

Roni Kobrosly Ph.D.

☎ 1-512-965-0601 | ✉ roni.kobrosly@gmail.com | 🌐 www.ronikobrosly.github.io | 📷 [ronikobrosly](#) | 📺 [ronikobrosly](#)

Summary

Data org leader with proven success unifying analytics, ML, generative AI, and data engineering to create business impact. Excellent communication skills (multiple-time conference speaker), strategic thinking, with the credibility of strong technical skills (particularly around causality and GenAI with unstructured text data), and open-source contribution. 9+ years of people management and mentoring in cross-functional teams. Record of success in growing data teams and ICs, with extensive experience sourcing, interviewing, and hiring effective data practitioners.

Relevant Work Experience

Capital One

Washington, D.C.

DIRECTOR OF DATA SCIENCE FOR ENTERPRISE OBSERVABILITY

March 2023 - Present

- Led enterprise-wide ML-based anomaly detection efforts resulting in early detection of dozens of business-critical breakages.
- Prototyped and ultimately delivered on enterprise-wide forecasting of SLIs and error budgets.
- Delivered prototype of novel, causality-based, algorithm for determining root cause of cascading failures in connected microservices.

DrFirst

Washington, D.C.

HEAD OF DATA SCIENCE (DIRECTOR)

October 2020 - March 2023

- Built data evaluation pipeline around company's prescription messaging data product, which brought in \$25 million from clients since 2020.
- Data Science team prototyped and developed the 2022 Edison Award-winning SmartSuite for Pharmacy, a ML-powered API service for parsing pharmacy medication instructions, which receives millions of API requests a day.
- Planned, built, and maintaining several NLP data products that have generated > \$1M in revenue since 2020. Employed BERT and LSTMs for text classification, OCR, and named entity recognition.
- Tripled team size and led a multifunctional team of data scientists, data analysts, and python engineers.
- Architected company's first cloud-based, ML, CI/CD platform, resulting in dramatically faster ML deployments.

Dia and Co

Austin, Texas

SENIOR MACHINE LEARNING ENGINEER

Oct 2018 - February 2020

- Led discussions on long-term project to improve warehouse operations, which increased available inventory by 9 percent. Implemented an optimization algorithm and deployed to Amazon AWS.
- Planned and implemented Dia's first ML-powered multi-channel attribution framework.

Charles Schwab

Austin, Texas

SENIOR MANAGER, DATA SCIENTIST

May 2016 - Sept 2018

- Identified previously undetected malicious and bot web client web sessions using Spark MLlib's Doc2Vec and k-means clustering. Created big data pipeline to analyze 10TB batches of server log data using Spark.
- Replaced \$150k/year Ipsos social media brand health reporting with an in-house tool created in Python. Obtained data through Twitter and Reddit APIs and custom-made news scraping pipeline. Employed topic modeling, sentiment analysis, and anomaly detection algorithms.
- Completed other high-impact projects with HR, security, customer support, and client experience teams. Led engagement for each project, including helping shape project scope, coordinating directly with stakeholders, presenting results, and iterating.

Insight Data Science

New York, New York

PROGRAM DIRECTOR FOR DATA SCIENCE

June 2015 - May 2016

- Advised and scoped impactful data science projects for early-stage startups in a number of industries.
- Provided strategic, product, and hands-on technical advising for 60+ data scientists.
- Conducted behavioral and technical interviews of nearly 100 data scientists.

Technical Skills

Open-Source Contributor: Creator and maintainer of [causal-curve \(clickable link\)](#), a widely-used python package.

Languages: Python, R, Go, SQL (ANSI, Snowflake, HiveQL, etc.), HTML, Scala.

GenAI: Huggingface, LangChain, chain-of-thought prompting, text similarity search, summarization, RAG chunking best practices.

Data: Classical statistics, classical machine learning, deep learning (PyTorch and Tensorflow for transformers, LSTMs, and CNNs), A/B testing, causal inference, Bayesian inference, time series forecasting, natural language processing (NLP), computer vision.

Engineering: AWS and GCP, GitHub Copilot, data platform design, DBT, Agile, git and git-flow, Docker, Kubernetes, CI/CD pipelines (CircleCI, Jenkins, TravisCI), NoSQL (DynamoDB, MongoDB), Spark, API development (flask, Django, Tensorflow Serving), MLOps.

Education

B.A. in Biological Sciences, Rice University

M.P.H. in Epidemiology, University of Michigan

Ph.D. in Epidemiology, University of Rochester