

# COLIN QUIRK

email: [colinquirk1993@gmail.com](mailto:colinquirk1993@gmail.com) | website: [colinquirk.com](http://colinquirk.com) | github: [@ColinQuirk](https://github.com/ColinQuirk)

## WORK EXPERIENCE

### *SparkNeuro – Lead Data Scientist (Director of Machine Learning)* 2020-Present

- Managed a team of 3 data scientists on the creation of XGBoost models to predict whether patients had cognitive impairment and Alzheimer's disease in support of a clinical study and published the results of our models at a medical conference (AUC: 0.90, sensitivity: 0.88, specificity: 0.91).
- Helped plan and build a production level MLOps infrastructure built on AWS Lambda to ensure our data science processes were trackable, testable, and repeatable.
- Designed and created a python dashboard with input from business stakeholders in order to allow for tracking our clinical study progress by combining data from multiple sources into clear and understandable figures.
- Assisted with the design and writing of our clinical study's statistical analysis plan and protocol and performed power analysis and simulations to guide internal thinking.

### *Allstate – Data Science Intern* Summer 2019

- Created a model using XGBoost to predict likelihood of a quote binding to a high level of performance given the difficulty of the task (AUC: 0.77) while being vigilant about the possibility of data leakage. After ending my internship project ahead of schedule, I extended this work and created a new model for other types of quotes.
- Presented an analysis based on simulations I ran suggesting that applying our team's models to a different business problem could earn over 1 million dollars more than the current solution.
- Provided simple entry point scripts for my model building code so that my team could refit my models as needed and assisted with the deployment of my model via pyspark and airflow.
- Added the ability to visualize SHAP values from our models into our shared machine learning libraries.
- Gave a company wide, 1-hour tech talk about the usefulness of animation as a data visualization tool.

## ADDITIONAL PROJECTS

### *Classifying Individuals with Eyetracking Data using Deep Learning*

- Cleaned over 20GB of eye tracking data and trained a CNN to predict which individual each segment of eye tracking data belonged to at over 95% accuracy.

### *Spatial Biases Influence Encoding of Information into Working Memory*

- Reanalyzed experiments with over 450,000 trials by combining data from multiple sources and created multiple open source python libraries which have been used by other researchers to run their own experiments.

### *Started Writing an Open-Source Statistics Textbook*

- Began writing an open source statistics textbook ([bookdown.org/cquirk/LetsExploreStatistics](http://bookdown.org/cquirk/LetsExploreStatistics)) that displayed javascript simulations inline with the text.

### *Youtube Channel for Teaching R*

- Started a youtube channel ([www.youtube.com/watch?v=uJKpz9T7mAg](https://www.youtube.com/watch?v=uJKpz9T7mAg)) to teach viewers R for free. My video on regular expressions has had over 6,000 views.

## EDUCATION

University of Chicago

*Ph.D. in Psychology (ABD)\**

2023\*

Hampshire College

*B.A. in Cognitive Neuroscience*

2015

## TECHNICAL SKILLS

Python, R, SQL, Regression, Random Forest, XGBoost, CNNs, AWS, pyspark, docker, linux, git

For a full list of my publications, please see [colinquirk.com/publications](http://colinquirk.com/publications)