Tarek Badih

(510) 570-4578 | tbadih2004@berkeley.edu | LinkedIn | Portfolio Website | GitHub

Summary

ML engineer with production experience building and deploying deep learning systems including computer vision, NLP pipelines, and generative models using PyTorch and TensorFlow. Skilled at implementing end-to-end ML pipelines from data ingestion through model training and real-time inference. Passionate about generative models, multimodal AI, and physically grounded simulation, with hands-on experience applying biologically inspired algorithms to complex real-world systems.

Education

University of California, Berkeley Dec 2025
Bachelor of Arts, Data Science — ML, Statistics, Optimization *Berkeley, CA*

Skills

Languages: Python, SQL, C, JavaScript, Java
ML Frameworks: PyTorch, TensorFlow, scikit-learn, JAX (familiar)
ML & Data: pandas, NumPy, Matplotlib, Seaborn, Feature Engineering, Distributed Data Pipelines
Domains: Computer Vision, NLP, Generative Models, Recommendation Systems, Graph Algorithms, Biologically Inspired Computation
Systems: Automated ML Pipelines, Real-Time Inference, Model Deployment, A/B Testing, Prompt Engineering
Tools: Git, Jupyter Notebook, VS Code

Experience

Helping Hands Apr 2025 – Aug 2025
AI & Machine Learning Intern *Hybrid, Berkeley, CA*

- Designed and trained deep learning recommendation models with end-to-end pipelines from data ingestion through real-time inference deployment, improving access to resources across underserved communities.
- Deployed a computer vision classifier for image categorization, reducing manual review time by **40%** through scalable inference integration into backend systems.
- Integrated NLP sentiment models for real-time message processing, designing the inference pipeline and downstream decision logic for automated, targeted outreach.
- Built automated ML pipelines in Python covering ingestion, cleaning, feature engineering, and model retraining, delivering analyst-ready datasets daily on production data.

BLCK UNICRN May 2024 – Aug 2024
Data Science Intern *Los Angeles, CA*

- Trained random forest and gradient boosting models on large-scale behavioral datasets, achieving **92% predictive accuracy** for user behavior forecasting with rigorous feature engineering.
- Built audience segmentation and dynamic pricing models for VR experiences, contributing to a **12% increase in ticket sales** through data-driven product decisions.
- Designed and analyzed A/B tests on promotional strategies, applying statistical rigor to validate model-driven decisions and lifting average ticket revenue by **8%**.
- Processed large-scale datasets using SQL and Python with rigorous feature engineering to improve model reliability and training efficiency.

Projects

Biodesign Challenge 2024 — UC Berkeley Finalist — Project Link Jun 2024 – Jul 2024

- Implemented slime mold network simulations in Python as a biologically inspired generative model of physical network dynamics, applied to optimize real-world transit routing.
- Combined 3D city modeling with graph optimization algorithms to simulate and evaluate transport strategies, presenting findings to academic and industry experts at the national competition.
- Demonstrated that simple, scalable simulation methods grounded in physical principles can generate novel solutions to complex infrastructure problems, directly aligned with world modeling research.

AI Powered Food Logistics — ENJAZ Tech Competition (3rd Place) Nov 2023 – Dec 2023

- Led development of an AI-driven routing and demand prediction system using clustering algorithms and shortest-path heuristics to model real-world logistics dynamics.
- Built an NLP pipeline using prompt engineering to extract structured insights from unstructured datasets and automate documentation workflows.