

# Nathan Jones

nathancooperjones.com

✉ nathancooperjones@gmail.com

in nathancooperjones

🐙 github.com/nathancooperjones

## SKILLS

---

Experience with a fullstack machine learning pipeline, 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 the open source library, *Collie*.

*Languages:* Python, SQL, R

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

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

## EXPERIENCE

---

### ShopRunner

Data Scientist

Aug. 2019–present

- Created and open-sourced *Collie*, a novel, deep learning recommendations library based on collaborative filtering and built with both flexibility and scalability in mind
- Productionalized 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 uses 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 original collaborative filtering model

### 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, built spaCy NER and fine-tuned BERT text classification models to automatically annotate data containing harmful language, deployed image and text annotation tools to be used in production by our research team, and roadmapped all 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

## 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

## 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

## 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