David S. Roth

davidroth.me 🛈
davidsamuelroth@gmail.com ✉
(201) 615-3785 📞

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

EDUCATION

NYU Center for Data Science, *M.S. Data Science*   *2021 - 2023*
Computational Linear Algebra, Inference and Representation, Deep Learning, Deep Reinforcement Learning, Deep Representation 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**   *February 2023 - August 2023*
NLP Engineer

Principal engineer on Edgi Librarian, an LLM-based tutoring, educational content recommendation and learning assessment chatbot. Responsibilities include conversation state management, training data curation, prompt engineering and implementing vector database content search.

⌂ **Zillow**   *September 2022 - December 2022*
Research Fellow

Built training and inference pipelines for multi-modal (image+text) graph neural networks using PyTorch and Deep Graph Library. Utilized message-passing and learned gating functions to compute dense node representations for property listings and predict attributes and node links over graphs containing 1M+ nodes.

● **Spotify**   *June 2022 - December 2022*
Machine Learning Engineering Intern

Built inference pipeline for entity disambiguation pipeline used to match recordings to registered musical works with Beam and Tensorflow. Delivered pipeline capable of scoring ∼500M match pairs daily.

Delivered usage analytics and case studies to partners in publishing, licensing and editorial to improve representation for smaller artists in recommendations and curated playlists.

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

Built data processing pipelines greenfield credit card offering at major NYC financial institution using Airflow, Amazon SQS and PySpark. Worked on small, agile team to rapidly build and scale fault-resistant systems to handle 20+ million daily transaction records. 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 15%.

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 language model to surface potential regulatory exposure in client contracts.

SKILLS

- Python (10 yrs), SQL (8 yrs), Scala (3 yr), Rust (1 yr)
- PyTorch, TensorFlow, ONNX, Neo4j
- Spark, Beam, Airflow, Dataflow
- Google Cloud Platform, AWS

INTERESTS

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