

John Freeman

John.C.Freeman17@gmail.com | jcf1.github.io

WORK EXPERIENCE

Consultant (Data Scientist)

Berkeley Research Group

Jan. 2022 – Present

Washington, DC (Remote)

- Implemented Retrieval-Augmented Generation (RAG) pipelines and custom agents that leverage LLMs with a vector database and/or SQL database to answer queries about an information source. The program can both answer questions about the content of the data and provide the context it used for its response to allow the user to verify that the answer was correct. I have also written programs that use the same techniques to draft sections or whole documents using the examples provided to it.
- Performed privileged modeling to help lawyers review documents in the e-Discovery phase to determine which ones should be withheld from production. I was responsible for cleaning the data, generating features from the text and meta-data, and training models to produce privilege scores that reduced the number of documents our clients had to manually review.
- Wrote and deployed software to completely automate previously manual processes for various clients. I was responsible for the implementation of business logic, the creation and format of our results, quality control throughout development, and support for issues the client had after deployment.
- Worked with many practices within BRG on cleaning and parsing various types of data using python and OCR. I was solely responsible for writing the code, processing the data, QC'ing the output, and providing a finished result in a timely matter. The types of documents I processed included PDFs, messy output from a client's custom program, and unconventional spreadsheets.

Senior Associate (Data Scientist)

Berkeley Research Group

Nov. 2019 – Dec. 2021

Chicago, IL (Remote)

- Responsible for combining multiple datasets and generating metrics for a settlement model that provided summary valuation estimates for a large class action lawsuit (~35,000 observations). Our model helped our client provide offers while negotiating settlements with multiple plaintiff law firms.
- Created a process for a client to help them better track their business development efforts since their company's projected revenue was not matching their actual revenue. Our findings led our client to completely redesign their internal business software and update their projected revenue.
- Created a model to detect whether an opioid transaction from a drug distributor to a pharmacy should be flagged as potentially fraudulent. Our model found a significant number of additional transactions that should have been flagged as potentially fraudulent in their dataset.

Junior Data Scientist

Vista Analytics (acquired by Berkeley Research Group in Nov. 2019)

Sep. 2017 – Oct. 2019

Washington, DC

- Designed and maintained an end-to-end entity pipeline that autonomously extracts, ranks, and assigns types to significant entities in a large document set. The results were used in dashboards that we provided to clients performing document review to help them find related documents and discover new insights in their data.
- Wrote a program to download and analyze thousands of 10-K and 10-Q SEC to predict if the stock price of a company will increase or decrease in the near-future. We gave our results to an investment firm client for back testing which resulted in a significant increase in their performance using our analysis.

Teaching Assistant

Williams College

Sept. 2015 – May 2017

Williamstown, MA

- Helped monitor lab sessions, grade problem sets, and hold TA Sessions to assist students with homework assignments.

EDUCATION

Williams College

Bachelor of Arts in Computer Science

Williamstown, MA

Sept. 2013 – Jun. 2017

TECHNICAL SKILLS

Languages: Python, Java, Typescript, JavaScript, C/C++, Scala, SQL, HTML/CSS

Frameworks: React, React Native, Svelte, Node.js

Developer Tools: Git, AWS, VS Code

Certifications: Large Language Model Ops (LLM Ops) Certificate (AI Makerspace, 2023)