YIJIE (EJ) ZHOU

yz926@cam.ac.uk ◇ ej-zhou.github.io GitHub Google Scholar

Research Interests: NLP - Interpretability, Alignment, Multilinguality, Uncertainty & Calibration

EDUCATION

University of Cambridge, PhD in Computation, Cognition and Language — Cambridge, UK
Language Technology Lab, Supervised by Anna Korhonen — Oct 2024 – Present

University of Oxford, Visiting Student — Oxford, UK
St Hilda's College, Linguistics and Computer Science — Oct 2023 – Jun 2024

Cornell University, Exchange Student & Research Assistant — Ithaca, NY, USA
College of Engineering & CommCollabTech Lab, GPA: 4.082/4.3 — Jan 2023 - May 2023

Zhejiang University, Bachelor *(Hons)* of Computer Science — Hangzhou, China
Chu Kochen Honors College, Overall GPA: 3.99/4, Major GPA: 4.0/4.0, Top 2% — Sept 2020 - Jun 2024

EXPERIENCE

Alibaba Qwen, Research Intern — Hangzhou, China
Qwen Multilingual Team, Advisor: Baosong Yang — Jun 2026 – Present

Yale University, Research Assistant — New Haven, CT, USA
Yale NLP Lab, Advisor: Arman Cohan — Apr 2023 - Oct 2023

Westlake University, Research Intern — Hangzhou, China
WestlakeNLP Lab, Advisor: Yue Zhang & Yulong Chen — Sept 2022 - Feb 2023

SELECTED PUBLICATION

[1] **Ej Zhou**, Lucas Resck, Zheng Hui, Anna Korhonen: *LLM Evaluators are Biased towards Languages*, in ICLR 2026 ICBINB & in submission to COLM 2026.

[2] **Ej Zhou**, Suchir Salhan, Catherine Arnett, Anna Korhonen: *Cross-Lingual Alignment Without Joint Training: Do Monolingual Language Models Converge on Universal Representations?*, in ICLR 2026 UCRL & in submission to EMNLP 2026.

[3] **Ej Zhou**, Caiqi Zhang, Tiancheng Hu, Chengzu Li, Nigel Collier, Ivan Vulić, Anna Korhonen: *Beyond the Final Layer: Intermediate Representations for Better Multilingual Calibration in Large Language Models* [paper], **spotlight talk** in COLM 2025 MELT & in NeurIPS 2025 MechInterp, in submission to EMNLP 2026.

[4] Zheng Hui, Sanhanat Sivapiromrat, **Ej Zhou**, Arnau Marin-Llobet, Nigel Collier: *Rethinking Intent Concealment: When Jailbreak Wrappers Backfire on Aligned LLMs*, in submission to EMNLP 2026.

[5] **Ej Zhou***, Zijie Zheng*: *What If Chinese Were Latinized? A Counterfactual Study of Script, Tokenization, and Language Modeling*, in ICML 2026 Culture x AI & in submission to EMNLP 2026.

[6] **Ej Zhou***, Songbo Hu*, Yinhong Liu*, Evgeniia Razumovskaia, Xiaobin Wang, Alexander Fraser, Ivan Vulić, and Anna Korhonen: *Dial* HealthDial *for Advice: A Multilingual and Multi-Parallel Spoken Dialogue Dataset for Knowledge-Grounded Information Seeking* [paper] [code] [video], in ACL 2026 Findings. * denotes equal contribution.

[7] Jiaqi Weng, Han Zheng, Hanyu Zhang, **Ej Zhou**, Qinqin He, Jialing Tao, Hui Xue, Zhixuan Chu, Xiting Wang: *Safe-SAIL: Towards a Fine-grained Safety Landscape of Large Language Models via Sparse Autoencoder Interpretation Framework* [paper], in ACL 2026 Findings.

[8] Pedro Ortiz Suarez **et al.** (incl. **Ej Zhou**): *CommonLID: Re-evaluating State-of-the-Art Language Identification Performance on Web Data* [paper] [blog], contributing author, in ACL 2026.

[9] **Ej Zhou**, Weiming Lu: *Bias Beyond English: Evaluating Social Bias and Debiasing Methods in a Low-Resource Setting* [paper], in ACL 2025 GeBNLP & NLPCC 2025.

[10] Giulia Occhini, Kumiko Tanaka-Ishii, Anna Barford, Refael Tikochinski, Songbo Hu, Roi Reichart, **Yijie Zhou**, Hannah Clause, Ulla Petti, Ivan Vulić, Ramit Debnath, Anna Korhonen: *Artificial Intelligence is Creating a New Global Linguistic Hierarchy* [paper] [article], in submission to Nature Communications.

[11] Rohan Phanse, **Yijie Zhou**, Kejian Shi, Wencai Zhang, Yixin Liu, Yilun Zhao, Arman Cohan: *MSRS: Evaluating Multi-Source Retrieval-Augmented Generation* [paper] [code], in COLM 2025.

[12] Yulong Chen, Huajian Zhang, **Ej Zhou**, Xuefeng Bai, Yueguan Wang, Ming Zhong, Jianhao Yan, Yafu Li, Judy Li, Michael Zhu and Yue Zhang: *Revisiting Cross-Lingual Summarization: A Corpus-based Study and A New Benchmark with Improved Annotation* [paper] [code], in ACL 2023.

SERVICE & EVENTS

Teaching - Supervisor & Lecturer, Li 18 Computational Linguistics 2025-2026, Cambridge

Conferences — EACL 2026, EMNLP 2025, ACL 2025, NLPCC 2025, ACL 2024, EACL 2024, EMNLP 2023, ACL 2023

Volunteering - EAG London 2026, EACL 2024

Committee Member - Cambridge World Cinema Society, Cambridge Language and Culture Society

Events — Cambridge Language Sciences Annual Symposium (Posters*2), SparkLab S1 (Mentor), OxGen 2023 Summit & OxGen AI Hackathon 2023 (Special Award), Polyglot Conference 2023, Language Event 2023

SELECTED AWARDS

- Cambridge Trust Scholarship (Fully Funded) 2024-2028
- Excellence Honors Scholarship, Zhejiang University *(Top 1% in CKC Honors College)* 2021
- Hengyi Scholarship *(5/˜24,000 in Zhejiang University)* 2021

LANGUAGES

- I do multilingual research because I love languages. I speak to varying degrees: *Mandarin*, *English*, *Wu*, *French*, *Russian*, *Japanese*, *German*; I'm learning *Polish* and *Portuguese*.
- I have a 700+ day streak on Duolingo.