

Qamil Mirza

(+1) 341-388-6936 | qamilmirza@berkeley.edu | <https://www.linkedin.com/in/qamil> | [portfolio](#)

EDUCATION

University Of California, Berkeley

Berkeley, CA

Bachelor of Arts in Data Science & Statistics; GPA: 3.97

August 2023 – May 2027

Relevant Coursework: Foundations of Data Science, Data Structures, Discrete Math & Probability, Multivariable Calculus, Abstract Linear Algebra, Real Analysis, Computational Chemistry, Data Science for Biology

EXPERIENCE

Machine Learning Research Assistant

Jun 2025 – Present

UC Berkeley PATH (Partners for Advanced Transportation Technology)

Berkeley, CA

- Built an active-learning CV pipeline (YOLOv12 → frame triage → Label Studio → retrain) for road-distress detection (potholes, vertical/lateral cracks, faded markings, manholes, blurred signage).
- Processed **50,000+** frames; auto-filtered candidate frames and standardised ontology/QA guidelines, reducing manual review time by **80%** per video.
- Achieved **0.994 mAP@0.5** on a **7k image** internal validation set for road-distress detection via iterative retraining and label-quality improvements.

AI Search Optimization Engineer

June 2025 – Aug 2025

Maxis Berhad

Kuala Lumpur, Malaysia

- Developed a fully customizable spin-the-wheel React application, now **deployed in 15,200+ Maxis retail stores** nationwide to support customer engagement campaigns
- Designed and deployed a serverless AWS-based document storage platform, **migrating 1.5 million documents** across multiple departments, eliminating third-party dependency and **reducing annual infrastructure costs by RM100,000+**
- Hardened backend APIs through comprehensive security audits, remediating high-risk vulnerabilities including SQL injection points and outdated cryptographic algorithms (MD5, SHA1)
- Rewrote document parsing logic and integrated a new parser into the company's internal AI assistant, improving dependency injection for scalable self-hosting and **enhancing knowledge access for 3,800+ employees**.

Machine Learning Research Assistant

Jan 2025 – May 2025

Merck Group

Berkeley, CA

- Built a modular, end-to-end model development pipeline by organising the project on GitHub and scripting automated workflows for data extraction, feature engineering, model training, and evaluation
- Engineered Directed Message Passing Neural Networks (D-MPNNs) to predict drug potency (IC50/EC50), leveraging protein–ligand interaction fingerprints; **results submitted to the NeurIPS 2025 AI4Science Workshop**
- **Streamlined molecular docking workflow for 18,000+ ligands**, optimizing docking parameters to reduce protocol noise by **25%** and significantly improving signal quality in downstream modeling

PROJECTS

ML-Guided Directed Evolution of Activation Domain | Python, PyTorch, Transformers

April 2025

- Trained regression models (Random Forest, Gradient Boosting) on amino acid features and ESM-2 embeddings to predict gene activation strength of intrinsically disordered 40-amino-acid protein sequences
- Engineered feature representations using net charge, hydrophobicity, motif counts, and per-residue disorder to combine domain-specific priors with learned embeddings
- Implemented in silico directed evolution to iteratively mutate low-activity sequences, filtering candidates by predicted activity and mean disorder score ≥ 0.5 for downstream selection.

Optimizing Equity Derivative Hedging Strategies | Python, Gurobi, Pandas, NumPy, Jupyter

Feb 2025

- Designed a Mixed-Integer Linear Programming (MILP) model to hedge equity derivatives across SPY, QQQ, and IWM, ensuring a minimum +\$10M exposure under all market scenarios
- Minimized daily and cumulative option premium costs while maintaining non-negative exposure throughout a simulated trading horizon
- Awarded **2nd place out of 20+ teams** in the Berkeley IEOR + Wells Fargo BADSS Case Competition for technical rigor and practical feasibility of the solution
- Set up modular codebase with Conda environments, shell scripts, and reproducible result pipelines for collaborative model development and analysis

TECHNICAL SKILLS

Languages: Python, JavaScript/TypeScript, SQL, Java, Go

Frameworks & Tools: React, Node.js, Flask, FastAPI, Docker, Jest, Git

ML & Data: PyTorch, TensorFlow, Scikit-Learn, Gurobi, pandas, NumPy, ESM-2, MetaPredict

Cloud & DevOps: AWS (EC2, S3, Lambda, Step Functions, DynamoDB, RDS, EFS, Amplify, Load Balancer), Supabase, Docker

Interests: Backend Systems, Distributed Systems, API Design, Databases, Observability, Performance, Reliability