

# MIKE ZHANG

Copenhagen, Denmark

[mikz@itu.dk](mailto:mikz@itu.dk) ♦ [linkedin.com/in/jjzha](https://www.linkedin.com/in/jjzha) ♦ [github.com/jjzha](https://github.com/jjzha)

## EDUCATION

**IT University of Copenhagen /** Copenhagen, DK  
**Ludwig Maximilian University of Munich**, Research Visitor Munich, DE  
PhD, Natural Language Processing Dec. 2023

Thesis: Natural Language Processing for Labor Market Information. Advised by prof. [Barbara Plank](#).

**University of Groningen** Groningen, NL  
MA, Information Science Jan. 2020

Thesis: The Effect of Translationese in Machine Translation Test Sets. Advised by dr. [Antonio Toral Ruiz](#).

**University of Groningen** Groningen, NL  
BSc, Information Science Jun. 2018

Thesis: Labeled Bilexical Dependencies for Deception Detection. Advised by prof. [Barbara Plank](#).

## WORK EXPERIENCE

**Dataprovider.com** Groningen, NL  
Data Engineer Jan. 2020 – Aug. 2020

- Created requested features into the proprietary search engine by maintaining code quality with PEP8 standards, basic unit- and integration tests, Git CI/CD, and code reviews;
- Spearheaded improvement of the preprocessing steps for structuring 72M unstructured company addresses with Python and Google's Geocoding API;
- Implemented a classifier predicting company SIC codes via character-level features using Scikit-Learn, improved predictions by 70% accuracy.

## PUBLICATIONS (SELECTED)

**Mike Zhang**, Kristian Nørgaard Jensen, Sif Dam Sonniks, and Barbara Plank. “**SkillSpan: Hard and Soft Skill Extraction from Job Postings.**” *To appear at the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. 2022.

**Mike Zhang**, Kristian Nørgaard Jensen, and Barbara Plank. “**Kompetencer: Fine-grained Skill Classification in Danish Job Postings via Distant Supervision and Transfer Learning.**” *To appear at the 13th Edition of its Language Resources and Evaluation Conference (LREC)*. 2022.

Dennis Ulmer, Elisa Bassignana, Max Müller-Eberstein, Daniel Varab, **Mike Zhang**, Christian Hardmeier, and Barbara Plank. “**Experimental Standards for Deep Learning Research: A Natural Language Processing Perspective.**” *Machine Learning Evaluation Standards workshop at ICLR 2022*. 2022. **Outstanding Paper Award**.

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

Kristian Nørgaard Jensen, **Mike Zhang**, and Barbara Plank. “**De-identification of Privacy-related Entities in Job Postings.**” *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. 2021.

Further work can be found at [Google Scholar](#).

## SERVICES

**Reviewer** NAACL SRW (2022), ARR (2021, 2022), W-NUT (2021), CoNLL (2021), ACL (2019)  
**Volunteer** EMNLP (2021), CLIN (2019)

## SKILLS

**Programming** Python, Java, R, HTML, CSS  
**Tools** Git, Doccano, Scikit-Learn, PyTorch  
**Hard Skills** Scientific Writing, Research Management, Public Speaking, Data Analysis  
**Soft Skills** Time Management, Teamwork, Communication, Problem Solving, Leadership, Accountability  
**Languages** Dutch (Native), English (Fluent), Spanish (Basic), Mandarin (Basic), Danish (Basic)