

CASS ZHIXUE ZHAO

44 7516862694
zhixue.zhao@sheffield.ac.uk
S1 4BP Sheffield
[Personal Site](#)
[Google Scholar](#)

I am Cass Zhixue Zhao, a lecturer in Natural Language Processing (NLP) at the Department of Computer Science, University of Sheffield. Prior to this role, I was a Postdoc researcher, supervised by Prof. Nikos Aletras, focusing on model interpretability and model compression. Back in 2020, I worked as a research assistant within the same department, working on NIHR-funded NLP projects for systematic reviews of public health research. My Ph.D. research applied transfer learning to hate speech detection tasks and investigated model bias, which was funded by the University of Sheffield.

My research interests include model explanation, model hallucination, model compression, and language model reasoning capability. I am dedicated to conducting innovative and responsible research. I have a keen interest in interdisciplinary research, exploring the application of AI, particularly NLP models, across different domains. I've served on the Program Committee for AAAI, ACL, and EMNLP. Apart from research, I find joy in teaching and have been honored with an Associate Fellowship in Higher Education Certificate.

Education

01/2018~08/2022	The University of Sheffield , PhD, Information School (fully funded)
Research Topic:	Transfer Learning for Hate Speech Classification, Bias in pre-trained language models
09/2017~09/2018	The University of Sheffield , M.Sc. Data Science, Distinction
Core Modules:	Data Analysis, Database Design, Data Mining and Visualization.
09/2008~/2012	Shanghai Institute of Technology , B.S. Food Engineering
Core Module:	Advanced Mathematics, Linear Algebra, Microbiology, Chemistry

Publications

Zhao, Z., N. Aletras (2023). [Incorporating Attribution Importance for Improving Faithfulness Metrics](#). The 61st Annual Meeting of the Association for Computational Linguistics. ACL2023 (oral presentation).

Zhao, Z., G. Chrysostomou, K. Bontcheva and N. Aletras (2022). [On the Impact of Temporal Concept Drift on Model Explanations](#). In Findings of the Association for Computational Linguistics: EMNLP 2022.

Zhao, Z., Zhang, Z., Hopfgartner, F. (2022). [Utilizing Subjectivity Level to Mitigate Identity Term Bias in Toxic Comments Classification](#). Online Social Networks and Media, 29, 100205.

Clowes, M., Stansfield, C., Thomas, J., Shemilt, I., Paisley, S., Stevenson, M., Zhao, Z., Marshall, I., Kell, G., (June 2022). [All is FAIR in health inequalities research](#): using machine learning to build a new database of health equity studies. European Association for Health Information and Libraries 2022.

Zhao, Z., Zhang, Z., Hopfgartner, F. (2021). [A Comparative Study of Using Pre-trained Language Models for Toxic Comment Classification](#). In Companion Proceedings of the Web Conference 2021 (pp. 500-507)

Zhao, Z., Zhang, Z., Hopfgartner, F. (2019). [Detecting Toxic Content Online and the Effect of Training Data on Classification Performance](#). In Proceedings of 20th International Conference on Computational Linguistics and Intelligent Text Processing, La Rochelle, France

Working Experience

11/2023~Now Department of Computer Science, The University of Sheffield

Lecturer in NLP

Teaching computer science modules at the undergraduate or graduate level. Curriculum development, course design, and assessment methods. Supervising and mentoring undergraduate and graduate students in research projects, theses, and career development. Research and grant applications.

01/2022~11/2023 Department of Computer Science, The University of Sheffield

Postdoc Researcher Model Explanation

This research is part of a large interdisciplinary project, "Social Explainable Artificial Intelligence", funded by the EPSRC. The overall goal of this project is to explore a wide range of research directions on responsible AI and model interpretability. I work independently and with supervision on programming and compiling papers. The main output is research publications.

01/2023~09/2023 (expected) Information School, The University of Sheffield

Postgraduate Research Supervisor

As a dissertation research supervisor, I provide advice on the dissertation project, regarding the research design, time management, programming, results analysis, and writing-up.

07/2022~08/2022 English and American Studies, The University of Manchester

Research Assistant Nineteenth-Century Nature Writing in English and Twenty-First-Century Environmentalism

This pilot project aims to open out research questions for a larger project on women's nature writing in the long nineteenth century. My duty focuses on pre-processing related text and exploratory research on the related corpus, such as temporal shift and geographic drift.

07/2021~12/2021 Department of Computer Science, The University of Sheffield

Research Assistant [Automatically mapping and assessing inequalities in public health research](#)

This is a multi-discipline project, collaborated with UCL and King's College, funded by the National Institute for Health Research. This project aims to develop automated methods to find, organize, and describe scientific literature relevant to public health and understand its findings in relation to inequalities in health. Particularly, NLP techniques will be used to automatically organize these documents into topics, identify the research method used, and identify whether it contains information about factors related to health inequalities at scale. My role was for programming and analysis.

11/2020~03/2021 Department of Computer Science, The University of Sheffield

Research Assistant Text Processing for Health Technology Assessment

The TePHTA is a cross-discipline study with researchers from the Computer Science Department and the School of Health and Related Research, University of Sheffield. This project aims to improve healthcare decision-making through natural language processing. I participate in a part of this project, the medical system review analysis, which utilizes machine learning to assist experts in health to gain insights from a massive volume of medical literature regarding employees' health situations and working conditions. I was in charge of programming and analysis.

12/2018~03/2021 Information School, The University of Sheffield, Sheffield

Teaching Assistant (for below modules)

- INF4002 Introduction to programming
- INF6027 Introduction to data science
- INF6060 Information retrieval
- INF6028 Data mining and visualization
- INF6032 Big data analytics
- INF6050 Database design
- INF6024 Researching social media
- INF109 Digital media and society

Society & Club

2023~Now **Communication officer** Staff Race Equality Network, The University of Sheffield

2021~2021 **Organizer** NLP Reading Group, The University of Sheffield

2019~2021 **Research students representative** Information Retrieval Research Group, The University of Sheffield

2019~2020 **Inclusion Officer** iSchool society, The University of Sheffield