# Practicum Sprint 3 - Status Check-In

Matthew Agard
magard3@gatech.edu

## 1 ACCOMPLISHMENTS

- Finalized my data source (DDI dataset)
- Uploaded DDI to AWS S3
- Developed AWS Glue job for metadata feature engineering
    - Script developed using solely PySpark, which was my goal

## 2 CHALLENGES

No significant challenges have been faced to date. However, a concern of mine remains regarding the dataset's size. There are only 656 records in total, and after feature engineering each record only contains 2 categorical features, an image and a target variable. Neural networks tend to generalize better to large datasets, so as stated in my project planning document I intend to artificially enhance the dataset's size with image augmentation to address this concern.

## 3 SPRINT PLANS

This upcoming sprint will consist entirely of watching videos and reading documentation to learn AWS SageMaker, followed by utilizing said service to build and deploy my neural network for image classification. I no longer need to configure an AWS Redshift data warehouse as I'd originally planned since I have only 1 data source. However, that works in my favor for staying on schedule given my severe underestimate of the AWS SageMaker learning curve.