

Anna Orosz (she/her)

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

Relevant Coursework: Deep Learning, Artificial Intelligence, Applied Machine Learning, Computational Linguistics, Big Data Analytics, Databases, Statistics for Data Science, Master's Thesis (Research)

University of Pennsylvania | Bachelor of Arts in Mathematics and Computer Science

Relevant Coursework: Algorithms & Data Structures, Unix, Python, Computer Architecture, Statistics for Data Science, Graph Theory, Analysis (I & II), Abstract Algebra (I & II), Linear Algebra, Game Theory, Calculus, Language and Automata

Professional Experience

LinkedIn | Artificial Intelligence - Machine Learning Engineer Intern | New York, NY

May 2021 – August 2021

- Build company embedding space for LinkedIn's ~20 million organizations
- Execute data preprocessing in *Scala* with *Apache Spark* to create directed weighted graph with ~100 million edges
- Train word2vec (gensim) and GNN (graphSAGE) models to produce company embeddings in 50-d using *TensorFlow*
- Use objective (e.g. logistic regression) and subjective (Appen) evaluation on embeddings

MBiton – Mercedes-Benz Innovation Lab | Artificial Intelligence Intern | Berlin, Germany

June 2020 – August 2020

- Built time series model detecting & classifying data outages among millions of Mercedes-Benz data
- Researched and evaluated state-of-the-art approaches (SARIMAX, LSTM, PROPHET) in *Azure Databricks* with *PySpark*
- Innovated on frontline of automotive world and develop next generation of self-driving Mercedes-Benz automobiles

LogMeIn | Machine Learning Intern | Budapest, Hungary

January 2018 – July 2019

- Developed award-winning NLP Hackathon project for training QA systems using state-of-the-art Deep Neural Networks
 - Won 1st prize Tech Innovation + Audience favorite prize + 2nd place Best Business Value Innovation prizes
- Researched term weighting methods, class hierarchy models facilitating email automation for Top 5 Indian Bank
- Enriched Q&A text corpus in languages (VN, GE, IN, etc.) by building sophisticated web-scraping services for Bold360 AI
- Specialized in Applied Research for **Natural Language Processing**, worked in *Python* with *Keras*, *Tensorflow*, *scikit-learn*

Facebook | Software Engineering Intern | Menlo Park, CA

May 2017 – August 2017

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

RapidMiner | Software Developer Intern | Budapest, Hungary

May 2016 – August 2016

- Developed software to make **Data Science** accessible to RapidMiner's customers
- Built operators RapidMiner Studio designed to process, analyze and alter data locally
- Designed Operators in RapidMiner Radoop - using Spark scripts - to process data on large scale with Apache Hadoop

Research

Persona-based Language Generation

- Build Language generation model with persona-based training texts obtained from Reddit
- Use GPT-3 to produce text with a user-defined set of characteristics
- Conducting research as a member of Professor Christopher Callison-Burch's Natural Language Processing lab

Teaching Experience

Teacher's Assistant @ CIS 530 – graduate-level Computational Linguistics:

Fall 2020

- Held weekly recitations, conducted 1-on-1 oral exams, wrote and graded programming assignments

Teacher's Assistant @ NETS 213 - Crowdsourcing and Human Computation:

Spring 2021

- Develop programming assignments and auto-graders, hold office hours

Teacher's Assistant @ CIS 521 - Artificial Intelligence:

Summer 2021

- Manage summer semester course, hold office hours, manage and grade assignments

Teacher's Assistant @ ESE 542 - Statistics for Data Science:

Summer 2021

- Lead summer semester course, hold recitations and office hours, manage and grade assignments

Proficiencies and Passions

Prog. Languages: Python, SQL, Java, C++, C, Bash/Shell, HTML/CSS/Javascript, OCaml, Assembly Language

Libraries: PyTorch, Numpy, Pandas, PySpark, scikit-learn, Keras, TensorFlow

Languages: German (Fluent), Hungarian (Native Speaker), French (Intermediate)

Interests: Computational Linguistics, Autonomous Vehicles, Machine Learning, Artificial Intelligence & Data Science

Activities & Hobbies: riding my Piaggio, horseback-riding, bouldering, salsa-dancing, ice-skating & skiing, cinematography