

Andy Cheon

San Francisco, CA

[linkedin.com/in/acheon](https://www.linkedin.com/in/acheon) | github.com/ajcheon | cheon.andy@gmail.com | 408-318-1892

EDUCATION

M.S. Data Science - University of San Francisco **June 2020**

Machine Learning, Deep Learning, SQL, Distributed Computing (PySpark),
Experimental Design (A/B Testing), Data Structures, Algorithms, Time Series Analysis

B.S. Bioengineering - University of California, Berkeley **May 2015**

PROFESSIONAL EXPERIENCE

Zyper, Data Science Intern (San Francisco, CA) **Oct 2019 - Present**

- Built and deployed an image classification convolutional neural network (CNN) with PyTorch to help brands efficiently recruit fans with desired aesthetic types on social media.
- Applied feature importance methods to identify top factors that drive engagement rates of user-generated content.
- Developed a user location prediction pipeline using NLP tools (NLTK, spaCy), improving existing predictor and optimizing the fan recruitment process.
- Discovered and visualized trends from group chat content from over 15 brand communities.

Genentech, Software Developer and Data Analyst (South San Francisco, CA) **Apr 2018 - June 2019**

- Built four interactive web applications with R Shiny for researchers to visualize and assess patient data from multiple clinical trials.
- Identified target genes for a developmental drug using weighted gene co-expression network analysis (WGCNA) from a database of 20,000+ genes.
- Analyzed RNA expression levels using cluster analysis methods in R from 60,000+ sequencing samples.

Pendulum Therapeutics, Software Engineer (San Francisco, CA) **July 2015 - Feb 2018**

- Built a web app and REST API with Django and PostgreSQL, providing tools for lab data management and analysis. Housed data from 350+ experiments.
- Developed data visualization tools with R Shiny to facilitate biological assay development.
- Streamlined laboratory data entry protocols with custom-built Google Apps Script plugins.

PROJECTS

Sparkle: a machine learning-based platform to encourage medication adherence

- Consists of integrated iOS, watchOS, and [web apps](#) allowing patients to automatically record medication intake via their AppleWatch, and doctors to view patients' medication adherence rates.
- Trained a gradient boosting classifier with Apache Spark on AWS EMR using smartwatch motion data to detect low pill counts in prescription bottles (achieved 0.82 F1 score).
- Paper accepted to the IEEE Engineering in Medicine and Biology Society Conference (EMBC) 2020.

Image captioner: a deep learning model that generates captions for images

- Built a CNN-LSTM *encoder-decoder* network that translates vector representations of image features into a sequence of words comprising the caption.
- Trained with an AWS GPU on 591,000 image-caption pairs from COCO dataset.

Scikit-learn emulator

- Implemented core machine learning algorithms from scratch with comparable performance to scikit-learn modules, including Ridge and Lasso regression, Naive Bayes, decision trees, random forest, and k-means clustering.

TECHNICAL SKILLS

Python (pandas, scikit-learn, NumPy, spaCy, PyTorch), PostgreSQL, R, Apache Spark (PySpark, SparkSQL), ggplot, matplotlib, AWS, Docker, web applications (Flask, Django, JavaScript), data scraping (Selenium, BeautifulSoup)