

# Anna Orosz

## Machine Learning Engineer

annaorosz.com | anna@orosz.pro | linkedin.com/in/anorosz/ | (215) 206-2244

### Education

---

#### University of Pennsylvania | Master of Science in Engineering in **Data Science**

🎓 *December 2021*

- Master's Thesis: "Generating Text-based Adventure Games" advised by Chris Callison-Burch
  - Built language generation model with training corpora extracted from text-based adventure games
  - Used GPT-3 to produce text with a user-defined set of characteristics
  - Applied fine-tuning, few-shot learning along a OpenAI's language models to auto-generate human-like text

#### University of Pennsylvania | Bachelor of Arts in Pure **Mathematics** and **Computer Science**

- Applied Machine Learning-focused coursework
- 10+ TA assignments for Machine Learning, AI, Data Science and Statistics classes

### Professional Experience

---

#### LinkedIn | Machine Learning Engineer | New York, NY

*March 2022 – Present*

- Ingest, infer and maintain company data for 60+ million companies as a member of Knowledge Graph Team
- Designed and implemented an end-to-end revenue ingestion pipeline for ingesting numerical data with 96% accuracy from SEC filings, enhancing company data with annual revenues for ~10k companies
- Owner of the company revenue pipeline. Facilitated launch of a new version which led to a \$6k/day increase in revenue through LinkedIn Marketing Solutions
- Collaborated to build URL ingestion pipeline resulting in 50M+ companies' URLs
- Implemented parked domain detection Spark job to remove ~500k low-quality URLs with 99% accuracy in Scala

#### LinkedIn | Machine Learning Engineer Intern | New York, NY

*May 2021 – August 2021*

- Used Scala and Spark to create a graph-representation (~100 million edges) of user-movement across companies
- Applied NLP technologies by training word2vec model to establish baseline performance of company embedding
- Pioneered a graph neural network model using graphSAGE to produce ~20 million company embeddings in 50 dimensions using TensorFlow
- Devised evaluation model for GNN performance on Fortune 500 companies by customizing k-means clustering

#### MBition – Mercedes-Benz | Artificial Intelligence Intern | Berlin, Germany

*June 2020 – August 2020*

- Visualized driving log data across all of Europe from Mercedes Benz cars collected throughout several years (10's of millions) to detect pattern between significant vs expected events
- Researched and evaluated time series models (SARIMAX, LSTM, PROPHET) in Azure Databricks with PySpark
- Developed robust ARIMA model capable of distinguishing between Mercedes Benz server/system outages and natural causes for unusual occurrences in driving logs

#### LogMeIn | Machine Learning Engineer Intern | Budapest, Hungary

*January 2018 – July 2019*

- Led award-winning Hackathon project by integrating BERT & fine-tuning the model with in-house data to develop an HR chatbot for internal usage and a QA system for LogMeIn's clientele
  - Won 1<sup>st</sup> prize Tech Innovation + Audience favorite prize + 2<sup>nd</sup> place Best Business Value Innovation prize
- Researched term weighting methods and class hierarchy models for email automation for a Top 5 Indian Bank
- Enriched Q&A text corpus in multiple languages by building sophisticated web-scraping services for Bold360 AI

#### Meta | Software Engineer Intern | Menlo Park, CA

*May 2017 – August 2017*

- Modernized internal client-library tool completing bulk data transfers - by operating with multi-tenancy, network utilization, scheduling, cross-dc connection-pooling in C++ - for Facebook's Data Scientists
- Overhauled internal tool COPTA for copying directories across HDFS clusters and FB's data centers
- Engineered detailed Scuba tables for Hyperloop and COPTA (100+ petabytes data)