Jake Pappo

Boston, MA • 781-799-7290 • jakepappo@college.harvard.edu • pappoja.github.io

EDUCATION

Harvard University, '25 | A.B. in Statistics (Data Science track) | GPA: 3.9
Coursework: Machine Learning, Data Science, ML Ops, Probability & Inference, Linear Algebra, Corporate Finance
Belmont Hill School, '20 | GPA: 4.54 | ACT: 36 | SAT Subject Test (for Math 2, Latin, History): 800
Regeneron Semi-Finalist, US History Term Paper Prize, Perfect Score on the NLE, Three-sport Varsity Athlete

EXPERIENCE

Sauron   *Jan 2024 – Aug 2024*
Machine Learning Engineer   *San Francisco, CA*
- Developed and deployed sensor fusion models for real-time object detection, tracking, and 3D mapping
- Led the creation of the company's data pipeline, from collection and labeling to re-training and evaluation
- Recruited as #2 employee by CEO Kevin Hartz (co-founder, Eventbrite and A* Capital) prior to $18M seed round

Harvard Business School   *Jan 2022 – Current*
Research Assistant under Professor Josh Lerner   *Boston, MA*
- Built a database for training LLMs (e.g, BERT, GPTs) to classify patents as financial or non-financial
- Formulated a framework and drafted case studies for sustainable private sector investments (book in progress)
- Researched the Web3- and crypto-equivalents of traditional startups and venture capital firms

Dent Wizard   *Jun 2023 – Aug 2023*
Business Intelligence
- Created dashboards to record, monitor, and analyze performance metrics for the company's nationwide locations
- Presented findings to executive management to inform data-based decisions in pursuit of $1B revenue goal

Trailer Central   *Jun 2022 – Aug 2022*
Data Scientist
- Developed an algorithm to calculate fair market values and assign price ratings to >1 million e-commerce listings
- Designed specialized reports based on industry data analysis for internal use and external commercialization

Mass General Hospital, Department of Neurosurgery   *Sep 2018 – May 2020*
Carter-Balaj Research Lab Member   *Boston, MA*
- Engineered a real-time minimally invasive brain cancer diagnostic that identifies mutant mRNA in glioma cells
- Selected as a national Regeneron Science Talent Search Scholar for my research

TECHNICAL SKILLS & PROJECTS

Advanced: Python, R, PowerBI; **Proficient:** SQL, C++, React
ML Techniques: NNs, CNNs, RNNs, transformers, LLMs, RAG; regression, random forests, XGBoost, PCA, clustering

Enhancing Satellite Image Classification with Non-Image Data   *Spring 2025*
- Extracted geospatial variables from Google Earth Engine (e.g., elevation, NDVI) to complement satellite imagery
- Developed a custom bimodal architecture that fuses the ResNet output layer with the structured non-image data
- Set a new best-performing benchmark on the EuroSAT dataset (98.69%) for land use and land cover classification

gAIn: An AI-Powered Health and Fitness Assistant   *Fall 2024*
- Co-developed a personalized fitness platform with an interactive chatbot to provide expert-level recommendations
- Fine-tuned Gemini and designed a RAG pipeline to combine expertise with user data (via wearable device APIs)
- Deployed the application on a Kubernetes cluster through GCP to ensure scalability across user systems

Sentiment Analysis of Audio Recordings   *Spring 2024*
- Predicted a speaker's emotion, out of 8 total, with deep learning models trained on 1,500 labelled recordings
- Achieved >50% accuracy with a custom lightweight transformer and >60% by fine-tuning Meta's Wav2Vec2

EXTRACURRICULARS

Harvard Financial Analysts Club
- Developed a foundational knowledge in accounting, financial statement analysis, and valuation methodologies
- Participated in stock pitch competitions, honing skills in researching, assessing, and presenting investments

Harvard Undergraduate Think Tank (Nuclear Advocacy Project)
- Conducted research and co-authored a policy memo using economic, technological, and historical analyses
- Networked with leaders in the field to coordinate a high-impact speaker event surrounding my team's work