Mike Zhang

Postdoctoral Researcher @ University of Copenhagen
Pioneer Centre for Artificial Intelligence
Copenhagen, Denmark
jjz@cs.aau.dk · jjzha.github.io · github.com/jjzha · hf.co/jjzha

EMPLOYMENT

Postdoctoral Researcher — Dec. 2025 – Present
▶ University of Copenhagen — Copenhagen, DK
Advisors: Desmond Elliott, Ali Basirat
Topic: Post-training for Low-resource Domains

Postdoctoral Researcher — Mar. 2024 – Nov. 2025
▶ Aalborg University — Copenhagen, DK
Advisors: Johannes Bjerva, Euan Lindsay
Topic: Natural Language Processing and Education

Ph.D. Research Intern — Sep. 2022 – Dec. 2022
▶ NEC Laboratories Europe — Heidelberg, DE
Advisors: Kiril Gashteovski, Goran Glavaš
Topic: Open Information Extraction

Data Engineer — Aug. 2019 – Aug. 2020
▶ Dataprovider.com — Groningen, NL

MOBILITY

Ph.D. Research Visitor — Oct. 2023 – Nov. 2023
▶ Swiss Federal Institute of Technology Lausanne (EPFL) — Lausanne, CH
Advisors: Syrielle Montariol, Antoine Bosselut

Ph.D. Research Visitor — Feb. 2023 – Jul. 2023
▶ National University of Singapore (NUS) — Singapore, SG
Advisor: Min-Yen Kan

EDUCATION

Ph.D. Computer Science — Mar. 2024
▶ IT University of Copenhagen — Copenhagen, DK
▶ Pioneer Center for Artificial Intelligence — Copenhagen, DK
▶ Ludwig Maximilian University of Munich — Munich, DE
Advisors: Barbara Plank, Rob van der Goot
Thesis: Computational Job Market Analysis with Natural Language Processing;
(**defended 7 March, 2024**).

M.Sc. Information Science — Jan. 2020
▶ University of Groningen — Groningen, NL
Thesis Advisor: Antonio Toral Ruiz

B.Sc. Information Science — Jul. 2018
▶ University of Groningen — Groningen, NL
Thesis Advisor: Barbara Plank

FUNDING

F5. **Aalborg University** // Co-applicant   2024
Awarded **91,930 DKK (12,500 EUR)** from Aalborg University to investigate and implement in-house chatbots.

F4. **Aalborg University** // Co-applicant   2024
Awarded **200,000 DKK (27,000 EUR)** from Aalborg University to investigate retrieval-augmented generation models for a local learning environment chat assistant.

F3. **Otto Mønsteds Fonden** // Main applicant   2024
Awarded **7,500 DKK (1,000 EUR)** from OMF to attend ACL 2024.

F2. **Danish e-infrastructure Consortium Computing** // Main applicant   2024
Awarded **75,000 GPU Core Hours** from the board of the DeiC e-resource committee to investigate and train large language model for abundant feedback in the education domain.

F1. **IT University of Copenhagen Mobility** // Main applicant   2022
Awarded **60,000 DKK (3,500 EUR)** for a research visit abroad in an external institution: National University of Singapore.

AWARDS

A3. **Best Practice Paper Nominee**   2025
Nominated (top-3) for best practice paper at SEFI 2025 for "Scaling Course Evaluations with Large Language Models: Semester-level Digestible Student Feedback for Program Leaders".

A2. **IT University of Copenhagen Research Award Nominee**   2025
Nominated (top-4) for the ITU Research Award for Snakmodel, an open-source Danish large language model.

A1. **Outstanding Paper Award**   2022
Awarded for an outstanding paper at the ML Evaluation Standards Workshop at ICLR 2022.

INVITED TALKS

T10.	CST Workshop, Copenhagen, DK	November 2025
T9.	University of Southern Denmark, Odense, DK	July 2025
T8.	AI and Labor Market (AI&LM) Workshop, Ghent, BE	April 2025
T7.	University of Amsterdam, Amsterdam, NL	December 2024
T6.	AI for the People Conference, Aalborg, DK	November 2024
T5.	AAU NLP Symposium, Copenhagen, DK	November 2024
T4.	AI x Education Network, Aalborg, DK	October 2024
T3.	NLP Workshop on Linguistic Variation, Copenhagen, DK	October 2023
T2.	WING Group, Singapore, SG	February 2023
T1.	GroNLP Group, Groningen, NL	October 2022

TEACHING	——— **Aalborg University Copenhagen** ———	

Certification:
- PBL Course   2024

Courses:
- *Machine Intelligence (DSNSWCB513)*, Guest Lecturer   2024
- *Machine Intelligence (DSNSWCB513)*, Guest Lecturer   2024
- *Semester Project (DAT9)*, Supervisor   2024-2025

——— **IT University of Copenhagen** ———

Certification:
- Introductory Teacher Development   2022

Courses:
- *Advanced Natural Language Processing (KSANLPD1KU)*, Guest Lecturer   2024
- *Seminars in Data Science (KSSEDAS1KU)*, Guest Lecturer   2022, 2023
- *Introduction to Natural Language Processing and Deep Learning (BSSEYEP1KU)*, Senior Teaching Assistant, Lecturer   2021, 2022
- *Communicating SOTA NLP Research to a Broader Audience (Ph.D. Course)*, Co-Organizer   2021

——— **University of Groningen** ———

Courses:
- *Machine Learning (SOMINDW07)*, Head Teaching Assistant   2019, 2020
- *Learning from Data (LIX016M05)*, Head Teaching Assistant   2019
- *Social Media (LIX017B05)*, Teaching Assistant   2019

SUPERVISION

Ph.D. Students
- *Co-supervision*, R. Aavang   2025–
  "Guarantees of Factuality in LLM-based Extraction of Financial KPIs from Earnings Transcripts"

Master Students
- *M.Sc. Thesis (DAT10)*, E. Norvin, N. Reece, R. Birkmose   2025
  "Optimization Methods for Large Language Models"
- *Project Supervision (DAT9)*, E. Norvin, N. Reece, R. Birkmose   2024
  "Efficient Home Assistants"
- *M.Sc. Thesis*, S. Pasham and P. das Neves Rodrigues Mateus Cristóvão   2022
  "Matching University Curricula to Occupations"
- *Project Supervision (SD)*, P. das Neves Rodrigues Mateus Cristóvão   2021
  "Converting Job Requirements into Skills"

Bachelor Students
- *B.Sc. Thesis*, T. Nguyen   2023
  "Skill Extraction for Vietnamese Job Market Understanding"
- *B.Sc. Thesis*, A.S. Barwig   2022
  "Unsupervised Skill Extraction for Job Market Understanding"

Visitors:

- *Ph.D. Research Visitor*, N. Musazade   2025
  Working on the intersection of labour market analytics and education.

ACADEMIC SERVICE & COMMISSIONS OF TRUST	**Grant Reviewer**: (2025) AAPG General Call – ANR (France) **Conference Organizational Roles**: (2025) NoDaLiDa (Social Media Chair); (2024) EACL SRW (Program Chair) **Conference Area Chair**: (2025–) ACL; (2025–) NAACL **Conference Reviewer**: (2025–) ICLR; (2024–) NeurIPS; (2022–) NAACL; (2021–) EMNLP; (2021–) ARR; (2021–) Several Workshops; (2019–) ACL

EXPERTISE CONSULTING	**Lightcast**; Consulting on Computational Job Market Analysis	2025
	Nejo; Consulting on Computational Job Market Analysis	2025
	IOM United Nations; Consulting on Computational Job Market Analysis	2025
	Stennevad; Consulting on Integrating Tailored Chatbots	2024
	IDIAP; Consulting on Computational Job Market Analysis	2024

Last Updated: December 4, 2025.

PUBLICATIONS

🛡 → **Equal Contribution**

Working Papers

P5. Nurlan Musazade, József Mezei, **Mike Zhang**. "**UniSkill: A Dataset for Matching University Curricula to Professional Competencies**". *Under Review*. 2025

P5. Elisa Bassignana, **Mike Zhang**, Dirk Hovy, Amanda Cercas Curry. "**Do Large Language Models Adapt to Language Variation across Socioeconomic Status?**". *Under Review*. 2025

P4. **Mike Zhang**, Johannes Bjerva, Russa Biswas. "**Follow the Path: Reasoning over Knowledge Graph Paths to Improve LLM Factuality**". *Under Review*. 2025

P3. Israfel Salazar, Manuel Fernandez, Shayekh Bin Islam, Arshia Soltani Moakhar, Shivalika Singh, Fabian Farestam, Angelika Romanou, ... (other authors), **Mike Zhang**, ... (other authors), Desmond Elliott, Enzo Ferrante, Sara Hooker, and Marzieh Fadaee. "**Kaleidoscope: In-language Exams for Massively Multilingual Vision Evaluation**". *Under Review*. 2025

P2. Rasmus Aavang, Giovanni Rizzi, Rasmus Bøggild, Alexandre Iolov, **Mike Zhang**, Johannes Bjerva. "**HIFI-KPI: A Dataset for Hierarchical KPI Extraction from Earnings Filings**". *Under Review*. 2025

P1. **Mike Zhang**, Amalie Pernille Dilling, Léon Gondelman, Niels Erik Ruan Lyngdorf, Euan D Lindsay, Johannes Bjerva. "**SEFL: A Framework for Generating Synthetic Educational Assignment Feedback with LLM Agents**". *Under Review*. 2025

Conference Proceedings

C21. Nurlan Musazade, **Mike Zhang**, József Mezei. "**Cross-Lingual Sentence-Level Skill Identification in English and Danish Job Advertisements**". *The 8th International Conference on Natural Language and Speech Processing*. 2025

C20. (**Best Practice Paper Nominee**) **Mike Zhang**, Euan D. Lindsay, Maj-Britt Quitzau, Johannes Bjerva. "**Scaling Course Evaluations with Large Language Models: Semester-level Digestible Student Feedback for Program Leaders**". *The 53rd Annual Conference of the European Society for Engineering Education*. 2025

C19. (**Highlight 13.5%**) Ashmal Vayani, ... (14 authors), **Mike Zhang**, ..., (48 authors), Thamar Solorio, Monojit Choudhury, Ivan Laptev, Mubarak Shah, Salman Khan, Fahad Shahbaz Khan. "**All Languages Matter: Evaluating LMMs on Culturally Diverse 100 Languages**". *The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2025*. 2025

C18. Euan Lindsay, **Mike Zhang**, Aditya Johri, Johannes Bjerva. "**The Responsible Development of Automated Student Feedback with Generative AI**". *IEEE Global Engineering Education Conference*. 2025

C17. Margaret Mitchell, ..., (50 authors), **Mike Zhang**, Sydney Zink, Aurélie Névéol, Zeerak Talat. "**SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models**". *In Proceedings of The 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics*. 2025

C16. (**Spotlight 1.8%**) Angelika Romanou, ..., (55 authors), **Mike Zhang**, Imanol Schlag, Marzieh Fadaee, Sara Hooker, Antoine Bosselut. "**INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge**". *In Proceedings of The Thirteenth International Conference on Learning Representations*. 2025

C15. Jeanette Falk, Yiyi Chen, Janet Rafner, **Mike Zhang**, Johannes Bjerva, and Alexander Nolte "**How Do Hackathons Foster Creativity? Towards Automated Evaluation of Creativity at Scale** ". *In Proceedings of The 2025 Conference on Human Factors in Computing Systems*. 2025

C14. (**ITU Research Award Nominee**) **Mike Zhang**⊘, Max Müller-Eberstein⊘, Elisa Bassignana, Rob van der Goot. "**Snakmodel: Lessons Learned from Training an Open Danish Large Language Model**". *The Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies*. 2025

C13. Rob van der Goot, Anette Jensen, Emil Allerslev Schledermann, Mikkel Wildner Kildeberg, Nicolaj Larsen, **Mike Zhang**, Elisa Bassignana. "**MorSeD: Morphological Segmentation of Danish and its Effect on Language Modeling**". *The Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies*. 2025

C12. **Mike Zhang**, Euan Lindsay, Frederik Bode Thorbensen, Danny Bøgsted Poulsen, Johannes Bjerva. "**Leveraging Large Language Models for Actionable Course Evaluation Feedback**". *The 52nd Annual Conference of the European Society for Engineering Education*. 2024

C11. Shivalika Singh, Freddie Vargus, Daniel Dsouza, Börje F. Karlsson, Abinaya Mahendiran, Wei-Yin Ko, Herumb Shandilya, Jay Patel, Deividas Mataciunas, Laura OMahony, **Mike Zhang**, Ramith Hettiarachchi, Joseph Wilson, Marina Machado, Luisa Souza Moura, Dominik Krzemiński, Hakimeh Fadaei, Irem Ergün, Ifeoma Okoh, Aisha Alaagib, Oshan Mudannayake, Zaid Alyafeai, Vu Minh Chien, Sebastian Ruder, Surya Guthikonda, Emad A. Alghamdi, Sebastian Gehrmann, Niklas Muennighoff, Max Bartolo, Julia Kreutzer, Ahmet Üstün, Marzieh Fadaee, Sara Hooker. "**Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning**". *In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. 2024.

C10. **Mike Zhang**, Rob van der Goot, Min-Yen Kan, and Barbara Plank. "**NNOSE: Nearest Neighbor Occupational Skill Extraction**". *In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics*. 2024.

C9. **Mike Zhang**, Rob van der Goot, and Barbara Plank. "**Entity Linking in the Job Market Domain**". *In Findings of the Association for Computational Linguistics: EACL 2024*. 2024.

C8. Maria Barrett⊘, Max Müller-Eberstein⊘, Elisa Bassignana⊘, Amalie Brogaard Pauli⊘, **Mike Zhang**⊘, and Rob van der Goot⊘. "**Can Humans Identify Domains?**". *In Proceedings of the 2024 Joint International Conference on Computational Linguistics*, *Language Resources and Evaluation*. 2024.

C7. **Mike Zhang**, Rob van der Goot, and Barbara Plank. "**ESCOXLM-R: Multilingual Taxonomy-driven Pre-training for the Job Market Domain**". *In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. 2023.

C6. Elisa Bassignana⚲, Max Müller-Eberstein⚲, **Mike Zhang**⚲, and Barbara Plank. "**Evidence > Intuition: Transferability Estimation for Encoder Selection**". *In Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing*. 2022

C5. Dennis Ulmer, Elisa Bassignana, Max Müller-Eberstein, Daniel Varab, **Mike Zhang**, Rob van der Goot, Christian Hardmeier, and Barbara Plank. "**Experimental Standards for Deep Learning in Natural Language Processing Research**". *In Findings of the Association for Computational Linguistics: EMNLP 2022*. 2022.

C4. **Mike Zhang**⚲, Kristian Nørgaard Jensen⚲, Sif Dam Sonniks, and Barbara Plank. "**SKILLSPAN: Hard and Soft Skill Extraction from Job Postings**". *In Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 2022.

C3. **Mike Zhang**⚲, Kristian Nørgaard Jensen⚲, and Barbara Plank. "**KOMPETENCER: Fine-grained Skill Classification in Danish Job Postings via Distant Supervision and Transfer Learning**". *In Proceedings of the 13th Edition of its Language Resources and Evaluation Conference*. 2022.

C2. **Mike Zhang** and Barbara Plank. "**Cartography Active Learning**". *In Findings of the Association for Computational Linguistics: EMNLP 2021*. 2021.

C1. **Mike Zhang**⚲, Kristian Nørgaard Jensen⚲, and Barbara Plank. "**De-identification of Privacy-related Entities in Job Postings**". *Proceedings of the 23rd Nordic Conference on Computational Linguistics*. 2021.

Workshops
W10. **Mike Zhang**, Rob van der Goot. "**NLPnorth @ TalentCLEF 2025: Comparing Discriminative, Contrastive, and Prompt-Based Methods for Job Title and Skill Matching**". *CLEF*. 2025

W9. Rune Birkmose, Nathan Mørkeberg Reece, Esben Hofstedt Norvin, Johannes Bjerva, **Mike Zhang**. "**On-Device LLMs for Home Assistant: Dual Role in Intent Detection and Response Generation**". *The 10th Workshop on Noisy and User-generated Text*. 2025

W8. Max Müller-Eberstein⚲, **Mike Zhang**⚲, Elisa Bassignana, Peter Trolle, Rob van der Goot "**DAKULTUR: Evaluating the Cultural Awareness of Language Models for Danish with Native Speakers**". *The 3rd Workshop on Cross-Cultural Considerations in NLP*. 2025

W7. Khanh Cao Nguyen, **Mike Zhang**, Syrielle Montariol, and Antoine Bosselut. "**Rethinking Skill Extraction in the Job Market Domain using Large Language Models**". *In Proceedings of the 1st Workshop on Natural Language Processing for Human Resources*. 2024.

W6. Antoine Magron, Anna Dai, **Mike Zhang**, Syrielle Montariol, and Antoine Bosselut. "**JobSkape: A Framework for Generating Synthetic Job Postings to Enhance Skill Matching**". *In Proceedings of the 1st Workshop on Natural Language Processing for Human Resources*. 2024.

W5. Elena Senger, **Mike Zhang**, Rob van der Goot, and Barbara Plank. "**Deep Learning-based Computational Job Market Analysis: A Survey on Skill Extraction and Classification from Job Postings**", *In Proceedings of the 1st Workshop on Natural Language Processing for Human Resources*. 2024.

W4. **Mike Zhang**, Kristian Nørgaard Jensen, Rob van der Goot, and Barbara Plank. "**Skill Extraction from Job Postings using Weak Supervision**". *RecSys in*

HR'22: The 2nd Workshop on Recommender Systems for Human Resources, *in conjunction with the 16th ACM Conference on Recommender Systems*. 2022

W3. (**Best Paper Award**) Dennis Ulmer, Elisa Bassignana, Max Müller-Eberstein, Daniel Varab, **Mike Zhang**, Christian Hardmeier, and Barbara Plank. "**A Common Set of Experimental Standards for Deep Learning Research: A Natural Language Processing Perspective**". *Machine Learning Evaluation Standards workshop at ICLR 2022.* (Non-Archival) 2022.

W2. **Mike Zhang** and Antonio Toral. "**The Effect of Translationese in Machine Translation Test Sets**". *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*. 2019.

W1. **Mike Zhang**⊗, Roy David⊗, Leon Graumans⊗, and Gerben Timmerman⊗. "**Grunn2019 at SemEval-2019 Task 5: Shared Task on Multilingual Detection of Hate**". *Proceedings of the 13th International Workshop on Semantic Evaluation*. 2019.

Technical Reports

T2. Long Phan, Alice Gatti, Ziwen Han, Nathaniel Li, Josephina Hu, ... (other authors), **Mike Zhang**, ... (other authors), Alexandr Wang, Dan Hendrycks. "**Humanity's Last Exam**". *Technical Report*. 2025

T1. Longxu Dou, Qian Liu, Fan Zhou, Changyu Chen, ... (30 authors), **Mike Zhang**, Shiqi Chen, Tianyu Pang, Chao Du, Xinyi Wan, Wei Lu, Min Lin. "**Sailor2: Sailing in South-East Asia with Inclusive Multilingual LLM**". *Technical Report*. 2025