Cheng Xu

Email: xucheng19991201@gmail.com Homepage: https://xucheng.space/
Last updated: Aug 20, 2025

EDUCATION

University College Dublin, Dublin, Ireland — Sep 2022 - Aug 2026 (Expected)
Ph.D., Computer Science
Supervisor: Prof. M-Tahar Kechadi
Thesis Title: Multimodal Fuzzy System for Fake News Detection

Huaibei Normal University, Huaibei, P.R. China — Sep 2018 - Jun 2022
B.Eng., Data Science and Big Data Technology
Mentor: Prof. Qi Fan & Prof. Longfeng Shen
GPA: 90.34/100, Graduated with Provincial Honor (Highest for graduates)

PUBLICATIONS

Full list of publications on Google Scholar.

** indicates corresponding author, † indicates co-first authorship, ‡ indicates a student I mentored.*

Conferences Paper

[7] **Cheng Xu**, Nan Yan, Shuhao Guan, Yuke Mei, M-Tahar Kechadi. SSA: Semantic Contamination of LLM-Driven Fake News Detection (*EMNLP 2025*)

[6] **Cheng Xu**, Nan Yan, Shuhao Guan, Changhong Jin, Yuke Mei, Yibing Guo, M-Tahar Kechadi. DCR: Quantifying Data Contamination in LLMs Evaluation (*EMNLP 2025*)

[5] **Cheng Xu**, Nan Yan. TripleFact: Defending Data Contamination in the Evaluation of LLM-driven Fake News Detection (*ACL 2025*)

[4] Shuhao Guan, Moule Lin, **Cheng Xu**, Xinyi Liu, Jinman Zhao, Jiexin Fan, Qi Xu, Derek Greene. PreP-OCR: A Complete Pipeline for Document Image Restoration and Enhanced OCR Accuracy (*ACL 2025*)

[3] Shuhao Guan, **Cheng Xu**, Moule Lin, Derek Greene. Effective Synthetic Data and Test-Time Adaptation for OCR Correction (*EMNLP 2024*)

[2] Yang Fan‡, **Cheng Xu***, Shuhao Guan, Nan Yan, Yuke Mei. Advancing Arabic Sentiment Analysis: ArSen Benchmark and the Improved Fuzzy Deep Hybrid Network (*CoNLL 2024*)

[1] **Cheng Xu**, M-Tahar Kechadi. Fuzzy Deep Hybrid Network for Fake News Detection (*SOICT 2023*)

Journal Paper

[3] **Cheng Xu***, M-Tahar Kechadi. An Enhanced Fake News Detection System with Fuzzy Deep Learning (*IEEE Access, 2024*)

[2] **Cheng Xu**, Jing Wang, Tianlong Zheng, Yue Cao, Fan Ye. Prediction of prognosis and survival of patients with gastric cancer by a weighted improved random forest model: an application of machine learning in medicine (*Archives of Medical Science, 2022*)

[1] **Cheng Xu**, Qingling Chen, Fan Ye, Qi Fan, Qing Wang. Selection of surgical procedures and analysis of prognostic factors in patients with primary gastric tumour based on Cox regression: a SEER database analysis based on data mining (*Gastroenterology Review, 2021*)

Preprint

[1] **Cheng Xu***, Shuhao Guan, Derek Greene, M-Tahar Kechadi. Benchmark Data Contamination of Large Language Models: A Survey (*arXiv:2406.04244*)

HONORS

Graduate with Honor of Anhui Province (Top 3%), Anhui Province Government — 2022
Graduate with Honor (Top 10%), Huaibei Normal University — 2022
"Top 100 Outstanding College Students" honorary title, Anhui Province Government — 2021
"Top Ten College Students" honorary title, Huaibei Normal University — 2021
First Prize Scholarship for Innovation (Top 15), Huaibei Normal University — 2021
First Prize Scholarship (Top 3%) Huaibei Normal University — 2019 - 2021
Academic Excellence Award, Huaibei Normal University — 2019 - 2021

WORK EXPERIENCE

Ph.D. Researcher, Insight SFI Research Centre for Data Analytics — Sep 2022 - Present

- Received rigorous scientific training, emphasizing critical thinking and problem-solving skills.
- Established a comprehensive false information recognition system capable of processing multimodal inputs, including news articles, textual context, and multimedia information, laying the groundwork for a robust and adaptable tool in the fight against the diffusion of fake news.
- Spearheaded research efforts in leveraging deep learning architectures to tackle the multifaceted challenge of fake news detection, resulting in state-of-the-art performance on benchmark datasets such as LIAR.
- Published findings in reputable academic venues and made code implementations publicly available, fostering collaboration and further advancements in the field of fake news detection and mitigation.

Research Assistant, Data Mining Lab of Huaibei Normal University — Nov 2018 - Oct 2021

- Data mining and research on cancer data using machine learning techniques.
- Demonstrated leadership skills by leading and contributing to five research projects.
- Managed laboratory operations and resources in the third year of my undergraduate, enhancing my organizational and project management skills.

PROJECT EXPERIENCE

COVID-19 Low-resource NLP Benchmark Development

Project Leader — Jan 2023 – Aug 2024

- Managed a four-person team as Project Manager for the development of the COVID-19 Benchmark.
- Led the project resulting in the work like AROT-COV23 and Arsen, and published the papers at conferences such as CoNLL and AfricaNLP, which are prestigious venues in the field of NLP research.

Cancer Damage Prediction Using Data Mining and Machine Learning

Project Leader — Sep 2020 - Jun 2022

- Led and supervised a cancer damage prediction project that was funded by the National University Student Innovation and Entrepreneurship Program in P.R. China, under Grant 202010373032.
- Recommended the project achievements for participation in the 15th China University Students Innovation and Entrepreneurship Annual Conference.
- Spearheaded the entire development cycle of the project, including data preparation, collection, cleaning, imputation, modeling, and evaluation.
- Applied various machine learning models on the datasets, including both classic models (Random Forest, Decision tree) and our improved models.

TEACHING

TA/Demonstrator, University College Dublin — Sep 2023 – Present

Responsibilities: Assisted with lab work, graded assignments, supported students in learning activities for the following modules:

- COMP10040 Introduction to Computer Architecture (2023 Fall, 2024 Fall)
- COMP10050 Software Engineering Project 1 (2025 Spring)
- COMP10120 Computer Programming II (2024 Spring)
- COMP20230 Data Structure & Algorithms (Conv) (2024 Spring)
- COMP20280 Data Structures (2025 Spring)
- COMP30760 Data Science in Python (2024 Fall)
- COMP40370 Data Mining (2023 Fall, 2024 Fall, 2025 Fall)
- COMP40610 Information Visualisation (2024 Fall)
- COMP41680 Data Science in Python (2025 Spring)
- COMP47580 Recommender Systems & Collective Intelligence (2024 Spring)
- COMP47590 Advanced Machine Learning (2024 Spring)
- COMP47650 Deep Learning (2024 Spring, 2025 Spring)
- COMP47780 Cloud Computing (2023 Fall, 2024 Fall, 2025 Fall)
- COMP47980 Generative AI: Language Models (2024 Spring, 2025 Spring)

ACADEMIC SERVER

Conference Reviewer: NeurIPS, ICLR, ICML, AAAI, AISTATS

Journal Reviewer: Archives of Medical Science