

# BAILEY WEI

New York, NY | 646-371-6126 | Work Authorization: U.S. Citizen  
[bw489@cornell.edu](mailto:bw489@cornell.edu) | [linkedin.com/in/baileywei](https://www.linkedin.com/in/baileywei) | [github.com/bw-tech](https://github.com/bw-tech) | [baileywei.com](https://baileywei.com)

## EDUCATION

<b>Cornell University</b> , New York, NY	May 2020
<b><i>Master of Engineering in Operations Research and Information Engineering – Data Science Concentration</i></b>	
Honors/Awards: Cornell Merit Scholarship	
Relevant Coursework: Deep Learning, Natural Language Processing, Data Science in the Wild, Applied Machine Learning, Optimization Methods, Modeling Under Uncertainty	
<b>University at Buffalo</b> , Buffalo, NY	May 2019
<b><i>Bachelor of Science in Industrial Engineering</i></b>	
Honors/Awards: Cum Laude, Dean's List	

## TECHNICAL SKILLS

<b>Programming Languages:</b>	Python, R, SQL, Swift, C++, MATLAB
<b>Frameworks and Libraries:</b>	Matplotlib, NLTK, Pandas, PySpark, PyTorch, Scikit-Learn, Spark, Transformers
<b>Operations Research:</b>	Gurobi Optimizer, Six Sigma Green Belt
<b>Additional Tools:</b>	AWS, Azure, CUDA, Google Colab, Tableau

## EXPERIENCE

<b>Cornell Tech, <i>iOS Engineering Intern</i></b> , New York, NY	Dec. 2019 – Present
<ul style="list-style-type: none"><li>Developed an iOS application using Microsoft Computer Vision API to generate captions for taken photos</li><li>Integrated Amazon S3 to create data storage solution for real-time result and feedback collection</li><li>Enhanced accessibility for vision-impaired users by implementing VoiceOver and other audio cues</li></ul>	
<b>Nielsen, <i>Data Science Intern</i></b> , Columbia, MD	Jun. 2019 – Aug. 2019
<ul style="list-style-type: none"><li>Utilized U.S. Census and internal data to develop random forest, k-NN, and logistic regression models to identify hard-to-survey households</li><li>Increased F1 score by 150% and reduced household false positive rate by 50% over existing models</li><li>Enhanced data preprocessing methodology through reducing discarded radio surveys to save \$300K annually</li><li>Promoted continuous learning and integration of relevant industry practices by leading weekly forums for a team of 11 scientists</li></ul>	
<b>Sentient Science, <i>Predictive Analytic Modeling Intern</i></b> , Buffalo, NY	Feb. 2019 – May 2019
<ul style="list-style-type: none"><li>Developed a visualization library in Python to automate graph creation, reducing deliver times of KPIs to leadership by 10%</li><li>Validated and predicted wind turbine failures by creating regression models to support data scientists</li><li>Optimized data aggregation from Amazon S3 by reducing data acquisition time by 15% per visualization</li></ul>	

## PROJECTS

<b>Fine-Grained Sentiment Classifier</b> , arXiv:2005.13619, (Google Colab, Python, PyTorch)	Spring 2020
Created a classifier to predict sentiment ratings on the Stanford Sentiment Treebank dataset <ul style="list-style-type: none"><li>Evaluated several BERT-like models to determine accuracy and complexity trade-offs</li><li>Utilized Google Colab GPUs to establish state-of-the-art accuracy, 60.2%, using RoBERTa-Large models</li></ul>	
<b>Fake Twitter News Detection</b> , (AWS, NLTK, Python, PyTorch, Scikit-Learn)	Spring 2020
Identified fake news related to politics and the U.S. election through deep learning solutions <ul style="list-style-type: none"><li>Developed pipeline to extract and preprocess a dataset of over 6,000,000 unique tweets</li><li>Utilized AWS along with NLTK and PyTorch to create BERT embeddings which reached 99.81% accuracy</li></ul>	
<b>Cornell Tech, BigCo Studio Challenge: Intersection</b> , (HTML, JavaScript, Python, SQL)	Spring 2020
Developed a trivia web application to integrate with Intersection kiosks to push user engagement and collect additional data <ul style="list-style-type: none"><li>Analyzed various product ideas to best facilitate kiosk-to-phone experience and increase kiosk attractiveness</li><li>Validated long-term user retention through conducting several experiments on live Intersection kiosks</li></ul>	
<b>Cornell Tech, Product Studio Challenge: JPMorgan &amp; Co.</b> (Python, Scikit-Learn, Swift, Tableau)	Fall 2019
Developed a recommender system to predict changing neighborhoods and demographics <ul style="list-style-type: none"><li>Utilized Instagram, Twitter, and Yelp data to further push accuracy of neural network models</li><li>Built iOS application to implement recommender system and visualize long-term neighborhood shifts</li></ul>	
<b>Large-Scale Image Search Engine</b> , (Python, Scikit-Learn)	Fall 2019
Created a model to retrieve relevant images based on text inputs <ul style="list-style-type: none"><li>Utilized TF-IDF, Word2Vec, ridge regression, ResNet, and neural networks to classify 85% of correct images</li><li>Developed color-based feature vectors for additional data to drive predictive power and accuracy</li></ul>	