Janosch Haber

Lead NLP Research Scientist | PhD in Computer Science: Computational Linguistics

✉ janoschhaber@gmail.com | ⌂ janoschhaber.com | in janoschhaber

Summary

Lead NLP Research Scientist looking for an opportunity that combines AI research challenges with social impact. Proven track record of academic research skills, hands-on expertise in developing AI solutions from initial data collection to commercial deployment, and on-the-job experience in leading teams and projects to reach their goals.

Experience

Lead NLP Research Scientist   *London, United Kingdom*
Rewire   *Jun 2022 - PRESENT*

- At Rewire, we develop state-of-the-art, custom-tailored AI models to detect toxic language such as hate, abuse, threats and explicit content. As Lead Research Scientist I co-ordinate projects and contribute to engineering context-sensitive language models all the way from initial data collection to deployment in our clients' applications.
- Skills: Natural Language Processing (NLP), Data Science, Project Management, Deep Learning, Python, Pandas, PyTorch, MLOps, AWS

Co-Founder, Board Member and Webdesigner
Stichting Correspondents of the World (CotW)   *Jan 2020 - PRESENT*

- CotW is a non-profit aimed at bringing together people from different backgrounds so they can share their perspectives and better understand one-another. I'm a co-founder and board member, balancing everyday organisational tasks with shaping our long-term mission.
- Skills: Organisation Management, Team Management, Team Leadership

Course Assistant for Natural Language Processing   *London, United Kingdom*
Queen Mary University of London (QMUL)   *Sep 2019 – Sep 2021*

- As Course Demonstrator for the NLP module I prepared course materials, helped students with questions about their assignments (Python programming, text processing, classification) and assisted the course coordinator in scoring assessments and exams.

Student Assistant for the BSc Artificial Intelligence   *Amsterdam, The Netherlands*
University of Amsterdam (UvA)   *Jan 2015 - Jun 2017*

- During my Bachelors and Masters I was a Student Assistant for a number of AI Bachelor's courses. I assisted the course coordinators by helping out students with their course assignments and grading assessments and exams.

Education

PhD in Computer Science: Computational Linguistics   *London, United Kingdom*
Queen Mary University of London (QMUL)   *Sep 2018 - Jul 2022*

- Supervised by Prof. Dr. Massimo Poesio, Dr. Julian Hough and Prof. Dr. Patrick Healey
- For my PhD I investigated nuances in the interpretation of ambiguous words - both in the human language processor and in context-sensitive AI language models. I collected a large-scale dataset of similarity ratings and with it was able to show that humans as well as current computer models do distinguish between some senses while merging others.
- My work has been published in top-tier conferences (see below), my thesis can be found here.
- Research funded by the Disagreements in Language Interpretation (DALI) project, ERC Grant 695662, and in part supported by the Enrichment Scheme of The Alan Turing Institute.

Master of Science (MSc) in Artificial Intelligence   *Amsterdam, The Netherlands*
University of Amsterdam (UvA)   *Sep 2016 - Jul 2018*

- Cum Laude; Final Grade 8.8
- Thesis supervised by Prof. Dr. Raquel Fernández and Dr. Elia Bruni
- Thesis Title: How should we call it? - Introducing the PhotoBook Conversation Task and Dataset for Training Natural Referring Expression Generation in Artificial Dialogue Agents.
- Thesis research funded by the Facebook ParlAI Research Award.

Bachelor of Science (BSc) in Artificial Intelligence   *Amsterdam, The Netherlands*
University of Amsterdam (UvA)   *Sep 2012 - Jul 2015*
- Cum Laude; Final Grade 8.3
- Thesis supervised by Dr. Roberto Valenti

Scholarships & Awards

Enrichment Student   *London, United Kingdom*
The Alan Turing Institute   *Oct 2021 – Jun 2022*

Research Scholarship
DALI project, funded by European Research Council Grant 695662   *Oct 2019 - Sep 2021*

Master's Thesis Research Funding
Facebook ParlAI Research Award   *Jan 2018 - Jul 2018*

Full Undergraduate and Graduate Scholarships
Evangelische Studienstiftung Villigst e.V.   *Feb 2013 - Jul 2018*

Publications

Polysemy - Evidence from Linguistics, Behavioural Science and Contextualised Language Models.
Janosch Haber and Massimo Poesio.
Computational Linguistics. In Press

Computational Models of Anaphora.
Massimo Poesio, Juntao Yu, Silviu Paun, Abdulrahman Aloraini, Pengcheng Lu, Janosch Haber, and Derya Cokal.
Annual Review of Linguistics. 2023

Word Sense Distance and Similarity Patterns in Regular Polysemy.
Janosch Haber.
PhD Thesis. Queen Mary University of London. 2022

Patterns of Lexical Ambiguity in Contextualised Language Models.
Janosch Haber and Massimo Poesio.
Findings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021

Word Sense Distance in Human Similarity Judgements and Contextualised Word Embeddings.
Janosch Haber and Massimo Poesio.
Proceedings of the Conference on Probability and Meaning (PaM2020) – SIGSEM Workshop. 2020

Assessing Polyseme Sense Similarity through Co-predication Acceptability and Contextualised Embedding Distance.
Janosch Haber and Massimo Poesio.
Proceedings of the 9th Joint Conference on Lexical and Computational Semantics (*SEM). 2020

Classification of low-agreement pronouns through collaborative dialogue: A proof of concept.
Janosch Haber and Massimo Poesio.
Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue (SemDial). 2020

The PhotoBook Dataset: Building Common Ground through Visually Grounded Dialogue.
Janosch Haber, Tim Baumgärtner, Ece Takmaz, Lieke Gelderloos, Elia Bruni, and Raquel Fernández.
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). 2019

How should we call it? - Introducing the PhotoBook Conversation Task and Dataset for Training Natural Referring Expression Generation in Artificial Dialogue Agents.
Janosch Haber.
Master's Thesis. University of Amsterdam. 2018