

Mike Zhang

Curriculum Vitae · Copenhagen, Denmark

✉ mike.zhang@di.ku.dk · 🎓 [Google Scholar](https://scholar.google.com/citations?user=jjzha) · 🌐 jjzha.github.io · 🖥️ github.com/jjzha · 😊 hf.co/jjzha

EMPLOYMENT	University of Copenhagen <i>Postdoctoral Researcher in Natural Language Processing</i> with Desmond Elliott & Ali Basirat	Dec. 2025 – Present Copenhagen, DK
	Aalborg University <i>Postdoctoral Researcher in Natural Language Processing</i> with Johannes Bjerva & Euan D. Lindsay	Mar. 2024 – Nov. 2025 Copenhagen, DK
	NEC Laboratories Europe <i>Ph.D. Research Intern</i>	Sep. 2022 – Dec. 2022 Heidelberg, DE
	Dataprovider.com <i>Data Engineer</i>	Aug. 2019 – Aug. 2020 Groningen, NL
MOBILITY	Swiss Federal Institute of Technology Lausanne <i>Ph.D. Research Visitor</i> ; with Syrielle Montariol & Antoine Bosselut	Oct. 2023 – Nov. 2023 Lausanne, CH
	National University of Singapore <i>Ph.D. Research Visitor</i> ; with Min-Yen Kan	Feb. 2023 – Jul. 2023 Singapore, SG
EDUCATION	<i>Ph.D. Computer Science (obtained on 7 March, 2024)</i> IT University of Copenhagen Pioneer Center for Artificial Intelligence (Affiliated Researcher) Ludwig Maximilian University of Munich (Affiliated Researcher) with Barbara Plank & Rob van der Goot Thesis: Computational Job Market Analysis with Natural Language Processing	Sep. 2020 – Mar. 2024 Copenhagen, DK Copenhagen, DK Munich, DE
	<i>M.Sc. Information Science (Cum Laude)</i>	Sep. 2018 – Jan. 2020
	<i>B.Sc. Information Science</i> University of Groningen	Sep. 2015 – Jul. 2018 Groningen, NL
HONORS/ AWARDS/ FUNDING	A9. NVIDIA Academic Grant Program // Co-applicant Awarded 32,000 GPU Core Hours for multilingual pretraining of LLMs.	2026
	A8. Best Practice Paper Nominee Nominated for best practice paper at SEFI 2025.	2025
	A7. IT University of Copenhagen Research Award Nominee Nominated for a Research Award for Snakmodel, an open-source Danish LLM.	2025
	A6. Aalborg University // Co-applicant Awarded 91,930 DKK (12,500 EUR) to investigate and implement local chatbots.	2024
	A5. Otto Mønstedts Fonden // Main applicant Awarded 7,500 DKK (1,000 EUR) from OMF to attend ACL 2024.	2024
	A4. Aalborg University // Co-applicant Awarded 200,000 DKK (27,000 EUR) to train RAG models.	2024
	A3. Danish e-infrastructure Consortium Computing // Main applicant Awarded 75,000 GPU Core Hours to train LLMs in the education domain.	2024

	A2. IT University of Copenhagen Mobility // Main applicant Awarded 60,000 DKK (3,500 EUR) for a research stay abroad.	2022
	A1. Outstanding Paper Award At the ML Evaluation Standards Workshop at ICLR 2022.	2022
INVITED TALKS	T12. CLASP Seminar	May 2026, Gothenburg, SE
	T11. DIKU Bits	Mar. 2026, Copenhagen, DK
	T10. CST Workshop	Nov. 2025, Copenhagen, DK
	T9. University of Southern Denmark	Jul. 2025, Odense, DK
	T8. AI and Labor Market (AI&LM) Workshop	Apr. 2025, Ghent, BE
	T7. University of Amsterdam	Dec. 2024, Amsterdam, NL
	T6. AI for the People Conference	Nov. 2024, Aalborg, DK
	T5. AAU NLP Symposium	Nov. 2024, Copenhagen, DK
	T4. AI x Education Network	Oct. 2024, Aalborg, DK
	T3. NLP Workshop on Linguistic Variation	Oct. 2023, Copenhagen, DK
	T2. WING Group	Feb. 2023, Singapore, SG
	T1. GroNLP Group	Oct. 2022, Groningen, NL
TEACHING	Natural Language Processing (DAKI5) , Guest Lecturer	2025
	Machine Intelligence (DSNSWCB513) , Guest Lecturer	2024
	Semester Project (DAT9) , Supervisor	2024, 2025
	Advanced Natural Language Processing (KSANLPD1KU) , Guest Lecturer	2024
	Seminars in Data Science (KSEEDAS1KU) , Guest Lecturer	2022, 2023
	Intro to NLP and Deep Learning (BSSEYEP1KU) , Senior TA, Lecturer	2021, 2022
	Communicating Research to a Broader Audience (Ph.D. Course) , Co-Organizer	2021
SUPERVISION	Ph.D. Students	
	R. Aavang; Information Extraction (<i>Co-supervisor</i>)	2027
	N. Musazade; Entity Linking (<i>Ph.D. visitor</i>)	2025
	Master Students	
	E. Norvin, N. Reece, R. Birkmose; Efficient Methods and Optimization	2025
	S. Pasham, P. das Neves Rodrigues Mateus Cristóvão; Entity Linking	2022
	Bachelor Students	
	T. Nguyen; Information Extraction	2023
	A.S. Barwig; Open Information Extraction	2022
	P. das Neves Rodrigues Mateus Cristóvão; Entity Linking	2021
ACADEMIC SERVICE	Grant Reviewer: AAPG General Call – ANR ($\pm\text{€}600\text{K}$)	2025, 2026
	Journal Reviewer: PNAS	2026
	Conference Organizational Roles: EACL SRW, NoDaLiDa	2024, 2025
	Conference Area Chair: ACL, NAACL	2025–
	Conference Reviewer: ARR, ICLR (2025), NeurIPS (2025)	2021–
PRO BONO	Lightcast; Consulting on Computational Job Market Analysis	2025
	IOM United Nations; Consulting on Computational Job Market Analysis	2025
	IDIAP; Consulting on Computational Job Market Analysis	2024

Full list of Peer-reviewed Publications (🏆 = Equal Contributions)

I co-authored 40 peer-reviewed publications: 1 journal article, 28 conference papers, and 11 workshop papers. According to Google Scholar per April 29, 2026, my h-index is 18 and I accrued a total of 1,791 citations.

Journal Articles

- J1. [A benchmark of expert-level academic questions to assess AI capabilities](#)
Center for AI Safety, Long Phan, Alice Gatti, Ziwen Han, Nathaniel Li, Josephina Hu, Hugh Zhang, Chen Bo Calvin Zhang, Mohamed Shaaban, John Ling, Sean Shi, Michael Choi, Anish Agrawal, ..., **Mike Zhang**, et al. (2026)
Nature.

Conference Proceedings

- C28. [Effective Performance Measurement: Challenges and Opportunities in KPI Extraction from Earnings Calls](#)
Rasmus T. Aavang, Rasmus Tjalk-Bøggild, Alexandre Iolov, Giovanni Rizzi, **Mike Zhang**, Johannes Bjerva (2026)
Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics: Industry Track.
- C27. [CommonLID: Re-evaluating State-of-the-Art Language Identification Performance on Web Data](#)
Pedro Ortiz Suarez, Laurie Burchell, Catherine Arnett, ..., **Mike Zhang**, et al. (2026)
Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics.
- C26. [Follow the Path: Reasoning over Knowledge Graph Paths to Improve LLM Factuality](#)
Mike Zhang, Johannes Bjerva, Russa Biswas (2026)
Findings of the 64th Annual Meeting of the Association for Computational Linguistics.
- C25. [UniSkill: A Dataset for Matching University Curricula to Professional Competencies](#)
Nurlan Musazade, József Mezei, **Mike Zhang** (2026)
Proceedings of the 15th edition of the Language Resources and Evaluation Conference.
- C24. [HIFI-KPI: A Dataset for Hierarchical KPI Extraction from Earnings Filings](#)
Rasmus Aavang, Giovanni Rizzi, Rasmus Bøggild, Alexandre Iolov, **Mike Zhang**, Johannes Bjerva (2026)
Proceedings of the 15th edition of the Language Resources and Evaluation Conference.
- C23. [SEFL: A Framework for Generating Synthetic Educational Assignment Feedback with LLM Agents](#)
Mike Zhang, Amalie Pernille Dilling, Léon Gondelman, Niels Erik Ruan Lyngdorf, Euan D. Lindsay, Johannes Bjerva (2026)
Proceedings of the 15th edition of the Language Resources and Evaluation Conference.
- C22. [Kaleidoscope: In-language Exams for Massively Multilingual Vision Evaluation](#)
Israfel Salazar, Manuel Fernández Burda, Shayekh Bin Islam, Arshia Soltani Moakhar, Shivalika Singh, Fabian Farestam, Angelika Romanou, Danylo Boiko, Dipika Khullar, **Mike Zhang**, et al. (2026)
In the Fourteenth International Conference on Learning Representations.
- C21. [Cross-Lingual Sentence-Level Skill Identification in English and Danish Job Advertisements](#)
Nurlan Musazade, **Mike Zhang**, József Mezei (2025)
The 8th International Conference on Natural Language and Speech Processing.
- C20. [Scaling Course Evaluations with Large Language Models: Semester-level Digestible Student Feedback for Program Leaders](#)
🏆 **Best Practice Paper nominee**
Mike Zhang, Euan D. Lindsay, Maj-Britt Quitzau, Johannes Bjerva (2025)
The 53rd Annual Conference of the European Society for Engineering Education.
- C19. [All Languages Matter: Evaluating LLMs on Culturally Diverse 100 Languages](#)
🏆 **Highlight 13.5%**

Ashmal Vayani, Dinura Dissanayake, Hasindri Watawana, Noor Ahsan, Nevasini Sasikumar, Omkar Thawakar, Henok Biadgign Ademtew, Yahya Hmaiti, Amandeep Kumar, Kartik Kuckreja, Mykola Maslych, Wafa Al Ghallabi, Mihail Mihaylov, Chao Qin, Abdelrahman M Shaker, **Mike Zhang**, et al. (2025)

The IEEE/CVF Conference on Computer Vision and Pattern Recognition.

C18. [The Responsible Development of Automated Student Feedback with Generative AI](#)

Euan Lindsay, **Mike Zhang**, Aditya Johri, Johannes Bjerva (2025)

IEEE Global Engineering Education Conference.

C17. [SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models](#)

Margaret Mitchell, Giuseppe Attanasio, Ioana Baldini, Miruna Clinciu, Jordan Clive, Pieter Delobelle, Manan Dey, Sil Hamilton, Timm Dill, Jad Doughman, ..., **Mike Zhang**, et al. (2025)

In Proceedings of The 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics

C16. [INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge](#)

 **Spotlight 1.8%**

Angelika Romanou, Negar Foroutan, Anna Sotnikova, Zeming Chen, Sree Harsha Nelaturu, Shivalika Singh, Rishabh Maheshwary, Micol Altomare, Mohamed A. Haggag, ..., **Mike Zhang**, et al. (2025)

In Proceedings of the Thirteenth International Conference on Learning Representations

C15. [How Do Hackathons Foster Creativity? Towards Automated Evaluation of Creativity at Scale](#)

Jeanette Falk, Yiyi Chen, Janet Rafner, **Mike Zhang**, Johannes Bjerva, Alexander Nolte (2025)

In Proceedings of The 2025 Conference on Human Factors in Computing Systems.

C14. [Snakmodel: Lessons Learned from Training an Open Danish Large Language Model](#)

 **ITU Research Award nominee**

Mike Zhang , Max Müller-Eberstein , Elisa Bassignana, Rob van der Goot (2025)

The Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies.

C13. [MORSED: Morphological Segmentation of Danish and its Effect on Language Modeling](#)

Rob van der Goot, Anette Jensen, Emil Allerslev Schledermann, Mikkel Wildner Kildeberg, Nicolaj Larsen, **Mike Zhang**, Elisa Bassignana (2025)

The Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies.

C12. [Leveraging Large Language Models for Actionable Course Evaluation Feedback](#)

Mike Zhang, Euan Lindsay, Frederik Thorbensen, Danny Bøgsted Poulsen, Johannes Bjerva (2024)

The 52nd Annual Conference of the European Society for Engineering Education.

C11. [Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning](#)

Shivalika Singh, Freddie Vargus, Daniel Dsouza, Börje F. Karlsson, Abinaya Mahendiran, Wei-Yin Ko, Herumb Shandilya, Jay Patel, Deividas Mataciunas, Laura OMahony, **Mike Zhang**, et al. (2024)

In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics.

C10. [NNOSE: Nearest Neighbor Occupational Skill Extraction](#)

Mike Zhang, Rob van der Goot, Min-Yen Kan, Barbara Plank (2024)







In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics.

C9. [Entity Linking in the Job Market Domain](#)

Mike Zhang, Rob van der Goot, Barbara Plank (2024)

In Findings of the Association for Computational Linguistics: EACL 2024.

C8. [Can Humans Identify Domains?](#)

Maria Barrett , Max Müller-Eberstein , Elisa Bassignana , Amalie Brogaard Pauli , **Mike Zhang** , Rob van der Goot  (2024)

In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation.

- C7. [ESCOXLM-R: Multilingual Taxonomy-driven Pre-training for the Job Market Domain](#)
Mike Zhang, Rob van der Goot, Barbara Plank (2023)
In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics.
- C6. [Evidence > Intuition: Transferability Estimation for Encoder Selection](#)
 Elisa Bassignana , Max Müller-Eberstein , **Mike Zhang** , Barbara Plank (2022)
In Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing.
- C5. [Experimental Standards for Deep Learning in Natural Language Processing Research](#)
 Dennis Ulmer, Elisa Bassignana, Max Müller-Eberstein, Daniel Varab, **Mike Zhang**, Rob van der Goot, Christian Hardmeier, Barbara Plank (2022)
In Findings of the Association for Computational Linguistics: EMNLP 2022.
- C4. [SKILLSPAN: Hard and Soft Skill Extraction from Job Postings](#)
Mike Zhang , Kristian Nørgaard Jensen , Sif Dam Sonniks, Barbara Plank (2022)
In Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics.
- C3. [KOMPETENCER: Fine-grained Skill Classification in Danish Job Postings via Distant Supervision and Transfer Learning](#)
Mike Zhang , Kristian Nørgaard Jensen , Barbara Plank (2022)
In Proceedings of the 13th Edition of its Language Resources and Evaluation Conference.
- C2. [Cartography Active Learning](#)
Mike Zhang, Barbara Plank (2021)
In Findings of the Association for Computational Linguistics: EMNLP 2021.
- C1. [De-identification of Privacy-related Entities in Job Postings](#)
Mike Zhang , Kristian Nørgaard Jensen , Barbara Plank (2021)
Proceedings of the 23rd Nordic Conference on Computational Linguistics.

Anthologies and Books

- W11. [Do Large Language Models Adapt to Language Variation across Socioeconomic Status?](#)
 Elisa Bassignana, **Mike Zhang**, Dirk Hovy, Amanda Cercas Curry (2026)
In Proceedings of The Thirteenth Workshop on NLP for Similar Languages, Varieties and Dialects.
- W10. [NLPnorth @ TalentCLEF 2025: Comparing Discriminative, Contrastive, and Prompt-Based Methods for Job Title and Skill Matching](#)
Mike Zhang, Rob van der Goot (2025)
In Proceedings of the Conference and Labs of the Evaluation Forum.
- W9. [On-Device LLMs for Home Assistant: Dual Role in Intent Detection and Response Generation](#)
 Rune Birkmose, Nathan Mørkeberg Reece, Esben Hofstedt Norvin, Johannes Bjerva, **Mike Zhang** (2025)
In Proceedings of The 10th Workshop on Noisy and User-generated Text.
- W8. [DAKULTUR: Evaluating the Cultural Awareness of Language Models for Danish with Native Speakers](#)
 Max Müller-Eberstein , **Mike Zhang** , Elisa Bassignana, Peter Trolle, Rob van der Goot (2025)
In Proceedings of The 3rd Workshop on Cross-Cultural Considerations in NLP.
- W7. [Rethinking Skill Extraction in the Job Market Domain using Large Language Models](#)
 Khanh Cao Nguyen, **Mike Zhang**, Syrielle Montariol, Antoine Bosselut (2024)
In Proceedings of the 1st Workshop on Natural Language Processing for Human Resources.
- W6. [JobSkape: A Framework for Generating Synthetic Job Postings to Enhance Skill Matching](#)
 Antoine Magron, Anna Dai, **Mike Zhang**, Syrielle Montariol, Antoine Bosselut (2024)
In Proceedings of the 1st Workshop on Natural Language Processing for Human Resources.

- W5. [Deep Learning-based Computational Job Market Analysis: A Survey on Skill Extraction and Classification from Job Postings](#)
Elena Senger, **Mike Zhang**, Rob van der Goot, Barbara Plank (2024)
In Proceedings of the 1st Workshop on Natural Language Processing for Human Resources.
- W4. [Skill Extraction from Job Postings using Weak Supervision](#)
Mike Zhang, Kristian Nørgaard Jensen, Rob van der Goot, Barbara Plank (2022)
RecSys in HR'22: The 2nd Workshop on Recommender Systems for Human Resources, in conjunction with the 16th ACM Conference on Recommender Systems.
- W3. [A Common Set of Experimental Standards for Deep Learning Research: A Natural Language Processing Perspective](#)
 **Best Paper Award**
Dennis Ulmer, Elisa Bassignana, Max Müller-Eberstein, Daniel Varab, **Mike Zhang**, Christian Hardmeier, Barbara Plank (2022)
Machine Learning Evaluation Standards workshop at ICLR 2022. (Non-Archival)
- W2. [The Effect of Translationese in Machine Translation Test Sets](#)
Mike Zhang, Antonio Toral (2019)
Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers).
- W1. [Grunn2019 at SemEval-2019 Task 5: Shared Task on Multilingual Detection of Hate](#)
Mike Zhang , Roy David , Leon Graumans , Gerben Timmerman  (2019)
Proceedings of the 13th International Workshop on Semantic Evaluation.

Pre-prints

- P2. [WorkRB: A Community-Driven Evaluation Framework for AI in the Work Domain](#)
Matthias De Lange, Warre Veys, Federico Retyk, Daniel Deniz, Warren Jouanneau, **Mike Zhang**, Aleksander Bielinski, Emma Jouffroy, Nicole Clobes, Nina Baranowska, David Graus, Marc Palyart, Rabih Zbib, Dimitra Gkatzia, Thomas Demeester, Tjil De Bie, Toine Bogers, Jens-Joris Decorte, Jeroen Van Hautte (2026)
ArXiv Pre-print.
- P1. [Sailor2: sailing in south-east Asia with inclusive multilingual LLMs](#)
Longxu Dou, Qian Liu, Fan Zhou, Changyu Chen, Zili Wang, Ziqi Jin, Zichen Liu, Tongyao Zhu, Cunxiao Du, Penghui Yang, Haonan Wang, Jiaheng Liu, Yongchi Zhao, Xiachong Feng, Xin Mao, Man Tsung Yeung, Kunat Pipatanakul, Fajri Koto, Min Si Thu, Hynek Kydlíček, Zeyi Liu, Qunshu Lin, Sittipong Sripaisarnmongkol, Kritaphad Sae-Khow, Nirattisai Thongchim, Taechawat Konkaew, Narong Borijindagoon, Anh Dao, Matichon Maneegard, Phakphum Artkaew, Zheng-Xin Yong, Quan Nguyen, Wannaphong Phatthiyaphaibun, Hoang H. Tran, **Mike Zhang**, Shiqi Chen, Tianyu Pang, Chao Du, Xinyi Wan, Wei Lu, Min Lin (2025)
ArXiv Pre-print.