

# Yong Zheng-Xin

---

CONTACT INFORMATION	personal website: <a href="https://yongzx.github.io">yongzx.github.io</a> email: <a href="mailto:contact.yong@brown.edu">contact.yong@brown.edu</a>	
EDUCATION	<b>Brown University</b> , Providence, RI Ph.D. Student, Computer Science Advisor: Prof. Stephen H. Bach 07/2021 - Present	
	<b>Minerva University</b> , San Francisco, CA B.Sc., Computer Science (Major) and Business (Minor) Advisor: Prof. Patrick D Watson Major GPA: 4.0/4, Cumulative GPA: 4.0/4 09/2017 - 05/2021	
RELEVANT EMPLOYMENT	<b>Brown University</b> , Ph.D. Student <i>Supervisor: Stephen Bach</i> 07/2021 - Present	
	<b>Meta AI</b> , Research Scientist Intern <i>FAIR, Fundamental AI Research</i> 06/2024 - Present <ul style="list-style-type: none"><li>Responsible AI research for multilingual speech model MMS.</li><li><i>Supervisor: Jean Maillard</i></li></ul>	
	<i>GenAI Trust Team (Research Collaborator)</i> <ul style="list-style-type: none"><li>Multilingual safety research.</li><li><i>Point-of-Contact: Jianfeng Chi</i></li></ul>	
	<b>Cohere For AI</b> , Research Collaborator <i>Aya Responsible Release</i> 05/2023 - 02/2024 <ul style="list-style-type: none"><li>Safety red-teaming for multilingual LLM Aya-101.</li><li>Deliverable: <b>Aya Model (co-first author)</b></li><li><i>Supervisors: Sara Hooker &amp; Julia Kreutzer</i></li></ul>	
	<b>BigScience</b> , Research Collaborator/Lead <i>Multilingual Modeling Group</i> 08/2021 - 12/2022 <ul style="list-style-type: none"><li>Led language adaptation research of BLOOM to low-resource languages.</li><li>Helped with data collection for the earliest instruction-following models, namely To (English) and BLOOMZ (multilingual).</li><li>Deliverables: <b>BLOOM+1 (Research Lead), To, PromptSource, BLOOMZ, BLOOM</b></li><li><i>Supervisor: Vassilina Nikoulina</i></li></ul>	
	<b>FrameNet Project</b> , Research Intern <i>Google Summer of Code 2019 and 2020</i> 06/2019 - 12/2020 <ul style="list-style-type: none"><li>Expanded FrameNet through semi-supervised learning and anomaly detection.</li><li>Investigated cross-lingual alignment of semantic frames in FrameNet graph.</li><li>Deliverables: <b>SDEC-AD, Frame Shift Prediction</b></li><li><i>Supervisors: Collin F. Baker, Tiago T. Torrent, Oliver Czulo</i></li></ul>	
AWARDS	<b>Best Paper Award</b> at ACL 2024 2024 <b>Best Paper Award</b> at NeurIPS 2023 Socially Responsible Language Modeling (SoLaR) Workshop 2023 <b>3rd Best App Overall</b> at FirstNet Public Safety Hackathon 2018 <b>Best Use of ESRI</b> at FirstNet Public Safety Hackathon 2018	

<b>Grand Prize Winner</b> at #100Hacks Hackathon for Puerto Rico	2017
<b>Grand Prize Winner</b> at NCSV Innovate NUS Hackathon	2017
<b>Open Category Grand Prize</b> at SIA Airlines AppChallenge	2017
<b>Global Finalist</b> at Google Science Fair	2016
<b>Outstanding A-Level Cambridge Learner Award</b> for Perfect Score in Maths	2016

FEATURED  
PUBLICATIONS  
(\* INDICATES  
CO-FIRST  
AUTHORSHIP)

### Multilingual AI Safety

- [1] Preference Tuning For Toxicity Mitigation Generalizes Across Languages  
Xiaochen Li\*, **Zheng-Xin Yong\***, Stephen H. Bach  
*In submission.*
- [2] Aya Model: An Instruction Finetuned Open-Access Multilingual Language Model  
Ahmet Üstün\*, Viraat Aryabumi\*, **Zheng-Xin Yong\***, Wei-Yin Ko\*, Daniel D'souza\*,  
Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, Fred-  
die Vargus, Phil Blunsom, Shayne Longpre, Niklas Muennighoff, Marzieh Fadaee,  
Julia Kreutzer, Sara Hooker  
*ACL 2024. (Role in project: Multilingual safety red-teaming)*  
**Best Paper Award.** Work also featured in [The New York Times](#) and other media.
- [3] Low-Resource Languages Jailbreak GPT-4  
**Zheng-Xin Yong**, Cristina Menghini, Stephen H. Bach  
*NeurIPS 2023 Socially Responsible Language Modeling Workshop*  
**Best Paper Award.** Work also featured in [New Scientist](#) and other media.

### Low-Resource NLP and Multilingual Synthetic Data Generation

- [4] LexC-Gen: Generating Data for Extremely Low-Resource Languages with Large  
Language Models and Bilingual Lexicons  
**Zheng-Xin Yong**, Cristina Menghini, Stephen H. Bach  
*In submission.*
- [5] Prompting Multilingual Large Language Models to Generate Code-Mixed Texts:  
The Case of South East Asian Languages  
**Zheng-Xin Yong**, Ruochen Zhang, Jessica Zosa Forde, Skyler Wang, Arjun Subramo-  
nian, Holy Lovenia, Samuel Cahyawijaya, Genta Indra Winata, Lintang Sutawika,  
Jan Christian Blaise Cruz, Yin Lin Tan, Long Phan, Rowena Garcia, Tamar Solorio,  
Alham Fikri Aji  
*EMNLP 2023 CALCS Workshop.*  
**Work also featured on WIRED.**
- [6] BLOOM+1: Adding Language Support to BLOOM for Zero-Shot Prompting  
**Zheng-Xin Yong**, Hailey Schoelkopf, Niklas Muennighoff, Alham Fikri Aji, David  
Ifeoluwa Adelani, Khalid Almubarak, M Saiful Bari, Lintang Sutawika, Jungo Ka-  
sai, Ahmed Baruwa, Genta Indra Winata, Stella Biderman, Edward Raff, Dragomir  
Radev, Vassilina Nikoulina  
*ACL 2023.*

OTHER  
PUBLICATIONS

- [7] CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark  
David Romero, ..., **Zheng-Xin Yong**, ..., Tamar Solorio, Alham Fikri Aji (75 authors)  
*In submission*
- [8] SEACrowd: A Multilingual Multimodal Data Hub and Benchmark Suite for South-  
east Asian Languages  
Holy Lovenia, ..., **Zheng-Xin Yong**, Samuel Cahyawijaya (61 authors)  
*In submission*

- [9] A Safe Harbor for AI Evaluation and Red Teaming  
Shayne Longpre, ..., **Zheng-Xin Yong**, ..., Percy Liang, Peter Henderson (23 authors)  
*ICML 2024 Oral (Position Paper)*.
- [10] Representativeness as a Forgotten Lesson for Multilingual and Code-switched Data Collection and Preparation  
A. Seza Dođruöz, Sunayana Sitaram, **Zheng-Xin Yong**  
*EMNLP Findings 2023*.
- [11] Crosslingual Generalization through Multitask Finetuning  
Niklas Muennighoff, ..., **Zheng-Xin Yong**, ..., Edward Raff, Colin Raffel (19 authors)  
*ACL 2023*.
- [12] PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts  
Stephen H. Bach\*, Victor Sanh\*, **Zheng-Xin Yong**, ..., Alexander M. Rush (27 authors)  
*ACL Demo 2023*.
- [13] The Decades Progress on Code-Switching Research in NLP: A Systematic Survey on Trends and Challenges  
Genta Indra Winata, Alham Fikri Aji, **Zheng-Xin Yong**, Tamar Solorio  
*ACL Findings 2023*.
- [14] BLOOM: A 176B-Parameter Open-Access Multilingual Language Model  
BigScience Workshop (including **Zheng-Xin Yong** and 300+ authors)  
*Preprint 2023*
- [15] Multitask Prompted Training Enables Zero-Shot Task Generalization  
Victor Sanh\*, ..., **Zheng-Xin Yong**, ..., Alexander M. Rush (41 authors)  
*ICLR 2022 Spotlight*.

SELECTED TALKS

- The Alan Turing Institute, London Data Week: Multilingual AI Safety**, Jul 2024.
- Cohere For AI, C4AI Gatherings: Aya Multilingual Safety**, Feb 2024.
- Cohere For AI, Aya Grande Finale: Aya Responsible Release**, Feb 2024.
- Cohere For AI, Closing the Contribution Chapter: Malay Ambassador**, Dec 2023.

SELECTED PRESS & MEDIA

- GPT-4 gave advice on planning terrorist attacks when asked in Zulu** (New Scientist, 2023)
- ChatGPT Is Cutting Non-English Languages Out of the AI Revolution** (WIRED, 2023)

SERVICES

- Area Chair:**  
EMNLP 2023 (Multilingualism and Linguistic Diversity)
- Conference/Workshop Reviewer:**  
COLM 2024  
ARR 2021-Present  
ACL 2021-2023
- Outreach and Mentorship Service Programs:**  
Deep Learning Indaba Mentorship Service (2022 - 2023)  
Brown Computer Science exploreCSR (2022)  
Minerva-Masason Mentoring Program (2019 - 2021)