Madeleine N. van Zuylen

madeleine.vanzuylen@gmail.com | 206-915-8875 | Seattle, WA
LinkedIn, Google Scholar, GitHub, Personal Website

SUMMARY

I have extensive experience delivering model analysis and developing novel datasets for biomedical NLP research in 13+ publications, including in responsible and accessible tech, and internship experience in backend development building data pipelines. I have a strong bias towards action to deliver strong and clear results for production or publication. I'm curious, creative, and collaborative, and look forward to working on complex problems.

EDUCATION

Northeastern University May 2023
MS in Computer Science

- **Courses:** Intensive Foundations of CS, Object-Oriented Design, Data Structures, Algorithms and Applications in Computer Systems, Algorithms, Database Management Systems, Machine Learning, Deep Learning, NLP, Web Development, Scalable Distributed Systems

University of Notre Dame May 2017
BS in Applied Computational Mathematics and Statistics and Biochemistry

- **Courses:** Calculus III, Linear Algebra, Differential Equations, Intro to Probability, Numerical Analysis, Statistical Methods and Data Analysis, Mathematical and Computational Modeling, Discrete Structures

EMPLOYMENT

Software Development Engineering Intern | Amazon Web Services | File Storage Gateway
September 2022 - December 2022

- Automated core dumps and restart of a gateway processes when gateways became unavailable
- Provided insight into unresponsive gateways by designing metrics to track gateway availability

Software Development Engineering Intern | Redfin | Listing Ingestion Platform
May 2022 - August 2022

- Built Samza app to store house listing records in AWS S3 with versioning to provide insight into historical changes in the raw records for 665,000 listings, 62,000 agents, and 14,000 Brokers
- Added 13,500 listings from West Alabama Multiple Listing Service onto redfin.com representing an .8% increase in Redfin's total coverage

Data Science Analyst II | Allen Institute for AI | Semantic Scholar
June 2017 - May 2022

- Curated, managed, and analyzed large novel training datasets for **13 published machine learning research papers** including verifying scientific claims, PDF accessibility, and document summarization
  - Evaluated model performance and compared to benchmarks
- Analyzed and iterated on a **field of study classifier** to label scientific papers by domain
- Designed and **built annotation tasks** launched to crowdsource workers

SKILLS

- **Technical:** Java, Python, Git, AWS, S3, EC2, Lambda, SQL, MySQL Databases, MongoDB, Pandas, NumPy, Scikit-Learn, PyTorch, TensorFlow, Keras, Terraform, Atlantis, Samza, Grafana, Mockito, RabbitMQ

PUBLICATIONS

The Semantic Reader Project: Augmenting Scholarly Documents through AI-Powered Interactive Reading Interfaces
Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg…**Madeleine van Zuylen**, Lucy Lu Wang, Christopher Wilhelm, Caroline M Wu, Jiangjiang Yang, Angele Zamarron, Marti A. Hearst, Daniel S. Weld
ArXiv, 2023, 2 citations

The Semantic Scholar Open Data Platform
Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg…**Madeleine van Zuylen**, Lucy Lu Wang, Christopher Wilhelm, Caroline M Wu, Jiangjiang Yang, Angele Zamarron, Marti A. Hearst, Daniel S. Weld
ArXiv, 2023, 8 citations

MSˆ2: Multi-Document Summarization of Medical Studies
Jay DeYoung, Iz Beltagy, **Madeleine van Zuylen**, Bailey Kuehl, Lucy Lu Wang
EMNLP, 2021, 58 citations

Extracting a Knowledge Base of Mechanisms from COVID-19 Papers
Tom Hope, Aida Amini, David Wadden, **Madeleine van Zuylen**, Sravanthi Parasa, Eric Horvitz, Daniel Weld, Roy Schwartz, Hannaneh Hajishirzi
NAACL, Human Language Technologies, 2021, *17 citations*

SciA11y: Converting Scientific Papers to Accessible HTML
Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie (Yu-Yen) Cheng, Chelsea Hess Haupt, Matt Latzke, Bailey Kuehl, **Madeleine van Zuylen**, Linda M. Wagner, Daniel S. Weld
ASSETS 2021 Posters and Demonstrations, Artifact Award 1st Place

MedICaT: A Dataset of Medical Images, Captions, and Textual References
Sanjay Subramanian, Lucy Lu Wang, Sachin Mehta, Ben Bogin, **Madeleine van Zuylen**, S. Parasa, Sameer Singh, Matt Gardner, Hannaneh Hajishirzi
EMNLP, 2020, *19 citations*

SCIREX: A Challenge Dataset for Document-Level Information Extraction
Sarthak Jain, **Madeleine van Zuylen**, Hannaneh Hajishirzi, Iz Beltagy
ACL, 2020, *95 citations*

Quantifying Sex Bias in Clinical Studies at Scale With Automated Data Extraction
Sergey Feldman, Waleed Ammar, Kyle Lo, Elly Trepman, **Madeleine van Zuylen**, Oren Etzioni
JAMA Network Open, 2019, *119 citations*

Structural Scaffolds for Citation Intent Classification in Scientific Publications
Arman Cohan, Waleed Ammar, **Madeleine van Zuylen**, Field Cady
NAACL, Human Language Technologies, 2019, *172 citations*

Construction of the Literature Graph in Semantic Scholar
Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, Rodney Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler Murray, Hsu-Han Ooi, Matthew Peters, Joanna Power, Sam Skjonsberg, Lucy Lu Wang, Chris Wilhelm, Zheng Yuan, **Madeleine van Zuylen**, and Oren Etzioni
NAACL, Human Language Technologies, 2018, *370 citations*

A Dataset of Peer Reviews (PeerReaD): Collection, Insights and NLP Applications
Dongyeop Kang, Waleed Ammar, Bhavana Dalvi Mishra, **Madeleine van Zuylen**, Sebastian Kohlmeier, Eduard Hovy, Roy Schwartz
NAACL, Human Language Technologies, 2018, *156 citations*