

# Oliver Spohngellert

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

## EXPERIENCE

### DEEP HEALTH | MACHINE LEARNING INTERN

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

- Designed an error analysis pipeline to enhance understanding of Breast Cancer Detection model's weak points using Python.
- Developed a system for taking in 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 in order 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 in order 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, MA

## PROJECTS

### PREDICTING THE OUTCOME OF SMASHBROS

Group project for Supervised Machine Learning. 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.

### LARGE-SCALE ANALYSIS OF GENE SENSITIVITY

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

## PUBLICATIONS

**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>

## EDUCATION

### NORTHEASTERN UNIVERSITY

CANDIDATE FOR MASTER'S OF SCIENCE  
IN DATA SCIENCE

Expected May 2021 | Boston, MA

Cum. GPA: 3.67 / 4.0

BACHELOR OF SCIENCE IN COMPUTER  
SCIENCE WITH MINOR IN MATHEMATICS

May 2019 | Boston, MA

Dean's List, Dean's Scholarship

Cum. GPA: 3.73 / 4.0

## SKILLS

### PROGRAMMING

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

### CONCEPTS

Deep Learning, Supervised Learning, Unsupervised Learning, Functional Programming, Software Development, NoSQL

## COURSEWORK

### GRADUATE

Supervised Machine Learning  
Unsupervised Machine Learning  
Reinforcement Learning  
Statistical Methods in Computer Science  
Data Management and Processing  
Algorithms

### UNDERGRADUATE

Large-Scale Information Handling  
Software Development  
Linear Algebra and Differential Equations  
Object-Oriented Design

## INTERESTS

Chess, Golf, Bicycling, Data Privacy, Cooking