

# Oliver Spohngellert

spohngellert.o@northeastern.edu | 917-750-1162  
github.com/spohngellert-o | oliverspohngellert.com | linkedin.com/in/spohngellert-o

---

## EDUCATION

### MASTER OF SCIENCE IN DATA SCIENCE | NORTHEASTERN UNIVERSITY

May 2021 | Boston, MA

**GPA:** 3.8/4.0 | **Relevant Courses:** Data Science Capstone, Deep Learning, Reinforcement Learning, Supervised Machine Learning, Unsupervised Machine Learning, Statistical Methods in Computing, Data Management, Algorithms

### BACHELOR OF SCIENCE IN COMPUTER SCIENCE, MINOR IN MATH | NORTHEASTERN UNIVERSITY

May 2019 | Boston, MA

**GPA:** 3.73/4.0 | **Relevant Courses:** Large Scale Storage/Retrieval, Software Development, Linear Algebra and Differential Equations, Object Oriented Design, Probability and Statistics

## SKILLS

**Programming:** Python, R, Scala, pandas, numpy, Keras, sklearn, matplotlib, tidyverse, git, SQL

**Concepts:** Deep Learning, Object Detection, Supervised Learning, Unsupervised Learning, Functional Programming, Software Development, NoSQL

## WORK EXPERIENCE

### DEEP HEALTH | MACHINE LEARNING INTERN

May 2020 – Aug 2020 (3+ months) | Cambridge, MA

- Designed an error analysis pipeline to enhance understanding of a Breast Cancer Detection model's weak points using Python.
- Developed a system for aggregating cancer annotations and confirmations from doctors, comparing results, and generating follow ups using Javascript and Python.

### NORTHEASTERN NDS2 LAB | RESEARCH ASSISTANT

Jan 2019 – Feb 2020 (1 year, 1 month) | Boston, MA

- Implemented multiple Deep Learning classifiers such as LSTMs to classify users of IoT systems using exclusively network traffic.
- Engineered features and implemented Isolation Forest and Kernel Density Estimator models to detect self-propagating malware such as WannaCry.
- Cleaned over 50 GB of simulated messages between self-driving cars to develop a system for identifying malicious messages, which achieved over 90% accuracy on all attacks.

### NETRA | DEEP LEARNING/MACHINE LEARNING CO-OP

Jan 2018 – July 2018 (6 months) | Boston, MA

- Optimized download, upload, and labeling sections of data pipeline via multiprocessing to decrease training time for the object detection system from 4 weeks to 2 weeks.
- Trained logo detection model using transfer learning, increasing number of classes by 500 while maintaining performance.
- Researched NLP approaches for content safety in videos, using FastText and Word2Vec models on video descriptions.

### INTUIT | SOFTWARE ENGINEER CO-OP

Jan 2017 – July 2017 (6 months) | San Diego, CA

## PUBLICATIONS & PROJECTS

**Large-Scale Analysis of Gene Sensitivity:** Individual project for Statistical Methods in Computer Science course. Used Broad Institute's L1000 dataset to identify genes that are most and least likely to change expression level under perturbation. Applied large scale ANOVA testing with Benjamini-Hochberg correction for multiple testing.

**Predicting The Outcome of Smashbros:** Group project for Supervised Machine Learning course. Scraped data from 10 tournaments, extracted and cleaned it in order to predict the outcome of games based on the current game state. Achieved 74% accuracy and 0.83 AUC using a Logistic Regression model.

**The House That Knows You** Extracted data from IoT network traffic logs to classify the user of the IoT devices during a given time period, with top overall F1 score of 0.86. <https://arxiv.org/abs/1908.00592>