David Wadden

Curriculum Vitae

dave.wadden@gmail.com
dwadden.github.io
Google Scholar
Semantic Scholar

Research Interests

My research aims to reliably adapt foundation models for high-impact applications, particularly in science and health. I'm currently interested in creating instruction-following datasets and benchmarks for specialized domains, developing methods to improve the factual accuracy and attributability of model outputs, and understanding the link between training data and model capabilities.

Education

2016-2023 **PhD Computer Science**, *University of Washington*, Seattle
Advisor: Hannaneh Hajishirzi

2006–2010 **BA Physics**, *Amherst College*, Amherst
Phi Beta Kappa, With Distinction

Professional Experience and Internships

Since 2023 **Research Scientist**, Allen Institute for AI, Seattle, WA
AllenNLP and Semantic Scholar teams.

Summer 2021 **Research Intern**, Google, Seattle, WA
Hosts: Nikita Gupta, Kenton Lee, and Kristina Toutanova.

Fall 2019 - Spring 2020 **Research Intern**, Allen Institute for AI, Seattle, WA
Hosts: Kyle Lo and Lucy Lu Wang.

2011-2014 **Associate Computational Biologist**, The Broad Institute, Cambridge, MA
Supervisor: Aravind Subramanian.

Publications

Conference

2024 Yanai Elazar*, Jiayao Zhang*, David Wadden*, Bo Zhang, Noah A. Smith. **Estimating the Causal Effect of Early ArXiving on Paper Acceptance.** *CLeaR (Causal Learning and Reasoning) 2024*. *Denotes equal contribution

2024 Jian Guan, Jesse Dodge, David Wadden, Minlie Huang, Hao Peng. **Language Models Hallucinate, but May Excel at Fact Verification.** *NAACL 2024*.

2024 Orion Weller, Kyle Lo, David Wadden, Dawn J Lawrie, Benjamin Van Durme, Arman Cohan, Luca Soldaini. **When do Generative Query and Document Expansions Fail? A Comprehensive Study Across Methods, Retrievers, and Datasets.** *EACL Findings 2024*.

2023 Yizhong Wang, Hamish Ivison, Pradeep Dasigi, Jack Hessel, Tushar Khot, Khyathi Chandu, David Wadden, Kelsey MacMillan, Noah A. Smith, Iz Beltagy, Hannaneh Hajishirzi. **How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources.** *NeurIPS (Datasets and Benchmarks) 2023*.

2023 Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, David Wadden, Khendra G. Lucas, Adam S. Miner, Theresa Nguyen, Tim Althoff. **Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction.** *ACL 2023*.
🏆 **Outstanding paper**.

2022 David Wadden, Nikita Gupta, Kenton Lee, Kristina Toutanova. **Entity-centric query refinement.** *AKBC 2022*.
🏆 **Best Paper Honorable Mention**.

2022 David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Iz Beltagy, Lucy Lu Wang, Hannaneh Hajishirzi. **SciFact-Open: Towards open-domain scientific claim verification.** *EMNLP Findings 2022*.

2022 David Wadden, Kyle Lo, Lucy Lu Wang, Arman Cohan, Iz Beltagy, Hannaneh Hajishirzi. **MultiVerS: Improving scientific claim verification with weak supervision and full-document context.** *NAACL Findings 2022*.

2022 Dustin Wright, David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Isabelle Augenstein, Lucy Lu Wang. **Generating Scientific Claims for Zero-Shot Scientific Fact Checking.** *ACL 2022*.

2021 David Wadden, Tal August, Qisheng Li, Tim Althoff. **The Effect of Moderation on Online Mental Health Conversations.** *ICWSM 2021*.
🏆 **Best Paper for Outstanding Study Design**.

2021 Rahul Nadkarni, David Wadden, Iz Beltagy, Noah A. Smith, Hannaneh Hajishirzi, Tom Hope. **Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study.** *AKBC 2021*.

2021 Aida Amini, Tom Hope, David Wadden, Madeleine van Zuylen, Eric Horvitz, Roy Schwartz, Hannaneh Hajishirzi. **Extracting a Knowledge Base of Mechanisms from COVID-19 Papers.** *NAACL 2021*.

2020 David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, Hannaneh Hajishirzi. **Fact or Fiction: Verifying Scientific Claims.** *EMNLP 2020*.

2019 David Wadden, Ulme Wennberg, Yi Luan, Hannaneh Hajishirzi. **Entity, Relation, and Event Extraction with Contextualized Span Representations.** *EMNLP 2019*.

2019 Yi Luan, David Wadden, Luheng He, Hannaneh Hajishirzi, Mari Ostendorf. **A General Framework for Information Extraction using Dynamic Span Graphs.** *NAACL 2019*.

Preprints

2024 Fangyuan Xu, Kyle Lo, Luca Soldaini, Bailey Kuehl, Eunsol Choi, David Wadden. **KIWI: A Dataset of Knowledge-Intensive Writing Instructions for Answering Research Questions**.

2024 Hamish Ivison, Yizhong Wang, Valentina Pyatkin, Nathan Lambert, Matthew Peters, Pradeep Dasigi, Joel Jang, David Wadden, Noah A. Smith, Iz Beltagy, Hannaneh Hajishirzi. **Camels in a Changing Climate: Enhancing LM Adaptation with Tulu 2**.

Workshop

2021 David Wadden, Kyle Lo. **Overview and Insights from the SciVer Shared Task on Scientific Claim Verification.** *Scholarly Document Processing Workshop @ NAACL 2021*.

Journal

2019 The ALS Stratification Consortium. **Stratification of amyotrophic lateral sclerosis patients: a crowdsourcing approach**. *Scientific Reports* 9, Article number: 690 (2019).

2018 Oana M Enache, David L Lahr, Ted E Natoli, Lev Litichevskiy, David Wadden, Corey Flynn, Joshua Z Gould, Jacob K Asiedu, Rajiv Narayan, Aravind Subramanian. **The GCTx format and cmap{Py, R, M, J} packages: resources for the optimized storage and integrated traversal of dense matrices of annotated dense matrices**. *Bioinformatics*, bty784.

2017 Aravind Subramanian et al. (David Wadden author #17 of 52). **A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles**. *Cell* 171(6):1437-1452.e17, 2017.

2017 Ian Smith, Peyton Greenside, Ted Natoli, David L. Lahr, David Wadden, Itay Tirosh, Rajiv Narayan, David E. Root, Todd R. Golub, Aravind Subramanian, John G. Doench. **Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map**. *PLoS Biol* 15(11):e2003213.

Honors and Awards

2023 Outstanding Paper, ACL 2023. For *Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction*.

2022 Best Paper Honorable Mention, AKBC 2022. For *Entity-Centric Query Refinement*.

2021 Best Paper Award for Outstanding Study Design, ICWSM 2021. For *The Effect of Moderation on Online Mental Health Conversations*.

2010 William Warren Stifler Prize in Physics, Amherst College.

2007 Bassett Physics Prize, Amherst College.

Open-source software

DyGIE++ (main author): Multitask system for entity, relation, and event extraction (550+ GitHub ★).

SciFact (main author): Dataset and modeling baselines for the SciFact scientific claim verification dataset (200+ GitHub ★).

Open-Instruct (contributor): Open instruction-following language models. (900+ GitHub ★).

Service

Professional service

2021 Organizer, SciVer shared task. Hosted at SDP workshop @ NAACL 2021.

Reviewer: EMNLP, ACL, NAACL, ACL Rolling Review, Journal of Biomedical Informatics.

Departmental service

2019, 2020 PhD admissions committee member.

2020 Prospective PhD student committee member.

Mentoring

Win-Spr '20 Shanchuan Peter Lin, UW CSE MS student → Software engineer, Microsoft.

Win-Spr, '19 Ulme Wennberg, UW CSE MS exchange student → PhD student, KTH Royal Institute of Technology.

Technical skills

Programming languages: Python, JavaScript, R, C, Matlab.

Deep learning frameworks: Tensorflow, Jax, PyTorch, PyTorch Lightning, AllenNLP, Hugging Face Transformers.

Data science packages: Pandas, Numpy, Scikit-Learn, Matplotlib.

Community engagement

2019-Present **Mentor**, Minds Matter Seattle
Weekly tutoring and college advising for high school students from underrepresented backgrounds.