

# Mike Zhang

mikz@itu.dk · Copenhagen, Denmark

[jjzha.github.io](https://jjzha.github.io) · [github.com/jjzha](https://github.com/jjzha) · [linkedin.com/in/jjzha](https://linkedin.com/in/jjzha)

## Education

### IT University of Copenhagen

*PhD Computer Science*

Aug. 2023 (Expected)  
Copenhagen, Denmark

- Extracting information from multilingual job vacancy data with time-shift, as a result to create better job matching;
- Implementing approaches from Information Extraction, Active Learning, Distant Supervision to job postings;
- Interests: Information Extraction, Active Learning, Distant Supervision, Transfer Learning, and Domain Adaptation.

### University of Groningen

*MA Information Science, Cum Laude*

Jan. 2020  
Groningen, Netherlands

- **Master Thesis:** *Translationese in the WMT: The Effect of Translationese in Machine Translation Test Sets*. Grade: 9.0/10.0;
- Presented long paper at the Conference of Machine Translation (WMT) in August 2019.

### University of Groningen

*BSc Information Science*

Jun. 2018  
Groningen, Netherlands

- **Exchange Semester** at The Chinese University of Hong Kong, Hong Kong SAR (Spring 2018);
- **Bachelor Thesis:** *From Relations to Falsehood: Labeled Bilexical Dependencies for Deception Detection*. Grade: 8.0/10.0.

## Work Experience

### Dataprovider.com

*Data Engineer*

Jan. 2020 – Aug. 2020  
Groningen, Netherlands

- Built a classifier predicting SIC codes for company names, using Python and scikit-learn;
- Developed preprocessing steps for structuring company addresses;
- Maintained code quality with PEP8 standards, basic unit- and integration tests, Git CI/CD.

### Dataprovider.com

*Data Scientist Intern*

Aug. 2019 – Dec. 2019  
Groningen, Netherlands

- Developed algorithms in Python for the generation of domain names using core Natural Language Processing techniques;
- Obtained 1M+ existing and undiscovered domain names for indexing.
- Integrating model into production using Python, Elasticsearch and NGINX.

## Publications, \* = equal contribution

1. **M. Zhang** and B. Plank. Cartography Active Learning. To appear in In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
2. K. N. Jensen\*, **M. Zhang\***, and B. Plank. De-identification of privacy-related entities in job postings. In Proceedings of the Nordic Conference on Computational Linguistics (NoDaLiDa), 2021.
3. **M. Zhang** and A. Toral. The Effect of Translationese in Machine Translation Test Sets. In Proceedings of the Fourth Conference on Machine Translation (WMT), 2019.
4. **M. Zhang**, R. David, L. Graumans, and G. Timmerman. Grunn2019 at SemEval-2019 Task 5: Shared Task on Multilingual Detection of Hate. In Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval), 2019.

## Skills

---

**Programming:** Python (PyTorch, scikit-learn, Pandas, NumPy) · R · Java · HTML

**Tools:** Git · Elasticsearch · Agile · Scrum · ~~La~~TeX

**Languages:** Dutch (Native) · English (Fluent) · Spanish (Basic) · Mandarin (Basic) · Danish (Basic)

## Services

---

### Reviewer

The SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2021

The 57th Annual Meeting of the Association for Computational Linguistics (ACL) 2019

## Teaching

---

### IT University of Copenhagen

Communicating State-of-the-art NLP Research to a Broader Audience (PhD Course) Fall 2021

Introduction to Natural Language Processing and Deep Learning Spring 2021

### University of Groningen

Learning from Data Fall 2019

Machine Learning Fall 2019, 2020

Social Media Spring 2019