

Colin Jiang

colin08812@gmail.com | 908 548 5048 | colin-jiang.github.io

Education

Emory University

Atlanta, GA

Bachelor of Science, Computer Science and Mathematics

May 2020

Honors: *Chair's Award for Achievement, Phi Beta Kappa, Dean's List*

GPA: 3.9/4.0

Courses: Data Structures & Algorithms, Machine Learning, Artificial Intelligence, Theory of Computing, Data mining, Systems Programming, Discrete Structures, Database Systems, Linear Algebra, Numerical Analysis, Probability & Statistics

Skills

Coding Languages: Proficient in Python, SQL, JavaScript; Familiar with Java, HTML/CSS, Scala, C

Other: Next.js, DynamoDB, Presto, React, Jupyter, S3, CloudFront, MongoDB, Node.js, Spark, Scikit-learn, Hadoop

Experience

Facebook

Remote (based in NYC)

Data Engineer

August 2020 – Present

- Support ads ecommerce integrity team where we focus on improving FB ads ecommerce ecosystem and prevent misleading experiences. Work on collecting signal on advertisers including user comment sentiment ML models, user surveys, and invoice data to determine advertiser trust and measuring/enforcing on negative experiences
- Tech stack is primarily in-house with Python-based Airflow alternative, and Presto as query engine for logging data
- Created new measurement system for estimating prevalence of fake/irrelevant reviews and Q&A on FB/IG shops, involving sampling content for ops to review, generating metrics estimating prevalence of problem, accuracy of labelling, along with signals for content banning
- Worked with ML teams in identifying signals and creating ground truth data for classifying violating content in an effort to automate enforcement and reduce false positives in a new domain area

Facebook

Menlo Park, CA

Data Engineering Intern

May 2019 – August 2019

- Improved statistical approximations of classifier missed violating content under the Community Integrity team
- Implemented bias correction on bootstrap confidence intervals of metrics for classifier missed violating content
- Optimized and unified two separate sampling pipeline systems to reduce daily I/O and computational cost by 50%

University of Miami, REU

Coral Gables, FL

Deep Learning Vision Researcher

May 2018 – July 2018

- Proposed and investigated methods of improving image recognition under a low contrast environment
- Trained large-scale convolutional neural networks Improving accuracy of base model by ~4% for low contrast images

Projects

Onetricks.gg

August 2021

- ~40k monthly active users and growing, ~600k monthly pageviews, 220k lifetime unique users as of Dec 2022
- Owner and sole developer of a gaming analytics web app that analyzes strategies of top players and recommends strategies for users; current coverage for League of Legends with plans to expand to other popular games
- Site grew primarily through word of mouth and organic SEO, with no marketing expenses. Bootstrapped all UI/UX designs, marketing material, social media engagement, and ad campaigns (free ad credits)
- Novel discovery and ranking algorithm for top players and unique strategies, and syncing game data with Twitch VODs
- Web stack: fullstack Next.js and DynamoDB for storing millions of matches data from Twitch and Riot Games APIs. Daily Google Cloud cron jobs consume and analyze live data from Twitch and Riot APIs to have the most up to date game data. Designed to scale with user growth with ISR (caching) and efficient DynamoDB querying and optimized for web speed. Hosted on Vercel and AWS S3 + CloudFront CDN for image storage and caching

Twitter Hashtag Recommender System

May 2019

- Recommends top-k hashtags for a given tweet to help users/businesses target the most relevant audience
- Novel recommender system, using popularity weighting schema to capture viralness of Twitter posts