

Sara Vera Marjanović

PhD Student in Information Retrieval & NLP

[Github](#) | [Google Scholar](#) | [LinkedIn](#) | [Website](#)

+45 91 41 07 74 | Copenhagen, Denmark | savema@di.ku.dk

Second-year PhD student working to ensure fair and transparent language models, with research interests in interpretability, bias detection, and uncertainty.

Education

- 2023 – now **PhD fellow in Computer Science**, University of Copenhagen, Denmark
» Joint supervision by Christina Lioma and Maria Maistro in Information Retrieval and Isabelle Augenstein in NLP. Anticipated graduation in 04 / 2026.
» Investigating the behaviour of LLMs in high-uncertainty domains using XAI.
- 2019 – 2021 **MSc in IT & Cognition**, University of Copenhagen, Denmark
» Identified gender biases on bipartisan political discussions on Reddit combining techniques from Natural Language Processing and Network Science.
» Coursework in Large-Scale Data Science, Machine Learning, Natural Language Processing, Image Processing, and Robotics. cGPA: 11.8/12.0
- 2012 – 2016 **BSc in Neuroscience**, McGill University, Canada
» Courses in computational neuroscience, evolutionary biology, and computer science. cGPA: 3.91/4.00, Graduated with Distinction. Dean's Honour List: 2013, 2015

Work experience

- 2025 - now **Graduate Research Trainee**, Mila – Quebec Artificial Intelligence Institute, Canada
» Supervision by Dr. Siva Reddy and Dr. Karolina Stanczak in the McGill NLP Group
» Investigating applications of mechanistic interpretability for AI safety
- 2022 – 2023 **Consultant**, Capgemini Insights & Data, Denmark
» Implemented forecasting ML pipelines to production for clients in the public sector using cloud-computing services (Azure, AWS)
» Led business development projects to improve team competencies within MLOps, multimodal synthetic data, and quantum machine learning.
- 2020 – 2022 **Data Scientist**, FocusWRX, Denmark
» Forecasted user stress using Garmin watch (HRV) and Outlook data to determine significant stressors in work environments and present in-app insights to users.
» Conducted user analytics research and interviews to guide feature development.
- 2020 – 2022 **Research Assistant**, Copenhagen Center for Social Data Science, Denmark
» Modelled network transmission and narrative patterns of misinformation on Twitter (now X) as part of a multi-university, multi-disciplinary collaboration investigating the social impact of the coronavirus pandemic (as part of the HOPE project).

Publications

- Preprint* Hagström, **Marjanovic**, Yu, Arora, Lioma, Maistro, Atanasova & Augenstein. A Reality Check on Context Utilisation for Retrieval-Augmented Generation. Pending review.
<https://arxiv.org/abs/2412.17031>
- 2024 **Marjanovic**, Yu, Atanasova, Maistro, Lioma & Augenstein. DYNAMICQA: Tracing Internal Knowledge Conflicts in Language Models. *EMNLP 2024 Findings*.
<https://aclanthology.org/2024.findings-emnlp.838/>
- 2024 **Marjanovic**, Augenstein & Lioma. Investigating the Impact of Model Instability on Explanations and Uncertainty. *ACL 2024 Findings*. aclanthology.org/2024.findings-acl.705/

- 2022 **Marjanovic**, Stańczak, & Augenstein (2022). Quantifying Gender Biases Towards Politicians on Reddit. *PlosONE*. journals.plos.org/plosone/article?id=10.1371/journal.pone.0274317
- 2022 Kjær, **Marjanovic**, Johansen, Baglini & Adler-Nissen (2022). Misinformation, social status og latterliggørelse: En undersøgelse af danskeres spredning af og reaktioner på Covid-19 misinformation på Twitter. *Politica*.
- 2022 Johansen, **Marjanovic**, Kjær, Baglini & Adler-Nissen (2022). Ridiculing the “tin foil hats:” Citizen responses to COVID -19 misinformation in the Danish facemask debate on Twitter. *Harvard Misinformation Review*. misinforeview.hks.harvard.edu/article/ridiculing-the-tin-foil-hats-citizen-responses-to-covid-19-misinformation-in-the-danish-facemask-debate-on-twitter/

Teaching Experience

- 2023 **Teaching Assistant**, University of Copenhagen, Fair & Transparent Machine Learning
 » Designed practical sessions in model (LLM, VLM) probing, fairness metrics and XAI
 » Provided feedback and guidance to student presentations and projects
- 2020 **Teaching Assistant**, University of Copenhagen, Introduction to Social Data Science
 » Led hybrid practical lessons introducing python programming, pandas and scraping to first-year Masters students.
 » Taught biweekly quantitative analysis workshops.

Awards

- 2021 **Columbus Prisen** (as part of the HOPE project), *Forlaget Columbus*
 50000 DKK award for public institutions that uphold ideas of democracy, genuine political commitment and the importance of factual arguments.
- 2015 **Quinn Research Assistantship Award**, *University of British Columbia*
 6000 CAD research grant for undergraduate research within psychology, for a summer internship at the University of British Columbia.

Skills

Coding Languages: Python, R, LaTeX

Languages: English (C2), Serbo-Croatian (C1), Danish (C1), French (B2)

Tools: Pytorch, Tensorflow, Huggingface, Cloud-computing services (Microsoft Azure, AWS), Git, Pyspark, NLTK, scikit-learn, Pandas, Data visualization (Matplotlib), spaCy, wandb

Hobbies: Word puzzles, rock-climbing, endurance sports (e.g. long-distance running, cycle-touring), reading, needle crafts (e.g. crocheting)