

Fangru Lin

Tel: +44 7503270537 | E-mail: fangru.lin@balliol.ox.ac.uk

EDUCATION

University of Oxford

DPhil Linguistics, Philology, and Phonetics (Oxford e-Research Centre)

Oct. 2023 – Present

MPhil Linguistics, Philology and Phonetics

Oct. 2021– Jun. 2023

- Focus: **Natural Language Processing** (Overall Result: **Distinction**)
- Supervisor: Prof Janet Pierrehumbert, Prof Daniel Altshuler

Shanghai International Studies University (SISU)

B.A in Korean

Sept. 2017 – Jul. 2021

Certificate: Honors Program (Multilingual Advanced Interpreting and Translation)

Sept. 2017 – Jul. 2021

- Overall GPA: **3.83/4.0(90.91/100)**

SERVICE

Google DeepMind (Mountain View, CA)

Research Intern

Sept. 2025 - Dec. 2025

- Researched in the Gemini Deep Research team

Microsoft Research

Research Intern

Apr. 2024 - Oct. 2024

- Researched Large Language Models and produced a publication

Microsoft Corporation

Software Engineering Intern

Jul. 2022 - Oct. 2022

- Worked for Azure Storage in the Cloud+AI team

RECENT PUBLICATIONS

[Under review] Can Large Language Models Generalize Procedures Across Representations?
(<http://arxiv.org/abs/2602.03542>)

- Created parallel datasets with isomorphic procedures across three domains (code, graph, natural language)
- Experimented with popular SFT and RL methods, including GRPO and distillation, and found that none of the methods enable generalization across domains
- Developed an RL data curriculum that enables Qwen-2.5-1.5B to outperform GPT-4o-mini on a natural language planning task

[ACL 2025 main] One Language, Many Gaps: Evaluating Dialect Fairness and Robustness of Large Language Models in Reasoning Tasks (<https://arxiv.org/abs/2410.11005>)

- Collected ReDial, a dataset of Standardized English-African American Vernacular English (AAVE) parallel prompts in four canonical reasoning tasks (algorithm, math, logic, comprehensive reasoning)
- Found that SotA LLMs exhibit significant unfairness and brittleness in prompts expressed in AAVE
- Empirically showed the dialect unfairness and brittleness cannot be easily explained by AAVE data skewness and that simple prompt engineering method cannot mitigate the gap

[ICML 2024] Graph-enhanced Large Language Models in Asynchronous Plan Reasoning
(<https://arxiv.org/abs/2402.02805>)

- Automatically generated benchmark to assess Large Language Models' ability to execute complex plans at scale
- Proposed an off-the-shelf method to re-structure text inputs as graphs to improve LLM performance

[LREC-COLING 2024] Probing Large Language Models for Scalar Adjective Lexical Semantics and Scalar Diversity Pragmatics (<https://arxiv.org/abs/2404.03301>)

- Probed LLMs for their knowledge of scalar adjective lexical semantics and scalar diversity pragmatics
- Provided rationales for why LLMs do not have similar performance in the semantic and pragmatic tasks

SELECTED AWARDS

- Clarendon Scholarship (Oxford, 2023, full funding for graduate students)
- Jason Hu Scholarship (Oxford, 2023, full funding for graduate students)

ADDITIONAL INFORMATION

- Programming skills: Python (5-year experience), C# (3-month experience), Java (1-month experience)