

A passionate, curious-minded and hardworking ML Engineer / Data Scientist. Likes solving big, messy problems with big, messy data. Excited about bringing good ideas from representation learning over text, images and graphs to production scale.


Currently a second year MS student at NYU's Center for Data Science

(May '23 Exp.)


SKILLS

- **Languages** - Python (10 yrs), SQL (6 yrs), Java (2 yrs) Scala (1 yr), C++ (<1 yr)
- **ML/Data Science** - PyTorch, TensorFlow, NumPy, Pandas, DGL, huggingface
- **Data Engineering** - Beam, Dataflow, Airflow, Luigi, Spark
- **Cloud** - Experience with declarative infra and CI/CD on AWS and GCP.


EXPERIENCE

 **Spotify** New York, NY
Data Scientist I (Part-Time) September 2022 - December 2022


- Delivered usage analytics and case-studies for partners in publishing, licensing and editorial.

 **Zillow** New York, NY
Capstone Researcher (Internship) September 2022 - December 2022


- Evaluated methods for scalable inductive representation learning over large, sparsely connected multi-modal graph data.
- Wrote training, evaluation and experiment tracking pipeline using Deep Graph Library, Ray and PyTorch Lightning.

 **Spotify** New York, NY
Data Science / Machine Learning Engineering Intern June 2022 - September 2022

- Wrote featurization code to perform text vectorization and measure cosine-similarity on 30M+ daily records using TensorFlow/Beam/Luigi.
- Assessed error rates on lyric content across languages and markets. Designed and oversaw annotation trial to evaluate quality control approaches.

 **KPMG Data & Analytics** New York, NY
Associate Data Scientist January 2019 - August 2021

- *Health Insurance Client*: Built data pipelines to compute cohort statistics for demographic subsets of claims data and provide visualizations for analyst-facing dashboards.
- *Financial Services Client*: Data Engineer on greenfield credit card offering at major NYC financial institution. Worked on small, agile team to rapidly build and scale fault-resistant data processing pipeline capable of handling 10+ million daily transactions using Airflow, PySpark and BigTable.
- *GDPR NLP*: Leveraged RoBERTa embeddings to improve recall on semantic search over regulatory documentation relating to the EU's General Data Protection Regulation laws.

 **KPMG Cognitive Automation Labs** New York, NY
Summer Intern June 2018 - August 2018

- Built feature generation pipeline to perform cleaning, tokenizing and part-of-speech tagging on text data.

EDUCATION

NYU Center for Data Science 2021 - 2023
M.S. Data Science (Exp.)

NYU Courant Institute of Mathematical Sciences 2014 - 2018
B.A. Computer Science

NYU Leonard Stern School of Business 2014 - 2018
B.S. Statistics

INTERESTS

- Multi-instrumentalist and full-stack musician (writer/performer/producer) _rothie_