Xiang Dai

dai.xiang.au@gmail.com

I am currently a research scientist at CSIRO Data61. My research interests lie in natural language processing and digital health. My work focuses on several areas, including extracting entities and their relations from unstructured text, pre-training domain-specific language representation models, and applying NLP techniques to the medical domain.

Education

University of Sydney, Ph.D, 2017 - 2021

University of Sydney, Master, 2015 - 2017

South China University of Technology, Bachelor, 2003-2007

Experience

CSIRO Data61, Research scientist, Aug 2023 - Present

CSIRO Data61, Postdoc, Dec 2021 - Aug 2023

University of Copenhagen, Postdoc, Feb 2021 - Dec 2021

PhD Supervision

Farhana (Hanna) Pethan (Feb 2022-Dec 2024), University of Sydney

Mong Yuan Sim (March 2024-present), University of Adelaide

Selected Publications

A more complete publication list can be found on Google Scholar.

5. Mong Yuan Sim, Wei Emma Zhang, Xiang Dai, and Biaoyan Fang, "Can VLMs Actually See and Read? A Survey on Modality Collapse in Vision-Language Models", in Findings of ACL, 2025.
4. Xiang Dai, Sarvnaz Karimi, Abeed Sarker, Ben Hachey, and Cecile Paris, "MultiADE: A Multi-domain benchmark for Adverse Drug Event extraction", JBI, vol. 160, p. 104744, 2024.
3. Xiang Dai, Ilias Chalkidis, Sune Darkner, and Desmond Elliott, "Revisiting Transformer-based Models for Long Document Classification", in Findings of EMNLP, 2022.
2. Xiang Dai and Heike Adel, "An Analysis of Simple Data Augmentation for Named Entity Recognition", in COLING, 2020.
1. Xiang Dai, Sarvnaz Karimi, Ben Hachey, and Cecile Paris, "An Effective Transition-based Model for Discontinuous NER", in ACL, 2020.