

# Xiang Dai

Postdoc at  
CSIRO Data61

Email: [dai.xiang.au@gmail.com](mailto:dai.xiang.au@gmail.com)

## Summary

I am currently a Postdoc at CSIRO Data61, working on NLP. My research interests focus on information extraction in the biomedical domain. Some problems I work on include recognizing complex (discontinuous, overlapping) entities; pre-training domain-specific language representation models; and identifying health risks from clinical notes.

## Education

### **University of Sydney**

Ph.D., Natural Language Processing, 2017 - 2021

Supervisors: Sarvnaz Karimi, Ben Hachey, Cecile Paris, Joachim Gudmundsson

Thesis: Recognising Biomedical Names: Challenges and Solutions

### **University of Sydney**

Master's Degree, Information Technology and Information Technology Management, 2015 - 2017

Supervisors: Joachim Gudmundsson, Michael Horton

### **South China University of Technology**

Bachelor's Degree, Computer Science, 2003-2007

## Experience

### **CSIRO Data61** [Sydney, Australia]

Postdoc Researcher, Dec 2021 - Present

### **University of Copenhagen** [Copenhagen, Denmark]

Postdoc Researcher, Feb 2021 - Dec 2021

### **Saarland University** [Saarbrücken, Germany]

Visiting Researcher, Aug 2020 - Jan 2021, hosted by Dietrich Klakow

### **Bosch Center for Artificial Intelligence** [Stuttgart, Germany]

Research Intern, Mar - Jul 2020

### **University of Sydney** [Sydney, Australia]

Teaching Assistant, 2017 - 2019

### **CSIRO Data61** [Sydney, Australia]

Research Assistant, Feb - Jul 2017

Summer Scholar, Nov 2016 - Feb 2017

### **Lanke Technology Co., Ltd.** [Guangzhou, China]

Software Developer, 2010 - 2014

### **Qiaoqing Technology Co., Ltd.** [Guangzhou, China]

Software Developer, 2008 - 2010

## Publications

J: journal; C: conference; W: workshop

- C7 mDAPT: Multilingual Domain Adaptive Pretraining in a Single Model  
Rasmus Kær Jørgensen, Mareike Hartmann, **Xiang Dai**, Desmond Elliott  
Findings of the Association for Computational Linguistics: EMNLP 2021
- C6 SearchEHR: A Family History Search System for Clinical Decision Support  
**Xiang Dai**, Maciej Rybinski, Sarvnaz Karimi  
International Conference on Information and Knowledge Management (CIKM 2021)
- J1 Extracting Family History Information From Electronic Health Records: Natural Language Processing Analysis  
Maciej Rybinski, **Xiang Dai**, Sonit Singh, Sarvnaz Karimi, Anthony Nguyen  
JMIR Medical Informatics, 2021
- C5 An Analysis of Simple Data Augmentation for Named Entity Recognition  
**Xiang Dai**, Heike Adel  
International Conference on Computational Linguistics (COLING 2020)
- C4 Cost-effective Selection of Pretraining Data: A Case Study of Pretraining BERT on Social Media  
**Xiang Dai**, Sarvnaz Karimi, Ben Hachey, Cecile Paris  
Findings of the Association for Computational Linguistics: EMNLP 2020
- W5 NLNDE at CANTEMIST: Neural Sequence Labeling and Parsing Approaches for Clinical Concept Extraction  
Lukas Lange, **Xiang Dai**, Heike Adel, Jannik Strötgen  
Iberian Languages Evaluation Forum (IberLEF 2020)
- C3 An Effective Transition-based Model for Discontinuous NER  
**Xiang Dai**, Sarvnaz Karimi, Ben Hachey, Cecile Paris  
Annual Meeting of the Association for Computational Linguistics (ACL 2020)
- C2 NNE: A Dataset for Nested Named Entity Recognition in English Newswire  
Nicky Ringland, **Xiang Dai**, Ben Hachey, Sarvnaz Karimi, Cecile Paris, James R. Curran  
Annual Meeting of the Association for Computational Linguistics (ACL 2019)
- C1 Using Similarity Measures to Select Pretraining Data for NER  
**Xiang Dai**, Sarvnaz Karimi, Ben Hachey, Cecile Paris  
Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)
- W4 Shot Or Not: Comparison of NLP Approaches for Vaccination Behaviour Detection  
Aditya Joshi, **Xiang Dai**, Sarvnaz Karimi, Ross Sparks, Cecile Paris, C Raina MacIntyre  
EMNLP Workshop on Social Media Mining for Health Applications (SMM4H 2018)  
Shared task system paper: our system ranked 1st for task 4 of the SMM4H 2018 shared task among all 17 systems.
- W3 Recognizing Complex Entity Mentions: A Review and Future Directions  
**Xiang Dai**  
ACL Student Research Workshop (ACL-SRW 2018)
- W2 Medication and Adverse Event Extraction from Noisy Text  
**Xiang Dai**, Sarvnaz Karimi, Cecile Paris  
Australasian Language Technology Association Workshop (ALTA 2017)

- W1 Automatic Diagnosis Coding of Radiology Reports: A Comparison of Deep Learning and Conventional Classification Methods  
Sarvnaz Karimi, **Xiang Dai**, Hamed Hassanzadeh, Anthony Nguyen  
ACL Workshop on Biomedical Natural Language Processing (BioNLP 2017)

## Other Academic Activities

### Journal reviewing

ACM Transactions on Intelligent Systems and Technology (2018)  
International Journal of Medical Informatics (2018,2021)  
Language Resources and Evaluation (2020)  
Computer Speech And Language (2021)  
Informatics in Medicine Unlocked (2022)

### Conference PC (reviewing)

ACL 2020, 2021  
ACL Rolling Review 2021, 2022  
EACL 2021 (Outstanding Reviewer)  
EMNLP 2020 (Outstanding Reviewer), 2021  
ICLR 2022  
NAACL 2021  
NeurIPS 2021  
  
ALTA 2019, 2020, 2021  
Student Research Workshop at ACL 2020, 2021, 2022  
Student Research Workshop at AACL 2020  
Workshop on NLP for COVID-19 at EMNLP 2020  
Workshop on Weakly Supervised Learning at ICLR 2021

### Membership

ALTA (Australasian Language Technology Association) **executive committee**, 2019-2020  
ACL (Association for Computational Linguistics), 2018-present

## Honors and Awards

### Outstanding Reviewer

EMNLP, 2020  
EACL, 2021

### Postgraduate Research Support Scheme

USYD, 2019 and 2020

### Engineering and Information Technologies Research Scholarship

AUD 27K per annum, 3.5 years  
Faculty of Engineering and Information Technologies, USYD, 2018 - 2021

**Data61 top-up Scholarship**

AUD 10K per annum, 3.5 years

Data61, CSIRO, 2018 - 2021

**SIT Summer Scholarships**

AUD 6K

School of Information Technologies, USYD, November 2016 - February 2017

**Postgraduate Coursework High Honour Roll**

School of Information Technologies, USYD, April 2016, May 2017 and June 2018

**MIT and MITM Half-Fee Scholarship**

School of Information Technologies, USYD, April 2016

Last updated: March 25, 2022