Vera Neplenbroek

 Vera Neplenbroek | Vera Neplenbroek | Veranep
 veranep.github.io | vera@veraneplenbroek.com

EDUCATION

2023 - present	PhD on Safety of Conversational AI	**ILLC, University of Amsterdam**
	Advisors: Prof. Dr. Raquel Fernández & Dr. Arianna Bisazza	
2021 - 2023	Master's Degree (Artificial Intelligence)	**University of Amsterdam**
	GPA: 9/10	
2018 - 2021	Bachelor's Degree (Liberal Arts and Sciences)	**Amsterdam University College**
	GPA: 3.93/4.0	

WORK EXPERIENCE

Data Scientist Intern   Jan 2020 - Jun 2021

Machine Learning Programs, Amsterdam

– Training machine learning models for premium prediction and propensity to claim.

– Incorporating domain knowledge on car insurance.

– Writing literate programming documents analyzing data sets, and model performance.

– Supervising groups of undergraduates during their one month internship projects.

TEACHING EXPERIENCE

Supervision

– Master thesis on 'Investigating the Effects of Code-Switching on Toxicity and Biases Exhibited by Large Language Models'.

– Internship on 'Explainable Customer Call Classification with Text Bottleneck Models'.

Teaching Assistant

– 4 courses in the MSc AI program: Natural Language Processing 1 (3x), Advanced Topics in Computational Semantics (2x), Knowledge Representation and Reasoning, and Automated Planning.

– Teaching tutorial sessions, assisting students with their assignments, and answering their questions.

– Giving (guest) lectures.

– Designing, supervising and grading group projects.

– Creating and grading of assignments and exams.

PUBLICATIONS

[1] Vera Neplenbroek et al. "Reading Between the Prompts: How Stereotypes Shape LLM's Implicit Personalization". In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Ed. by Christos Christodoulopoulos et al. Suzhou, China: Association for Computational Linguistics, Nov. 2025, pp. 20367–20400. ISBN: 979-8-89176-332-6. DOI: 10.18653/v1/2025.emnlp-main.1029. URL: https://aclanthology.org/2025.emnlp-main.1029/.

[2] Vera Neplenbroek et al. "Cross-Lingual Transfer of Debiasing and Detoxification in Multilingual LLMs: An Extensive Investigation". In: *Findings of the Association for Computational Linguistics: ACL 2025*. Ed. by Wanxiang Che et al. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 2805–2830. ISBN: 979-8-89176-256-5. URL: https://aclanthology.org/2025.findings-acl.145/.

[3] Anna Bavaresco et al. "LLMs instead of Human Judges? A Large Scale Empirical Study across 20 NLP Evaluation Tasks". In: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Ed. by Wanxiang Che et al. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 238–255. ISBN: 979-8-89176-252-7. URL: https://aclanthology.org/2025.acl-short.20/.

[4] Vera Neplenbroek et al. "MBBQ: A Dataset for Cross-Lingual Comparison of Stereotypes in Generative LLMs". In: *First Conference on Language Modeling*. 2024. URL: https://openreview.net/forum?id=X9yV4lFHt4.

SKILLS

Languages	Dutch, English
Programming	Python (Pytorch, Transformers, NLTK, Scikit-learn, Pandas, Numpy, Seaborn)
Tools	Git, Linux, slurm, LaTeX

TALKS & PRESENTATIONS

– **NatWest Group DS Seminar Series**
August 2025 · Online
– **HumanCLAIM Workshop**
March 2025 · Göttingen
– **CL Group Seminar**
March 2025 · Groningen
– **Workshop on New Perspectives on Bias and Discrimination in Language Technology**
Nov 2024 · Amsterdam

SERVICE

– **Reviewing**
ARR, ECIR 2026 - IR-for-Good track, EACL SRW 2026, GeBNLP 2025, BlackboxNLP 2025, CLIN 34
– **Volunteering**
COLM 2024, ACL 2025
– **Organizing**
  – Co-organizing the next edition of the Gender Bias in NLP (GeBNLP) workshop.
  – CLS (Computational Linguistics Seminar) of the University of Amsterdam.
  – NLPitch (Series of talks about ongoing and published work by NLP PhD students at the ILLC)