

## EDUCATION

<b>Utrecht University</b>	Utrecht, the Netherlands
Ph.D. in Computer Science, Advisor: Dr. Dong Nguyen and Prof. Albert Gatt	07.2020–08.2025
– Working on data-centric NLP and ML, with 30% teaching load and 70% research load.	
<b>East China Normal University</b>	Shanghai, China
M.Sc. in Computer Science, Advisor: Dr. Yuanbin Wu	09.2017–07.2020
– Thesis: “Research on Detecting Gender Stereotypes in Texts”, long paper in EMNLP 2019.	
<b>East China Normal University</b>	Shanghai, China
B.Sc. in Psychology	09.2013–07.2017
– Thesis: “Research on the Attention of Negative Emotional Stimuli”	

## EXPERIENCE

<b>LMU Munich</b>	Munich, Germany
Visiting researcher	01.2025 - 03.2025
Visiting PhD researcher under the supervision of Prof. Barbara Plank. Working on training data attribution for reasoning tasks.	
<b>Amazon</b>	Berlin, Germany
Applied Scientist Intern	11.2022 - 05.2023
Developed a user-data-driven approach for enriching product catalogs (knowledge graph). Under the mentorship of Prof. Ziawasch Abedjan. Published as a long paper at ECML-PKDD 2024.	
<b>Sogou Research</b>	Hangzhou, China
Research Intern	07.2019 - 01.2020
Developed a question generation system to improve the generation diversity. Used crowd-sourcing to annotate a Chinese question generation dataset consisting of 20k+ samples.	

## SELECTED PUBLICATIONS

1. **Yupei Du**, Yingjin Song, Hugh Mee Wong, Daniil Ignatev, Albert Gatt, and Dong Nguyen, “Disentangling the Roles of Representation and Selection in Data Pruning”, in *ACL 2025*
2. Yuqian Li, **Yupei Du**, Yufang Liu, Feifei Feng, Mou Xiao Feng, and Yuanbin Wu, “On Support Samples of Next Word Prediction”, in *ACL 2025*
3. Yingjin Song, **Yupei Du**, Denis Paperno, and Albert Gatt, “Burn After Reading: Do Multimodal Large Language Models Truly Capture Order of Events in Image Sequences?”, in *Findings of ACL 2025*
4. **Yupei Du**, Albert Gatt, and Dong Nguyen, “FTFT: efficient and robust Fine-Tuning by transFerring Training dynamics”, in *COLING 2025*
5. **Yupei Du**, Jacek Golebiowski, Philipp Schmidt, and Ziawasch Abedjan, “Guiding Catalogue Enrichment with User Queries”, in *ECML-PKDD 2024*.
6. Goya van Boven, **Yupei Du**, and Dong Nguyen, “Transforming Dutch: Debiasing Dutch Coreference Resolution Systems for Non-binary Pronouns”, in *FAccT 2024*.

7. **Yupei Du** and Dong Nguyen, “Measuring the Instability of Fine-Tuning”, in *ACL 2023*.
8. **Yupei Du**, Qi Zheng, Yuanbin Wu, Man Lan, Yan Yang, and Meirong Ma, “Understanding Gender Bias in Knowledge Base Embeddings”, in *ACL 2022*.
9. **Yupei Du**, Qixiang Fang, and Dong Nguyen, “Assessing the Reliability of Word Embedding Gender Bias Measures”, in *EMNLP 2021*.
10. **Yupei Du**, Yuanbin Wu, and Man Lan, “Exploring Human Gender Stereotypes with Word Association Test”, in *EMNLP 2019*.

## ACADEMIC SERVICE

---

- **Program Committee** 2021 - Current  
*ACL, EMNLP, NAACL, EACL, ACL Rolling Review, COLM (2024-)*
- **Outstanding Reviewer**  
*ACL 2022*

## SELECTED TEACHING ACTIVITIES

---

- **Teaching Assistant and Lecturer** at Utrecht University 2024, 2025  
*Transformers: Applications in Language and Communication (INFOMTALC)*
- **Teaching Assistant** at Utrecht University 2021 - 2024  
*Applications of Machine Learning (INKUB3APML)*
- **Teaching Assistant** at Utrecht University 2021 - 2024  
*Human-centered Machine Learning (INFOMHCML)*

## EXAMPLE SUPERVISION

---

- Goya van Boven, “Transforming Dutch: Debiasing Dutch Coreference Resolution Systems for Non-Binary Pronouns”, Master’s thesis, 2024. Long paper in FAccT 2024.
- Daan van der Weijden, “Dutch Data Augmentation For Low-Resource Tasks – Can Large Language Models be Used for Dutch Data Augmentation?”, Master’s thesis, 2023.
- “CREATED: Generating Viable Counterfactual Sequences for Predictive Process Analytics Authors”, Master’s thesis, 2023. Long paper in CAiSE 2023.
- Xinyu Zhang, “Study on Iterative Nullspace Projection Debiasing of Multiple Attributes of Word Embedding”, Master’s thesis, 2022.

## SCHOLARSHIPS AND AWARDS

---

- Utrecht University Teaching Award 2022  
In recognition of distinguished teaching of Applications of Machine Learning.
- ECNU Zhihui Award 2019  
Top 5% of the master’s students in East China Normal University.