David S. Roth

davidroth.me ℹ
davidroth6@gmail.com ✉
(201) 615-3785 ☎

A hardworking and curious-minded ML Engineer. Passionate about designing and building scalable machine learning systems to solve challenging, high-impact problems.

EDUCATION

NYU Center for Data Science, *M.S. Data Science*   *2021 - 2023*
Deep Learning, Computational Linear Algebra, Inference and Representation, Reinforcement Learning, Natural Language Understanding, Causal Inference, Responsible Data Science

NYU Courant Institute of Mathematical Sciences, *B.A. Computer Science*   *2014 - 2018*
Algorithms & Data Structures, Discrete Math, Linear Algebra, Computer Systems Organization, Large Scale Web Applications, Computer Vision, High Performance Machine Learning

NYU Leonard Stern School of Business, *B.S. Business / Statistics*   *2014 - 2018*
Probability & Statistics, Linear Regression and Forecasting Models, Regression and Multivariate Data Analysis, Stochastic Processes, Theory of Probability

EXPERIENCE

⟳ Edgi Learning   *March 2023 - Present*
NLP Engineer

Principal engineer on Edgi Librarian– a comprehensive learning, content recommendation and assessment chatbot powered by LLMs. Responsibilities include training data curation, model fine-tuning, prompt engineering, prompt chaining and managing vector database content search.

⌂ Zillow   *September 2022 - December 2022*
Research Intern

Evaluated methods for scalable model inference on multi-modal graph data. Wrote training, evaluation and inference pipelines to compute dense node representations using message passing and gated graph convolutional networks to predict listing attributes and node links over graphs containing $1M+$ nodes.

● Spotify   *June 2022 - September 2022*
Machine Learning Engineering / Data Science Intern

Built inference pipeline for entity disambiguation pipeline used to match recordings to registered musical works with Luigi, Tensorflow and Apache Beam. Delivered pipeline capable of scoring $\sim30M$ potential match pairs on a 2 hour SLA.

Assessed error rates on lyric content across languages and markets. Designed and oversaw annotation trial to evaluate quality control approaches, resulting in an 8% reduction of erroneous lyric assignments on instrumental tracks.

KPMG KPMG Data & Analytics   *January 2019 - August 2021*
Data Scientist, NLP

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 to handle $20+$ million daily transaction records using Airflow and Spark. Worked closely with anomaly detection and fraud teams to develop and deploy models to detect fraudulent transactions.

Contributed module to Cognitive Contract Modeling library utilizing FinBERT text embeddings to perform semantic search over regulatory documentation. Improved recall over previous model by 10%.

KPMG KPMG Cognitive Automation Labs   *June 2018 - August 2018*
Machine Learning Engineering Intern

Fine-tuned passage-level topic classification network using privacy documentation and features from pre-trained eLMo Bi-LSTM to surface potential privacy risks in client contracts.

SKILLS

- Python (10 yrs), SQL (8 yrs), Rust (1 yr), Scala (1 yr)
- PyTorch, TensorFlow, DeepSpeed, Ray, Deep Graph Library, huggingface
- Spark, Beam, Airflow, Luigi
- Google Cloud Platform

INTERESTS

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