Chenxi Whitehouse

RESEARCH SCIENTIST · NLP · LLM · META GenAI

✉ chenxi.whitehouse@gmail.com | ⌂ chenxwh.github.io | ⬛ chenxwh | 🔲 chenxwh

Professional Experience

RESEARCH

2024	**Research Scientist, Meta GenAI**, London, UK
2024	**Applied Research Scientist, Amazon AGI**, Cambridge, UK
2023	**Research Intern, Google DeepMind**, Amsterdam, Netherlands
2023	**Research Assistant, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI)**, Remote
2022	**Research Intern, Amazon**, Alexa AI, Cambridge, UK
2021	**Research Intern, Huawei**, Noah's Ark Lab, London, UK

ENGINEERING

2021-2024	**Machine Learning Engineer**, Replicate AI, Remote
2019-2020	**Software Engineer**, HP Inc, London, UK
2018	**Software Engineer Intern**, Siemens AG, Nuremberg, Germany

Education

University of Cambridge — *Cambridge, UK*
POSTDOC RESEARCH ASSOCIATE IN NATURAL LANGUAGE PROCESSING — *2024*

City, University of London — *London, UK*
PHD IN COMPUTER SCIENCE, NATURAL LANGUAGE PROCESSING — *2020 - 2023*
- Thesis: Towards Knowledge-Grounded Natural Language Understanding and Generation.
- Research Focus: Knowledge-intensive NLP, Multilingual NLP

University College London & University of Erlangen-Nuremberg — *London, UK & Erlangen, Germany*
MSc ELECTRICAL ENGINEERING & SIGNAL PROCESSING — *2015 - 2018*

Xi'an Jiaotong University — *Xi'an, China*
BSc INFORMATION & COMPUTER ENGINEERING — *2011 - 2015*

Selected Publications

PUBLISHED

1. Zhangdie Yuan, **Chenxi Whitehouse**[†], Eric Chamoun[†], Rami Aly[†], Andreas Vlachos. "PRobELM: Plausibility Ranking Evaluation for Language Models". In *First Conference on Language Modeling: CoLM 2024*. [†Equal Contribution]

2. David Romero, Chenyang Lyu, Haryo Akbarianto Wibowo, …, **Chenxi Whitehouse**, …, Thamar Solorio, and Alham Fikri Aji. "CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark". In *Advances in Neural Information Processing Systems: NeurIPS 2024*.

3. Holy Lovenia, Rahmad Mahendra, Salsabil Maulana Akbar, …, **Chenxi Whitehouse**, …, and Samuel Cahyawijaya. "SEA-Crowd: A Multilingual Multimodal Data Hub and Benchmark Suite for Southeast Asian Languages". In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024*.

4. **Chenxi Whitehouse**, Fantine Huot, Jasmijn Bastings, Mostafa Dehghani, Chu-Cheng Lin, and Mirella Lapata. "Low-Rank Adaptation for Multilingual Summarisation: An Empirical Study". In *Findings of the Association for Computational Linguistics: NAACL 2024*. *(Work conducted as research intern at Google DeepMind)*

5. Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, **Chenxi Whitehouse**, Osama Mohammed Afzal, Tarek Mahmoud, Alham Fikri Aji, and Preslav Nakov. "M4: Multi-generator, Multi-domain, and Multi-

lingual Black-Box Machine-Generated Text Detection". In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: **EACL 2024**. (Best Resource Paper Award)*

6. **Chenxi Whitehouse**, Monojit Choudhury, and Alham Fikri Aji. "LLM-powered Data Augmentation for Enhanced Cross-lingual Performance". In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: **EMNLP 2023**. (Work conducted as visiting researcher at MBZUAI)*

7. **Chenxi Whitehouse**, Clara Vania, Alham Fikri Aji, Christos Christodoulopoulos, and Andrea Pierleoni. "WebIE: Faithful and Robust Information Extraction on the Web". In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistic: **ACL 2023**. (Work conducted as research intern at Amazon)*

8. **Chenxi Whitehouse**, Tillman Weyde, and Pranava Madhyastha. "Towards a Unified Model for Generating Answers and Explanations in Visual Question Answering". In: *Findings of the Association for Computational Linguistics: **EACL 2023**.*

9. **Chenxi Whitehouse**, Fenia Christopoulou, and Ignacio Iacobacci. "EntityCS: Improving Zero-Shot Cross-lingual Transfer with Entity-Centric Code Switching". In: *Findings of the Association for Computational Linguistics: **EMNLP 2022**. (Work conducted as research intern at Huawei)*

10. **Chenxi Whitehouse**, Tillman Weyde, Pranava Madhyastha, and Nikos Komninos. "Evaluation of Fake News Detection with Knowledge-Enhanced Language Models". In: *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media: **AAAI-ICWSM 2022**.*

PREPRINT

11. Xiaochen Zhu, Georgi Karadzhov, **Chenxi Whitehouse**, and Andreas Vlachos. "Segment-Level Diffusion: A Framework for Controllable Long-Form Generation with Diffusion Language Models".

12. Chu-Cheng Lin, Xinyi Wang, Jonathan H. Clark, Yun Zhu, Han Lu, **Chenxi Whitehouse**, and Hongkun Yu. "Inducing Generalization across Languages and Tasks using Featurized Low-Rank Mixtures". *(Work conducted as research intern at Google DeepMind)*

Additional Experiences and Awards

2024	**Area Chair**, COLING 2025
2024	**Seventh FEVER Workshop Organisation**, co-located with EMNLP 2024, Miami, Florida
2024	**SemEval-2024 Task-8 Organisation**, co-located with NAACL 2024, Mexico City, Mexico
2023	**EMNLP 2023 Student Volunteer**, Singapore
2022	**EMNLP 2022 Student Volunteer**, Abu Dhabi
2022	**Lisbon Machine Learning School**, Instituto Superior Técnico, with Google Scholarship
2022	**Oxford Machine Learning Summer School**, University of Oxford
2020-2023	**PhD Studentship**, City, University of London
2016-2017	**Erasmus Scholarship**, University College London
2012-2015	**Outstanding Student Scholarship**, Xi'an Jiaotong University