Manraj Mondair

559-862-5139 · manraj@stanford.edu · linkedin.com/in/manrajmondair · github.com/manrajmondair

Education

Stanford University **Expected June 2028**
Bachelor of Science in Electrical Engineering and Computer Science *GPA: 3.9/4.0*
Activities: Neurotech Club (President), Google Developer Club, Stanford Consulting, Product Management Club (Leader), ASES/BASES
Awards: Coca-Cola Scholar (0.14%), Bank of America Student Leader (6%), California Science & Engineering Fair Finalist, Yale Research Award, President's Lifetime Achievement Award (4,000+ service hours), FIRST Robotics World Finalist
Relevant Coursework: Data Structures & Algorithms, Computer Organization & Systems, OOP, Machine Learning

Technical Skills

Languages: Python, C/C++, Java, TypeScript, JavaScript, SQL, Bash, R, MATLAB
Backend & Web: FastAPI, Flask, Node.js, Express, React, Next.js, REST APIs
Cloud & Systems: Linux, Docker, Kubernetes, Terraform, AWS (EC2, S3), GCP, Apache Kafka, Git, CI/CD (GitHub Actions)
Databases & Testing: PostgreSQL, Redis, MongoDB, Vector DBs (Pinecone, Milvus), BigQuery, Pytest, JUnit, Datadog
ML & AI: PyTorch, TensorFlow, Scikit-learn, ONNX, LangChain, LlamaIndex, Generative APIs (OpenAI, Gemini, Perplexity)

Work Experience

Salesforce (Slack) **Summer 2026**
Incoming Software Engineer Intern *San Francisco, CA*
- Building trusted agentic infrastructure for grounded retrieval, reasoning, and secure execution on enterprise data

Neural Interaction Lab, Stanford University **May 2025 – Present**
Software Engineer | Python (PyTorch, PyEDFlib, MNE, NumPy), CNN, Jupyter Notebook *Palo Alto, CA*
- Reducing scoring time 90%+ (50→5 min/case) by training 1D attention **CNN** (**PyTorch**) to auto-flag obstructive sleep apnea (OSA) from polysomnography (PSG) windows
- Lowering preprocessing 45→6 min (−87%) & eliminating leakage with deterministic PSG/EDF→tensor pipeline
- Cutting reruns 18%→5% (−72%) & speeding iteration ∼2.3× via schema validation + unit tests for per-window risk scoring

Aplazo (Series B FinTech Startup) **June 2025 – September 2025**
Software Engineer Intern | Python (Pandas, Scikit-Learn, Pyarrow), SQL, AWS *San Francisco, CA*
- Built **agentic LLM** pipeline for loan fraud detection, raising anomaly AUC 0.68→0.79 & F1 score 0.42→0.55 by assigning 0–1 risk scores to **BigQuery** events using **SQL**
- Cut API failures 3.8%→0.9% (−76%) & P95 latency 850→320 ms (−62%) via retries + backoff, throttling, compact JSON prompts, raw-response capture, and strict JSON schema
- Reduced fraud-ops triage 35→12 min/case (−66%) with ROC/threshold sweeps, K-Means triage clusters, top-K flagged exports, and per-event reports

McLean Hospital, Harvard Medical School **August 2022 – March 2023**
Software Engineer | R Shiny, SQL, REDCap API, PostgreSQL *Boston, MA*
- Cut data retrieval 4:00→0:20 (−92%) for 50+ researchers with **R Shiny/PostgreSQL** dashboard via **REDCap API** for an 800-record trauma registry (supports $3.2M NIH/NIMH)
- Reduced P95 page load 2,200→900 ms (−59%) via indexing & incremental sync (replaced CSV merges)
- Increased export 3→8/min (+167%) & lowered errors 4.8%→1.2% via schema validation + unit tests end-to-end

Safe & Intelligent Autonomy Lab, University of Southern California **May 2022 – August 2022**
Software Engineer | Python (PyTorch, TensorFlow), CNN, Computer Vision *Los Angeles, CA*
- Raised navigation success 62%→81% (+19 pp) by deploying **PyTorch CNN** perception stack & optimizing GPU P95 inference 48→22 ms to sustain 30 Hz control
- Cut CI time 24→9 min (−62%) & raised coverage 54%→82% by modularizing 10k+ LOC & parallelizing unit integration suites in **GitHub Actions**
- Reduced experiment reproduction time 4.5→1.2 hrs (−73%) & eliminated framework drift by unifying preprocessing + seed control across **TensorFlow** to **PyTorch** (with FAIR)

Research Experience

Argonne National Laboratory **August 2023 – March 2024**
Artificial Intelligence Researcher | Python (PyTorch, TensorFlow), LLM (GPT), Colab *Lemont, IL*
- Built **Python** data pipelines for preprocessing quantum computing literature for generative AI experiments
- Developed and tested model components for **Aurora**, supporting LLM experimentation and scientific AI workflows
- Partnered with **Argonne** and **Intel** researchers on large-scale machine learning research systems

Princeton University **September 2021 – June 2022**
Computer Science Researcher (Liberty Research Group) | Python, NLP, LLM *Princeton, NJ*
- Accelerated analysis 45→110 docs/hr (+144%) by building structured dataset from 300+ interviews w/ **Python/NLP**
- Raised tool-ID accuracy F1 0.71→0.86 via LLM extraction + validation pipeline, cutting manual labeling by 78%
- Presented 12-page IEEE paper at SC24, leading data analysis/results & documented evaluation for reproducibility

Projects

Zippr: Brain-Controlled Robotic Arm (1st Place, Science Corp Hackathon) | [GitHub Repo] **Apr 2026**
Python, PyTorch, C++, RNN, ONNX, FastAPI, IKPy, Streamlit
- Built an end-to-end brain-to-robotic-arm interface that decoded **64-channel** neural recordings into continuous **7-channel** motor commands for real-time control of a **6-DOF** robotic arm
- Trained and deployed a **2-layer GRU RNN** on-device via **ONNX**, achieving $\mathbf{R^2 = 0.847}$, **97.1%** gate accuracy, and **˜16 ms** inference latency on Science Corp's SciFi platform

Sniff: AI Shopping Agent | [GitHub Repo] **Feb 2026**
Next.js, React, TypeScript, OpenAI, Bright Data, Browserbase, Server-Sent Events
- Engineered an agentic shopping assistant that used multi-turn query refinement and multi-retailer search to surface relevant listings across Google Shopping sources in real time
- Built a parallel fraud-detection pipeline with **5** concurrent checks per listing—including **WHOIS**, **Google Safe Browsing**, **community sentiment**, **brand impersonation**, and **page red-flag** analysis—streamed live via **SSE** with trust scoring and best-pick recommendation