

# Cheng Xu

Email: [xucheng19991201@gmail.com](mailto:xucheng19991201@gmail.com)   Homepage: <https://xucheng.space/>  
Last updated: Nov 1, 2025

<b>EDUCATION</b>	<b>University College Dublin</b> , Dublin, Ireland Ph.D., Computer Science Supervisor: Prof. M-Tahar Kechadi Thesis: Towards Trustworthy Fake News Detection: From Interpretation to Robust Evaluation of NLP Models	Sep 2022 - Aug 2026 (Expected)
	<b>Huaibei Normal University</b> , Huaibei, P.R. China 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)	Sep 2018 - Jun 2022
<b>PUBLICATIONS</b>	<i>Full list of publications on <a href="#">Google Scholar</a>.</i> <i>* indicates corresponding author; † indicates co-first authorship, ‡ indicates a student I mentored.</i>	
	<b>Conferences Paper</b>	
	<ul style="list-style-type: none"><li>[7] <b>Cheng Xu</b>, Nan Yan, Shuhao Guan, Yuke Mei, M-Tahar Kechadi. <a href="#">SSA: Semantic Contamination of LLM-Driven Fake News Detection (EMNLP 2025)</a></li><li>[6] <b>Cheng Xu</b>, Nan Yan, Shuhao Guan, Changhong Jin, Yuke Mei, Yibing Guo, M-Tahar Kechadi. <a href="#">DCR: Quantifying Data Contamination in LLMs Evaluation (EMNLP 2025)</a></li><li>[5] <b>Cheng Xu</b>, Nan Yan. <a href="#">TripleFact: Defending Data Contamination in the Evaluation of LLM-driven Fake News Detection (ACL 2025)</a></li><li>[4] <b>Cheng Xu</b>, M-Tahar Kechadi. <a href="#">Fuzzy Deep Hybrid Network for Fake News Detection (SOICT 2023)</a></li><li>[3] Shuhao Guan, Moule Lin, <b>Cheng Xu</b>, Xinyi Liu, Jinman Zhao, Jiejin Fan, Qi Xu, Derek Greene. <a href="#">PreP-OCR: A Complete Pipeline for Document Image Restoration and Enhanced OCR Accuracy (ACL 2025)</a></li><li>[2] Shuhao Guan, <b>Cheng Xu</b>, Moule Lin, Derek Greene. <a href="#">Effective Synthetic Data and Test-Time Adaptation for OCR Correction (EMNLP 2024)</a></li><li>[1] Yang Fan<sup>‡</sup>, <b>Cheng Xu</b><sup>*</sup>, Shuhao Guan, Nan Yan, Yuke Mei. <a href="#">Advancing Arabic Sentiment Analysis: ArSen Benchmark and the Improved Fuzzy Deep Hybrid Network (CoNLL 2024)</a></li></ul>	
	<b>Journal Paper</b>	
	<ul style="list-style-type: none"><li>[3] <b>Cheng Xu</b>, M-Tahar Kechadi. <a href="#">An Enhanced Fake News Detection System with Fuzzy Deep Learning (IEEE Access, 2024)</a></li><li>[2] <b>Cheng Xu</b>, Jing Wang, Tianlong Zheng, Yue Cao, Fan Ye. <a href="#">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)</a></li><li>[1] <b>Cheng Xu</b>, Qingling Chen, Fan Ye, Qi Fan, Qing Wang. <a href="#">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)</a></li></ul>	
	<b>Preprint &amp; Under Review</b>	
	<ul style="list-style-type: none"><li>[1] <b>Cheng Xu</b><sup>*</sup>, Shuhao Guan, Derek Greene, M-Tahar Kechadi. <a href="#">Benchmark Data Contamination of Large Language Models: A Survey (arXiv:2406.04244)</a></li></ul>	
<b>FUNDING</b>	<b>Benchmark Data Contamination Problem in LLMs</b> Principal Investigator, OpenAI Research Access Program, USD 10,000	Jan 2025 - Feb 2026
	<b>Cancer Prognosis Prediction Based on Data Mining and Machine Learning</b> Principal Investigator, National Undergraduate Innovation Training Program (China), CNY 10,000	Sep 2020 - Jun 2022

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

<b>RESEARCH EXPERIENCE</b>	<b>Ph.D. Researcher</b> , Insight SFI Research Centre for Data Analytics	Sep 2022 - Present
	<ul style="list-style-type: none"> <li>▪ Received rigorous scientific training, emphasizing critical thinking and problem-solving skills.</li> <li>▪ 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.</li> <li>▪ 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.</li> <li>▪ 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.</li> </ul>	
	<b>Research Assistant</b> , Data Mining Lab of Huaibei Normal University	Nov 2018 - Oct 2021
	<ul style="list-style-type: none"> <li>▪ Data mining and research on cancer data using machine learning techniques.</li> <li>▪ Demonstrated leadership skills by leading and contributing to five research projects.</li> <li>▪ Managed laboratory operations and resources in the third year of my undergraduate, enhancing my organizational and project management skills.</li> </ul>	

<b>TEACHING</b>	<b>TA/Demonstrator</b> , University College Dublin	Sep 2023 – Present
Responsibilities: Lead or assist with lab work, graded assignments, supported students in learning activities for the following modules:		
	<ul style="list-style-type: none"> <li>▪ COMP10040 Introduction to Computer Architecture (2023 Fall, 2024 Fall)</li> <li>▪ COMP10050 Software Engineering Project 1 (2025 Spring)</li> <li>▪ COMP10120 Computer Programming II (2024 Spring)</li> <li>▪ COMP20230 Data Structure &amp; Algorithms (Conv) (2024 Spring)</li> <li>▪ COMP20280 Data Structures (2025 Spring)</li> <li>▪ COMP30760 Data Science in Python (2024 Fall, 2025 Fall)</li> <li>▪ COMP40370 Data Mining (2023 Fall, 2024 Fall, 2025 Fall)</li> <li>▪ COMP40610 Information Visualisation (2024 Fall, 2025 Fall)</li> <li>▪ COMP41680 Data Science in Python (2025 Spring)</li> <li>▪ COMP41780 Foundations of Deep Learning (2025 Fall)</li> <li>▪ COMP47580 Recommender Systems &amp; Collective Intelligence (2024 Spring)</li> <li>▪ COMP47590 Advanced Machine Learning (2024 Spring)</li> <li>▪ COMP47650 Deep Learning (2024 Spring, 2025 Spring)</li> <li>▪ COMP47780 Cloud Computing (2023 Fall, 2024 Fall)</li> <li>▪ COMP47980 Generative AI: Language Models (2024 Spring, 2025 Spring)</li> </ul>	

<b>ACADEMIC SERVER</b>	<b>Conference Reviewer:</b> NeurIPS, ICLR, ICML, AAAI, AISTATS <b>Journal Reviewer:</b> Archives of Medical Science
------------------------	--