

Madeleine N. van Zuylen

madeleine.vanzuylen@gmail.com | 1-206-915-8875 | Seattle, WA

[LinkedIn](#), [Google Scholar](#), [GitHub](#), [Personal Website](#)

EDUCATION

Northeastern University | Boston, MA
MS in Computer Science

May 2023

- **Courses:** Intensive Foundations of CS (Python), Object-Oriented Design (Java), Data Str, Algo & App in Cmp Sys (C), Algorithms (Java), Database Management Sys (R, SQL), Machine Learning (Python), Deep Learning, NLP, Web Development, Scalable Distributed Systems

University of Notre Dame | South Bend, IN
BS in Applied Computational Mathematics and Statistics and Biochemistry

May 2017

- **Courses:** Discrete Structures, Calc III, Linear Algebra, Differential Equations, Intro to Probability, Numerical Analysis, Stat Methods & Data Analysis, Mathematical/Comp Modeling

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 large novel training datasets for **13 published machine learning research papers** including verifying scientific claims, citation intent, bias in medical studies, information extraction, PDF accessibility, and document summarization
- 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, Scikit-Learn, PyTorch, TensorFlow, Terraform, Atlantis, R, Samza, Grafana, Mockito, Kafka, MongoDB

PUBLICATIONS

[The Semantic Reader Project: Augmenting Scholarly Documents through AI-Powered Interactive Reading Interfaces*](#)

ArXiv, 2023, 2 citations

[The Semantic Scholar Open Data Platform](#)

ArXiv, 2023, 8 citations

[MS²: Multi-Document Summarization of Medical Studies](#)

EMNLP, 2021, 58 citations

[Extracting a Knowledge Base of Mechanisms from COVID-19 Papers](#)

NAACL, Human Language Technologies, 2021, 17 citations

[Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific PDF accessibility for blind and low vision users](#)

ArXiv, 2021, 13 citations

[SciA11y: Converting Scientific Papers to Accessible HTML](#)

ASSETS 2021 Posters and Demonstrations, Artifact Award 1st Place

[MediCaT: A Dataset of Medical Images, Captions, and Textual References](#)

EMNLP, 2020, 19 citations

[SCIREX: A Challenge Dataset for Document-Level Information Extraction](#)

ACL, 2020, 95 citations

[Fact or Fiction: Verifying Scientific Claims](#)

EMNLP, 2020, 231 citations

[Quantifying Sex Bias in Clinical Studies at Scale With Automated Data Extraction](#)

JAMA Network Open, 2019, 119 citations

[Structural Scaffolds for Citation Intent Classification in Scientific Publications](#)

NAACL, Human Language Technologies, 2019, 172 citations

[Construction of the Literature Graph in Semantic Scholar](#)

NAACL, Human Language Technologies, 2018, 370 citations

[A Dataset of Peer Reviews \(PeerRead\): Collection, Insights and NLP Applications](#)

NAACL, Human Language Technologies, 2018, 156 citations

VOLUNTEER ACTIVITIES

King County Family Law Court Appointed Special Advocate (FL CASA)

January 2020 - December 2020

- Gathered information in high conflict child custody cases and reported to the court recommendations for a parenting plan
- Interviewed parents, collaterals, and child, toured the home, wrote two reports, and testified in court