

Lexin Zhou

✉ Email: lexinzhou@gmail.com — 🌐 Homepage: <https://lexzhou.github.io/>

Education

University of Cambridge

MPhil of Advanced Computer Science (Track: NLP + HCI)

- GPA: 4.0/4.0 (Distinction)
- Advisor: [Andreas Vlachos](#)

Cambridge, UK
2023 – 2024

Universidad Politécnica de Valencia

Bachelor of Science, Data Science

- GPA: 3.93/4.0 (ranked top 1 in the cohort)
- Advisor: [Jose Hernandez-Orallo](#)

Valencia, Spain
2019 – 2023

Selected Publications

[Larger and More Instructable Language Models Become Less Reliable](#)

Lexin Zhou, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri, José Hernández-Orallo
Nature, 2024

Media Coverage: *Nature*, *Forbes*, *MIT Tech. Review*, *IEEE Spectrum*, *El País*, *New Scientists*, etc. & 290K views on X/Twitter

[An LLM Feature-based Framework for Dialogue Constructiveness Assessment](#)

Lexin Zhou, Youmna Farag, Andreas Vlachos
EMNLP, 2024

Top 1% of Submissions (Avg. Rev. Score = 4.17)

[Predictable Artificial Intelligence](#)

Lexin Zhou, Pablo Moreno-Casares, Fernando Martínez-Plumed, John Burden, [...], José Hernández-Orallo
Artificial Intelligence Journal, 2024. Under the 2nd Round of Review.

Other Publications

[Reject Before You Run: Small Assessors Anticipate Big Language Models](#)

Lexin Zhou, Fernando Martínez-Plumed, José Hernández-Orallo, Cèsar Ferri, Wout Schellaert
Evaluation Beyond Metrics Workshop@IJCAI-2022

[A Framework for Categorising AI Evaluation Instruments](#)

Anthony G Cohn, José Hernández-Orallo, Julius Sechang Mboli, Yael Moros-Daval, Zhiliang Xiang, **Lexin Zhou**
Evaluation Beyond Metrics Workshop@IJCAI-2022

[Subphenotyping of Mexican Patients With COVID-19 at Preadmission to Anticipate Severity Stratification: Age-Sex Unbiased Meta-Clustering Technique](#)

Lexin Zhou, Nekane Romero, Juan Martínez-Miranda, J Alberto Conejero, Juan M García-Gómez, Carlos Sáez
JMIR Public Health and Surveillance, 2022

1st Prize on Research Publication with the Highest Impact Factor (IF=14.56) in 2022 at ITACA Institute

Research & Industry Experience

Microsoft Research

Research Assistant (advised by Xing Xie)

Nov 2024 – Aug 2025 (Expected)

DMIP@Valencian Research Institute for AI

Research Assistant (advised by Jose Hernandez-Orallo)

Jan 2022 – May 2023 & Aug 2024 – Nov 2024

Meta AI

Red Teaming Research Consultancy

Jan 2024 – Mar 2024

Kruger AI Safety Lab

Research Intern (advised by Gabriel Recchia)

June 2023 – Sep 2023

OpenAI

Red Teaming Research Consultancy

Sep 2022 – Mar 2023

Joint Research Centre, European Commission

AI Evaluation Research Consultancy

Jul 2022 – Aug 2022

Biomedical Data Science Lab

Research Collaborator (advised by Carlos Sáez)

Jun 2020 – Dec 2021

Honors & Awards

- 2023 Open Philanthropy Long-Term Future Scholarship for master's studies at Cambridge (50000\$), Open Philanthropy
- 19'-23' Best Academic Record Awards (ranked the 1st in the BSc data science cohort every year), Uni. Politécnica de Valencia
- 2023 1st Prize for the publication with the highest impact factor in 2022, ITACA Institute – Uni. Politécnica de Valencia
- 2022 Undergraduate Research Collaboration Fellowship, Ministry of Education – Government of Spain.
- 2022 Santander Bank Studies Progress Scholarship (top 0.0001% in the university), Santander Bank

Media Coverage: My work has received extensive media coverage from Nature, Financial Times, Forbes, MIT Technology Review, IEEE Spectrum, El País, New Scientists, among others.

Professional Service

- Invited Talk: "[Larger and More Instructable Language Models Become Less Reliable](#)", Microsoft Research, 2024.
- Invited Talk: "[An LLM Feature-based Framework for Dialogue Constructiveness Assessment](#)", Toshiba Cambridge, 2024.
- Conference Reviewer: AMMAS 2023, ACL 2023.
- Conference Organising Committee: The 1st kick-off event of [Predictable AI](#) conference, Valencia, 2023.
- Newsletter: [The AI Evaluation Substack](#).
- Volunteering: Over the past five years, I taught 20+ classmates to grasp difficult concepts at both undergraduate and master's levels, provided feedback to three undergraduate students on their master's applications, visited seniors in retirement home once, made a donation of 100 Euros to Effective Altruism, and providing translation assistance for my extended family and friends who struggle with Spanish over 100+ occasions in settings such as medical appointment and business affairs.

Miscellaneous

- Language: English, Spanish, Chinese, Catalan.
- Leisure Activities: Outside of science, I spend my time playing piano, swimming, practising tennis or hiking with friends, biking on the road, traveling with loved ones, or enjoy reading about miscellaneous philosophical/science content.