Amine Jelloul

510-313-8230 | amine_jelloul@berkeley.edu

Technical Skills

Machine Learning & AI: LLMs, multimodal models, representation learning, fine-tuning, evaluation design, experimentation, error analysis
ML Systems & MLOps: model deployment, monitoring, telemetry, experiment tracking, versioning, data validation, scalable training pipelines
Software Engineering: Python, backend systems, API design, modular architecture, testing, performance optimization
Data & Infrastructure: SQL, relational databases, feature pipelines, cloud ML systems, containerized deployments
Product & Collaboration: production ownership, cross-functional development, iterative model improvement

Professional Experience

Machine Learning Engineer — San Francisco, CA
Docugami — *Jan 2025 – Present*

- Designed and productionized LLM/VLM pipelines for document understanding and reasoning over complex PDFs, processing millions of documents and pages in production workflows.
- Built scalable multimodal systems combining OCR, layout parsing, retrieval, and transformer models to support structured extraction and question answering across heterogeneous document formats.
- Led evaluation and MLOps workflows including metric design, experiment tracking, dataset iteration, and prompt/retrieval optimization, improving task reliability by 30%+ and reducing failure cases in production.
- Implemented robust backend inference pipelines with structured outputs, versioning, and monitoring, enabling reliable deployment of AI features used by enterprise customers.

Machine Learning Engineer — San Francisco, CA
TextMe — *Sept 2024 – Jan 2025*

- Developed and deployed personalization and user segmentation models impacting millions of users, increasing engagement by 15%.
- Designed A/B experimentation pipelines evaluating ML-driven features, driving +10% ad conversion and +5% revenue uplift.
- Productionized ML experimentation workflows with Python and Weights & Biases, accelerating iteration cycles and improving reproducibility across models.

Machine Learning Researcher — Berkeley, CA
NeurIPS (Supervised by Yunkai Zhang) — *Aug 2023 – Jan 2024*

- Developed *Insight Miner*, a multimodal model generating domain-specific time-series insights from a dataset of 100k time-series/text pairs.
- Integrated large language models with classical time-series decomposition, outperforming baseline multimodal systems including LLaVA-style approaches.

Machine Learning Engineer — Paris, France
Orange — *Sep 2022 – Aug 2023*

- Built PyTorch forecasting systems for 100k+ telecom time series predicting network congestion, enabling optimized infrastructure planning and $1M+ annual savings.
- Reduced large-scale data processing time by 50% through optimized PySpark pipelines.

Software Engineering Intern — Paris, France
TOPI — *Jun 2022 – Aug 2022*

- Developed FastAPI backend serving 15k+ daily requests and integrated ML services into production APIs.

Data Scientist Intern — Paris, France
Acolyt — *Jun 2021 – Aug 2021*

- Built forecasting models reducing logistics costs by 20% and redesigned PostgreSQL schema improving processing performance by 50%.

Education

UC Berkeley — Berkeley, CA
Master's Degree in Operations Research — *GPA: 4.0*

Arts et Métiers Institute of Technology — Paris, France
BSc & MSc in Operations Research — *GPA: 4.0*