

Jason Portenoy

jason.portenoy@gmail.com

jasport.org

NYC-based data engineer/data scientist PhD with 10 years experience across industry and academia. 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

DataCite

Remote

Data Analyst

Nov 2024 – Present

- Data cleaning, analysis, expansion, and enhancement for the Make Data Count: A Central Corpus for All Data Citations project. (contract position)

OurResearch

Remote

Senior Data Engineer – OpenAlex

Feb 2023 – Sept 2024

- Development, maintenance, data science, and community outreach for OpenAlex, the world's largest free and open Scientific Knowledge Graph, with over 250 million scholarly works and terabytes of linked metadata.
- Collaborated in a small team to monitor, improve, and make strategic development decisions for cloud database and software infrastructure (Postgres, Elasticsearch, AWS Redshift, Databricks/Spark), and optimize backend systems and data pipelines using Python/SQLAlchemy.
- Engaged with external stakeholders and the research community to promote OpenAlex's mission and provide technical support through webinars, social media, and in-person events.

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, FastAPI, Elasticsearch, and Gensim language models.

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—using PyTorch to create embeddings for 3.5 million documents—and wrote research publications.
- 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](#)