MIKE ZHANG

Copenhagen, Denmark

jjz@cs.aau.dk ⋄ linkedin.com/in/jjzha ⋄ jjzha.github.io

EXPERIENCE

Postdoctoral Researcher 02.2024 – Current
Aalborg University Copenhagen Copenhagen, DK
Research Affiliation: TypNLP Group | Advisors: Johannes Bjerva & Euan Lindsay

- Researching at the intersection of NLP and Education;
- Developing large language models to generate actionable feedback to student assignments;

Postdoctoral Researcher 01.2024 – 02.2024
IT University of Copenhagen Copenhagen, DK
Research Affiliations: NLPnorth | Advisors: Rob van der Goot

- Lead and trained large language models for the Danish language;

Ph.D. Research Intern 10.2023 – 11.2023
Swiss Federal Institute of Technology Lausanne (EPFL) Lausanne, CH
Research Affiliation: EPFL's NLP Lab | Host: Antoine Bosselut

- Researched applying large language models to the job market domain;
- Co-supervised a master student on generating synthetic job posting data for skill matching;
- Published two workshop papers related to the topic.

Ph.D. Research Intern 02.2023 – 07.2023
National University of Singapore (NUS) Singapore, SG
Research Affiliation: WING Group | Host: Min-Yen Kan

- Conducted research with respect to Natural Language Processing and Information Retrieval related to job descriptions and related text sources;
- Investigated Nearest Neighbor Language Models for occupational skill extraction;
- Published results at a top-tier NLP conference.

Ph.D. Research Intern 09.2022 – 12.2022
NEC Laboratories Europe Heidelberg, DE
Research Affiliation: Human-centric AI Unit | Host: Kiril Gashteovski

- Improved precision of neural Open Information Extraction systems using data-centric approaches;
- Developed a data pre-processing and cleaning pipeline for large-scale Open Information Extraction.

EDUCATION

Ph.D. Natural Language Processing 2020 – 2024
IT University of Copenhagen / Ludwig Maximilian University of Munich   Copenhagen, DK / Munich, DE /
Research Affiliations: NLPnorth / MaiNLP / P1 | Advisors: Barbara Plank & Rob van der Goot

- Researched how to extract structured data from job ads, published results at top-tier NLP conferences;
- Managed a small team with a research assistant and annotator for developing datasets and models;
- Open-sourced (large) language models and public datasets on the Huggingface platform.
- Officially graduated from the Ph.D. in three years (the standard in Denmark) and published 14 (10 first author) peer-reviewed scientific papers during the Ph.D. including publications in top NLP venues and workshops (ACL, EMNLP, NAACL, EACL, LREC, COLING).

SELECTED PUBLICATIONS (* = EQUAL CONTRIBUTION)

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

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.

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.

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

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.

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.

COMMUNITY SERVICE

Chair:
NoDaLiDa/Baltic-HLT (Social Media Chair; 2025), EACL SRW (Co-chair; 2023–2024), LREC (Session Chair; 2022)

Program Committee:
ARR (2021–), ACL (2019–), EMNLP (2021–), NAACL (2022–), LREC (2022–), RecSysHR (2022–)

LEADERSHIP

Language Ambassador   9.2023 – 2.2024
Cohere For AI   Remote, DK
I was the Dutch language ambassador in the Aya project by Cohere for AI. By creating high-quality instruction data for large multilingual language models we made sure no language is left behind. I communicated with the region ambassadors on the progress of high-quality Dutch instruction examples and contributed ±500 examples.

AWARDS

Research Stay Abroad (IT University of Copenhagen)   2022
Received **3,500 EUR** for a research visit abroad in an external institution. The destination is the National University of Singapore (NUS).

Outstanding Paper Award (ML Evaluation Standards Workshop)   2022
Awarded **2,000 USD** for an outstanding paper award at the ML Evaluation Standards Workshop at ICLR 2022, see paper [5].

Student Scholarship Award (EMNLP 2021, 2022)   2021-2022
Received free attendance to EMNLP 2021 (eq. **300 USD**) in Punta Cana, DR.
Received free attendance to EMNLP 2022 (eq. **325 USD**) in Abu Dhabi, UAE.

Marco Polo Fund (University of Groningen)   2018
Awarded **1,000 EUR** for studying abroad outside of Europe. For this grant there is one awardee per semester. Destination was The Chinese University of Hong Kong (CUHK). Courses in Java Programming, Politics, Computer Networks.