Nate Jones

Staff Machine Learning Engineer

Chicago, IL | nathancooperjones@gmail.com
nathancooperjones.com | github.com/nathancooperjones
linkedin.com/in/nathancooperjones

Profile

Staff ML engineer with 7+ years of experience in recommendations and search, now leading LLM agent work across teams. Building LTK's shared multimodal embedding, retrieval, and serving stack and shipping models with A/B-tested gains in engagement and GMV.

Technical Skills

Languages: Python, SQL
ML specialties: Recommendations, semantic retrieval, multimodal embeddings, LLM agents, A/B testing
ML and agents: PyTorch, scikit-learn, ONNX, Core ML, OpenAI Agents SDK, OpenAI ChatKit, MCP
Infrastructure: OpenSearch, Airflow, AWS, Docker, FastAPI, Streamlit, Datadog

Experience

LTK Jan. 2023–Present

Staff Data Scientist & MLE Feb. 2025–Present

- Acting manager of the Data Science & Machine Learning team with **5 direct reports**.
- Technical lead for a proactive agent system on the LTK Brand Platform, with a planned beta launch in late 2026. Wrote the vision doc **5 distinct teams** build from, covering multi-agent design, memory, and evals.
- Technical lead across **3 distinct teams** for a consumer AI shopping assistant. Owned the architecture and built the agent harness from prototype to production, including MCP tools, streaming, structured UI, and guardrails. It has reached **325K users**, with weekly users growing from **7K to 12K**. In its first two weeks, **12%** of users clicked out to products. Its search version powers AI Overview at **245K requests weekly**.
- Pitched and shipped Dynamic Interests, an unsupervised model that finds evolving topics in post embeddings. Worked with backend to serve it at low latency to all users in Home feeds and interest chips. One month in, 5.3K users had added **32K interests** and **39.5%** of chip taps led to post opens.
- Built a creator-profile recommender and cut its batch run from **30+ hours to under 10 minutes**, fast enough for personalized notifications. In an A/B test, they raised GMV **6.3%** and clickouts **4%**.
- Led development of a shared multimodal embedding model for posts, products, and search queries across images, video, and text. It replaced two older embedding models and now powers **7+ other use cases**. Trained natively in FP16 to cut embedding storage by up to 50% with no quality loss.
- Built semantic search on these embeddings across millions of products, hundreds of thousands of posts, and thousands of creators. In an A/B test against the long-running Cohere-based search, it raised post CTR **2.45%**, product CTR **2.63%**, and video post clicks **9%**, while cutting search costs by **about 65%**.
- Built shared Go + ONNX serving with the backend team for 7+ production models, **cutting p99 latency in half** compared to Python APIs.

Senior Data Scientist & MLE II Jan. 2023–Feb. 2025

- Built personalized brand recommendations from creators' posts, favorites, and collaborations, served by a sub-100 ms API. Raised brand clickouts **68.4%** in an A/B test ($p < 0.001$) and shipped them to all creators.
- Led on-device product tagging ML with iOS engineers, using Core ML object detection and visual matching. Raised offline match rate from **about 20% to 80%** and helped cut models from 800 MB to under 100 MB. About **a third of LTK posts** use auto-tagging; on these posts, it adds **57%** of product links.
- Mentored four data scientists, and **two have since been promoted**. Set the team's coding standards and co-created the hiring case study used as the team nearly doubled.

Earlier Experience

Condé Nast July 2022–Jan. 2023
Data Scientist III

- Trained and handed off a production-ready multilingual transformer classifier for articles and recipes. It beat the existing text-vector model by **over 35%** in offline recommendation evaluation.
- Improved monitoring for newsletter recommendation pipelines, bringing the job failure rate to **under 1%**.

The Mom Project Aug. 2021–June 2022
Senior Data Scientist

- Built and deployed an Elasticsearch-backed talent–job recommendation API. An email A/B test showed **55% higher click-to-open rates** and **51% fewer unsubscribes** than the previous version of the model.
- Developed cold-start recommendations using modified out-of-vocabulary embedding imputation.
- Built a fuzzy duplicate-job detection library and API for matching and automated job approval.

ShopRunner May 2019–July 2021
Data Scientist Aug. 2019–July 2021
Data Science Intern May–Aug. 2019

- Created and open-sourced **Collie**, a deep learning library for collaborative filtering adopted at many other companies, including The Mom Project, LTK, and Spotify. Deployed product–product, retailer–retailer, and member–product recommendations. In an email A/B test, the member–product model raised click-to-open rates **184%** over the previously-deployed matrix factorization model.
- Built a real-time anomalous-login model with **under 50 ms** latency and developed "cross scoring," an evaluation method for unsupervised fraud detection.
- Built and deployed a real-time supervised credit-card fraud model that reduced chargebacks by **up to 52%** versus the existing rules-based heuristics.

Contract Experience

The Representation Project June 2021–Jan. 2025
Senior Data Scientist (Contract)

- Owned machine learning models, data analysis, and cloud infrastructure. Built and maintained a portal for uploading, tracking, and analyzing bias in media content. Contributed to **3 published studies**.

Geena Davis Institute May 2017–May 2021
Senior Data Scientist (Contract) June 2020–May 2021
Researcher (Contract) May 2017–June 2020

- Built a multitask computer vision pipeline to detect, classify, and cluster faces for media representation research, with human-level annotation accuracy. Developed spaCy NER and fine-tuned BERT classifiers, and managed AWS infrastructure and annotation tools. Contributed to **9 published studies**.

Education

Illinois Institute of Technology Chicago, IL
Master of Data Science, GPA 4.0/4.0 2019
B.S. Computer Science, GPA 4.0/4.0 2019