

Jason Portenoy

jason.portenoy@gmail.com
jasport.org

Experienced data engineer/data scientist with proficiency in full-stack development, databases, machine learning, and recommender systems. Looking to support a team doing meaningful and impactful work.

Work Experience

Allen Institute for AI (AI2)

Seattle, WA

Visiting Scientist / Full Stack Developer — Semantic Scholar Team

March 2020 – Present

- Built AI-enabled scholarly recommendation systems and wrote research publications.
- Using Pytorch and additional Python libraries, I fine-tuned a pretrained BERT language model to encode document similarity, which I used to create embeddings for 3.5 million documents.
- Built a custom React+D3 web app to provide recommendations and collect data for user studies.
- Currently, I am employed part-time deploying my prototype recommender system in production.

University of Washington

Seattle, WA

Data Engineer — Center for an Informed Public (CIP)

March 2022 – Sept 2022

- Managed and maintained the data infrastructure supporting the research and operations of the CIP, an interdisciplinary center for research and action around misinformation and disinformation.
- Responsibilities included oversight and management of a collection of physical and virtual servers, and pipelines and tools for data collection, storage, and analysis.
- Applied updates and security patches, built and maintained containers (Docker and Kubernetes) for various tasks, managed a Kubernetes cluster to run the Center's computing operations
- Set up and oversaw multiple databases (Postgres, Airflow) with billions of rows and dozens of terabytes of data.

Research Analyst / Full Stack Developer

Sept 2021 – March 2022

- With the Gordon and Betty Moore Foundation, developed grantexplorer.org—a tool to explore federal research grants over time—using React, D3, and FastAPI for interactive visualizations, and Elasticsearch and Gensim language models to intelligently assist with keyword queries.

Chan Zuckerberg Initiative

Redwood City, CA

Visiting Researcher

May 2019 – Sept 2019

- Analyzed networks of biomedical researchers to improve suggestions of new papers in feeds of research publications

Microsoft Research and AI

Seattle, WA

Summer Research Intern (Mentors: Paul Bennett, Ryen White, Eric Horvitz)

March 2022 – Sept 2022

- Developed methods to identify when people make commitments in their written communications, and understand what kind of commitments they make. Co-authored a patent based on this work.
- Used email data from Microsoft's internal terabyte-scale Cosmos data store to build ML classifiers.

Skills

- Python, Pandas, JavaScript, React, Typescript, R, SQL, D3, HTML, CSS, Linux, Git, Django/Flask, SciPy
- Machine Learning and AI, Databases, APIs, Network Analysis, Visualization, High Performance Computing

Education

University of Washington, Information School

Seattle, WA

PhD, MS Information Science

2013 – 2021

- [Dissertation](#): "Harnessing Scholarly Literature as Data to Curate, Explore, and Evaluate Scientific Research"
- Committee: Jevin D. West (Chair), Emma Spiro, Bill Howe, Benjamin Mako Hill
- Research projects with National Academy of Sciences, JSTOR, Pew Charitable Trusts, others
- TA Experience: Adv. Data Science Methods, Design Thinking, Info. Sys. Analysis & Management, Web Dev.

Brown University

Providence, RI

BS, Neuroscience

2004 – 2008

For publications and additional info, see [full CV](#)