project: **Statistical enhancement of multi-channel audio data.**

WORK, RESEARCH EXPERIENCE

Amazon AWS AI; Seattle. *Applied Scientist Intern* Oct. 2024 - Feb. 2025

- Develop Multi-Modal Summarization meta-Evaluation benchmark and evaluation methods.

Mircrosoft Research; Cambridge. *Research Intern* Jun. 2024 - Sep. 2024

- Developed an advanced Automatic Prompt Optimization pipeline to enhance Model Migration efficiency and improve Model Routing for Microsoft Copilot teams.

AI/ML, Apple; Cambridge. *Research Intern* Jun. 2023 - Oct. 2023

- Developed an innovative automatic Task-Oriented Dialog generation pipeline, incorporating advanced analysis of future virtual assistant response styles and simulating complex apps interactions. This work benefits both internal Apple Intelligence development and external NLP community.
- Authored a research paper detailing the dialog generation methodology, which was accepted by ACL 2024.

Gambit Research; London. *Full-time Quant-analyst* Oct. 2018 - Jul. 2020

- Working on various statistical football teams & games modelings, market price & volume predictions, strategies developing and back-testing.

Jigsaw Unintended Bias in Toxicity Classification, Kaggle competition Jun. 2019

- Got a 225th place out of 3165 teams (top 8%), a bronze medal, as a solo participant.

PUBLICATIONS

- **Yinhong Liu**, Zhijiang Guo, Tianya Liang, Ehsan Shareghi, Ivan Vulic and Nigel Collier. *Aligning with Logic: Measuring, Evaluating and Improving Logical Consistency in Large Language Models*. ICML 2025, Spotlight.
- **Yinhong Liu**, Han Zhou, Zhijiang Guo, Ehsan Shareghi, Ivan Vulic, Anna Korhonen and Nigel Collier. *Aligning with Human Judgement: The Role of Pairwise Preference in Large Language Model Evaluators*. COLM 2024.
- Sharan Maiya, **Yinhong Liu**, Ramit Debnath, Anna Korhonen. *Improving Preference Extraction In LLMs By Identifying Latent Knowledge Through Classifying Probes*. ACL 2025.
- Yingjian Chen, Haoran Liu, **Yinhong Liu**, et al. *GraphCheck: Breaking Long-Term Text Barriers with Extracted Knowledge Graph-Powered Fact-Checking*. ACL 2025.
- Yijiang River Dong, Tiancheng Hu, **Yinhong Liu**, Ahmet Üstün, et al. *When Personalization Meets Reality: A Multi-Faceted Analysis of Personalized Preference Learning*. Under review.
- Zongqian Li*, **Yinhong Liu***, Yixuan Su and Nigel Collier. *Prompt Compression for Large Language Models: A Survey*. NAACL 2025.
- **Yinhong Liu**, Yimai Fang, David Vandyke and Nigel Collier. *TOAD: Task-Oriented Automatic Dialog generation pipeline with diverse system response style*. ACL 2024.
- **Yinhong Liu**, Yixuan Su, Ehsan Shareghi and Nigel Collier. *Unlocking Structure Measuring: Introducing PDD, an Automatic Metric for Positional Discourse Coherence*. NAACL 2024.
- Han Zhou, Xingchen Wan, **Yinhong Liu**, Nigel Collier, Ivan Vulić, Anna Korhonen. *Fairer Preferences Elicit Improved Human-Aligned Large Language Model Judgments*. EMNLP 2024, main conference.
- Zhongshen Zeng, **Yinhong Liu**, Yingjia Wan, Jingyao Li, et al. **Mr-Ben: A meta-reasoning benchmark for evaluating system-2 thinking in llms**. NeurIPS 2024
- **Yinhong Liu**, Yixuan Su, Ehsan Shareghi and Nigel Collier. *Plug-and-play Recipe Generation with Content Planning*. EMNLP 2022, Generation, Evaluation & Metrics Workshop.
- **Yinhong Liu** and Guy Emerson. *Learning Functional Distributional Semantics with Visual Data*. ACL 2022, main conference.
- **Yinhong Liu** and Simon Godsill. *Multi-channel audio statistical restoration*. IEEE-ICSPCC, 2020.

PRIZE & EXTRA-CURRICULUM ACTIVITIES

Best Reviewer Award, EMNLP 2023, Natural Language Generation track.

PhD scholarship: Cambridge International Scholarship by Cambridge Trust.

Academic scholarship: Caldwell Scholarship by Corpus Christi College, University of Cambridge

Cambridge Chinese Drama Society: Protagonist in a university play, *Our Assassin*; Producer of two university plays *The Robin Murders* and *The Village*.

LANGUAGE & SKILLS

Languages: Chinese Mandarin (native), English (fluent), Chinese Cantonese (fluent).

Programming Languages: Python, SQL, Latex, C++.