

# Nathan Jones

nathancooperjones.com

✉ nathancooperjones@gmail.com

in nathancooperjones

🐙 github.com/nathancooperjones

## SKILLS

---

Experience building fullstack machine learning pipelines including data collection and processing, iterative model building, and deployment to production. Proficiency in building computer vision, natural language processing, tabular, and recommendations datasets and models. Creator and maintainer of **Collie**, an open source library for building and evaluating deep learning recommendation models.

*Languages:* Python, SQL, R

*Libraries:* Pandas, NumPy, Scikit-learn, PyTorch, Keras, Tensorflow, PySpark, FastAPI, Flask, Prodigy, spaCy

*Technologies:* Airflow, AWS EC2, AWS SQS, AWS S3, Databricks, Docker, Elasticsearch, Git, Kafka, Redshift, Snowflake

## EXPERIENCE

---

### The Mom Project

Senior Data Scientist

Aug. 2021–present

- Productionalized a recommendations algorithm backed by Elasticsearch to pair jobs with relevant talent, and talent with relevant jobs. In an email A/B test, the new model resulted in an click-to-open rate increase of 55% and an unsubscribe rate decrease of 51% when compared to the previous model
- Developed an innovative method for solving the cold start recommendation problem using a modified version of out-of-vocabulary embedding imputation
- Created a library and API for finding fuzzy duplicate jobs in our network, which is being used in both the new recommendations algorithm and automated job approval
- Spent ~25% of my time mentoring both a new data scientist and a senior data scientist transitioning from R to Python on engineering best practices, project planning, and Python
- Worked with the infrastructure team to develop deployment plans for our recommendations API using AWS Batch, SQS, and OpenSearch

### ShopRunner

Data Scientist

Aug. 2019–July 2021

- Created and open-sourced **Collie**, a novel, deep learning recommendations library based around collaborative filtering and built with both flexibility and scalability in mind
- Deployed improved product-product, retailer-retailer, and member-product recommendations systems. In an email A/B test, this new member-product recommendations system showed a 184% improvement in click to open rate compared to our previously deployed recommendations system
- Built a low-latency (<50ms), unsupervised fraud model for real-time use in detecting anomalous logins. Pioneered unsupervised learning at ShopRunner via a novel evaluation metric called “cross scoring”
- Made a real-time, supervised fraud model for credit card transactions data, capable of reducing chargebacks by up to 52% more than the original rules-based system in place
- Researched and implemented a stable solution for how our team used CUDA and GPUs within Airflow jobs

Data Science Intern

May–Aug. 2019

- Built the initial framework for a deep learning recommendations library to outperform our existing recommendations algorithm

## Geena Davis Institute

Senior Data Scientist (Contract)

June 2020–May 2021

- Created a novel multi-task object recognition pipeline to identify, classify, and cluster character faces in media on perceived identities of gender, skin tone, age, and body type. These automated findings helped identify disparities in media with human-level annotation accuracy
- Managed the Institutes’s AWS cloud infrastructure in hosting assets and deploying image and text annotation tools for the coding team to use
- Built spaCy NER and fine-tuned BERT text classification models to automatically annotate data containing harmful language
- Created a two-year roadmap for all current and future machine learning projects to be completed by teams working at the Institute
- Summarized important data stories in reports and presentations for clients to effectively make systemic changes in their content

Researcher (Contract)

May 2017–June 2020

- Collected, cleaned, and analyzed datasets to identify bias in media

## Computational Physiology Lab at the University of Houston

Undergraduate Research Assistant

June–Aug. 2018

- Ran the data visualization and explanatory modeling components for a novel study examining stress in the workplace

## TALKS

---

- “Understanding and Playing *Deal or No Deal* Better Than a Human.” Internal ShopRunner Engineering Conference. Sept. 2020
- “Detecting, Recognizing, and Analyzing Animated Faces.” Animal Crossing Artificial Intelligence Workshop. July 2020

## SELECTED WRITINGS

---

- **These Bored Apes Do Not Exist.** Flooding the NFT market by training a lightweight GAN + super-resolution model to generate Bored Ape Yacht Club NFTs that do not exist
- **Fetching Better Beer Recommendations with Collie.** A three-part tutorial walking through a full recommendations pipeline using Collie, including dataset preparation, architecture choice, model evaluation, and inference

## PUBLICATIONS

---

“Capturing Individual Differences in Stress Reactions to Email Interruptions Using Thermal Imaging” (co-author), *Association for Computing Machinery Conference on Human Factors in Computing Systems*. May 2019

## EDUCATION

---

### Illinois Institute of Technology

Master of Data Science. GPA 4.0/4.0. Summa cum laude.

Jan. 2018–Dec. 2019

B.S. Computer Science. GPA 4.0/4.0. Summa cum laude.

Aug. 2016–Dec. 2019