




# Mike Zhang

---

IT University of Copenhagen, Department of Computer Science ([NLPnorth](#))  
Rued Langgaards Vej 7, DK 2300, Copenhagen S, Denmark, Building 3E13  
+45-7218-5204  
[mikz@itu.dk](mailto:mikz@itu.dk) | [jjzha.github.io](https://jjzha.github.io) | [github.com/jjzha](https://github.com/jjzha)

ACADEMIC BACKGROUND	<i>Ph.D. Natural Language Processing</i>	2023
	<a href="#">IT University of Copenhagen</a> (ITU), Copenhagen, DK / <a href="#">Ludwig Maximilian University of Munich</a> (LMU), Munich, DE (Research Visitor)	
	<ul style="list-style-type: none"><li>• Ph.D. research in Information Extraction for Job Market Analysis advised by prof. <a href="#">Barbara Plank</a>.</li><li>• Leading a small research team with a research assistant and annotator.</li></ul>	
	<i>M.A. Information Science</i>	2020
	<a href="#">University of Groningen</a> (RUG), Groningen, NL	
	<ul style="list-style-type: none"><li>• Focus areas: Computational Linguistics, Semantics, Evaluation.</li><li>• Thesis: The Effect of Translationese in Machine Translation Test Sets. Advised by dr. <a href="#">Antonio Toral Ruiz</a>.</li></ul>	
	<i>B.Sc. Information Science</i>	2018
	<a href="#">University of Groningen</a> (RUG), Groningen, NL	
	<ul style="list-style-type: none"><li>• Focus areas: Computational Linguistics, Linguistics, Web Technology.</li><li>• Thesis: From Relations to Falsehood: Labeled Bilexical Dependencies for Deception Detection. Advised by prof. <a href="#">Barbara Plank</a></li><li>• Exchange Semester at <a href="#">The Chinese University of Hong Kong</a> (CUHK), HK.</li></ul>	
WORK EXPERIENCE	<i>Data Engineer</i>	Jan. 2020 – Aug. 2020
	<a href="#">Dataprovider.com B.V.</a> , Groningen, NL	
	<ul style="list-style-type: none"><li>• Implemented 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;</li><li>• Built a classifier predicting SIC codes for company names via character-level features, using Python and scikit-learn, improved predictions by 70% accuracy;</li><li>• Spearheaded improvement of the preprocessing steps for structuring 72M unstructured company addresses with Python and Google's Geocoding API.</li></ul>	
	<i>Data Scientist Intern</i>	Aug. 2019 – Dec. 2019
	<a href="#">Dataprovider.com B.V.</a> , Groningen, NL	
	<ul style="list-style-type: none"><li>• Developed five text generation algorithms in Python for the generation of undiscovered country code top-level domains for five countries: Found ~15K undiscovered domain names/hr;</li><li>• Obtained in total 1M+ existing, but undiscovered country code top-level domains for indexing in the search engine product.</li></ul>	
PEER- REVIEWED PUBLICATIONS	7. <b>Mike Zhang</b> ,  Kristian Nørgaard Jensen,  Sif Dam Sonniks, and Barbara Plank. "SKILLSPAN: Hard and Soft Skill Extraction from Job Postings." <i>In Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics</i> . 2022.	
 Equal Contribution		

6. **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.
5. 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*. 2022. **Outstanding Paper Award**
4. **Mike Zhang** and Barbara Plank. “Cartography Active Learning.” *Findings of the Association for Computational Linguistics: EMNLP 2021*. 2021.
3. 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*. 2021.
2. **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.
1. **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.

## GRANTS & AWARDS

**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) 2021  
 Received free attendance to EMNLP 2021 (eq. **300 USD**) at Punta Cana, DR.

**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.

## SERVICES

- **Reviewer:** NAACL SRW (2022), ARR (2021, 2022), W-NUT (2021), CoNLL (2021), ACL (2019)
- **Volunteer:** EMNLP (2021), CLIN (2019)
- **Talks:** Talk at the CL group of the University of Groningen, [GroNLP](#) (2021)

## CORE COMPETENCES

- **Programming:** Python (PyTorch, scikit-learn, Pandas, NumPy), R, Java, HTML
- **Tools:** Git, Elasticsearch, Agile, Scrum, L<sup>A</sup>T<sub>E</sub>X
- **Skills:** Scientific writing, research management, public speaking, data analysis
- **Languages:** Dutch (Native), English (Fluent), Spanish (Basic), Mandarin (Basic), Danish (Basic)