Jason Portenoy

jason.portenoy@gmail.com
jasport.org

> ***NYC-based data engineer/data scientist PhD with 2 years industry experience / 10+ years experience total. Proficiency in Python development, Jupyter notebooks, SQL, full-stack development, and machine learning and recommender systems. Looking to support a team doing meaningful and impactful work.***

Work Experience

OurResearch   *Fully Remote*
Senior Data Engineer — OpenAlex   Feb 2023 -- Sept 2024
- OpenAlex is the world's largest free and open Scientific Knowledge Graph (SKG), created by the nonprofit OurResearch. Representing the entire global research ecosystem, it comprises more than 250 million scholarly works, billions of links between them, and terabytes of linked metadata. My primary role was to understand the strengths and limitations of the data, and to help our users do the same.
- As a key team member of this six-person startup, I had an active role in most operations: maintenance and development of a variety of databases (Postgres, Elasticsearch, Redshift, Databricks/Spark); backend dev and data pipeline/quality improvement (Python, SQLAlchemy, Flask); writing technical and general documentation; working with the press to provide and analyze data; engaging with community through webinars, virtual and in-person presentations, social media, online forums, customer support tickets, and open houses; monitoring infrastructure and database operations; and tracking KPIs.

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.
- 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.

Allen Institute for AI (AI2)   *Seattle, WA*
Visiting Scientist / Full Stack Developer — Semantic Scholar Team   March 2020 -- July 2023
- 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.

Chan Zuckerberg Initiative — Visiting Researcher — 2019

Microsoft Research and AI — Summer Research Intern — 2017

Skills

Python, Pandas, Jupyter notebooks, Spark/Databricks, React, Vue, Typescript, SQL, D3, R, Docker, Kubernetes 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, Information Science   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 — BS, Neuroscience   2008

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