ADAM CATTO

agocatto@gmail.com • adamcatto.github.io • New York, NY, USA

EXPERIENCE

Data Scientist, Icahn School of Medicine at Mount Sinai **Jun 2022 – Present**

- First author on a deep-learning-driven behavioral neuroscience manuscript; built an end-to-end pipeline combining active-learning and pose estimation with time-series models (LSTM, TFT) to forecast near-term behavioral events.
- Re-implemented and trained from scratch Google DeepMind's AlphaGenome foundation model in PyTorch/DeepSpeed. Achieved comparable performance to DeepMind's model on over 1,000 human functional genomics assays with 100x less compute.
- Designed a novel deep learning task for neurogenomics, predicting cell-type-specific gene expression from personal genome sequence variants. Implemented and validated on Alzheimer's Disease (AD) dataset and discovered genetic drug targets for AD.
- Designed and trained a multi-scale 1D CNN for detecting neuron spikes in wavelet-decomposed electrophysiology signals, achieving an F1 of 81% and eliminating dozens of hours of manual annotation per experiment.
- Built a full-stack, end-to-end video object segmentation and tracking application integrating Meta's SAM3/SAM2 foundation models via a FastAPI + React/TypeScript stack; implemented click-to-annotate canvas overlays with real-time mask inference, cross-batch identity-preserving propagation over arbitrarily long videos using server-sent events streaming, and a user-friendly interface for remote annotation and review of segmentation results.
- Implemented a local LLM-based question-answering RAG agent with Llama 3.2 70B over thousands of neuroscience papers.
- Led analysis of 2D/3D spatial transcriptomics, whole-mouse-brain 3D microscopy, animal behavior video, and in-vivo/in-vitro electrophysiology datasets. Co-author on multiple manuscripts.
- Benchmarked self-supervised deep learning learning algorithms for medical imaging tasks.
- Integrated single-nucleus RNA-seq, 2D and 3D spatial transriptomics datasets to create a cell-type atlas of the mouse hippocampus.

Research Data Scientist Intern, Johnson & Johnson **Jun 2021 – Oct 2021**

- Discovered 9 digital biomarkers of 4 autoimmune disorders from wearable sensor data.
- Implemented distributed pipeline for terabyte-scale accelerometer data. Improved running time > 100x.
- Implemented signal processing algorithms to detect interpretable sleep and activity patterns in accelerometer data.
- Trained machine learning models to classify disease vs. healthy controls from interpretable time-series features.

Graduate Research Assistant, Research Foundation of CUNY **Aug 2020 – Jun 2022**

- Contributed to data preprocessing, machine learning model building, and visualization/analysis that won a national data challenge hosted by the National Institutes of Health: https://www.nichd.nih.gov/research/supported/challenges/decoding-maternal-morbidity. Wrote multiple technical machine learning preprints documenting the methods and results.
- Conceptualized, implemented, and validated two novel ensemble learning techniques for handling large amounts of missing data in tabular datasets, with and without imputation.
- Built POMDP-based sequential decision making clinical decision support system to optimize pregnancy treatment plans

SKILLS

- **Languages:** Python (Advanced), Shell Scripting, R, JavaScript
- **Machine Learning / Data Analysis:** PyTorch, PyTorch-Lightning, PyTorch-XLA, DeepSpeed, TorchAudio, TorchVision, PyTorch-Forecasting, OpenCV, Pandas, Scikit-Learn, HuggingFace, LlamaIndex, LangChain, RAG
- **Miscellaneous:** Linux, Google Cloud Platform, Large-Scale Deep Learning, Streamlit, LSF, Git

EDUCATION

CUNY Graduate Center, MS Data Science **2020 – 2022**

- **(GPA 3.90/4.00)** Graduate Coursework: Machine Learning, Data Visualization, Artificial Intelligence, Data Mining, Computational Biology, Digital Image Processing, Deep Learning for Genomics, Big Data Analytics
- **Thesis:** *Hierarchical Model Transfer Methods for Ensemble Learning with Large Amounts of Missing Data*

Stony Brook University, BS Applied Mathematics & Statistics, Philosophy **2015 – 2019**

PROJECTS

AI Bioinformatics Research Assistant

Used LlamaIndex, Ollama, and Llama3.2 to create a local RAG agent that autonomously plans and executes an analysis of next-generation biological sequencing data, demonstrating its reasoning and outputs in real-time in a Jupyter notebook.