

Karim Semaan

AI / ML Engineer

Boston, MA | semaankarim02@gmail.com | +1 (617) 606-0207 | [LinkedIn](#) | [GitHub](#) | karimsemaan.me

AI / ML engineer shipping production LLM pipelines and calibrated ML systems; AWS ML Specialty certified; M.S. in AI (Northeastern) in progress.

EXPERIENCE

BDO · Achrafieh, Lebanon Jun 2025 – Present (remote, part-time)

AI Solutions Lead

- Built and operate PayStream, an automated payroll platform now used daily by Fortune 500 FMCG and pharma payroll teams for reporting (technical detail under Selected Projects).
- Cut reporting time about 80% for 350+ employees by rolling out a real-time time-reporting and analytics platform.

Vocadian · Cambridge, MA Sep 2024 – Jun 2025

DevOps Engineer

- Accelerated deployments about 50% by building CI/CD pipelines with GitHub Actions.
- Closed off credential-leak risk for SOC 2 compliance by migrating secrets to AWS Secrets Manager.
- Migrated 10+ production databases to AWS RDS with zero data loss.

Orna Therapeutics · Watertown, MA Jan 2023 – Jul 2023

AWS Cloud / DevOps Engineer Co-op

- Reduced data discrepancies about 15% for reporting dashboards by building AWS Lambda data-validation pipelines.
- Streamlined Kubernetes releases by optimizing GitLab CI/CD and automating infrastructure with Terraform.

Harvard University · Cambridge, MA Jan 2022 – Jul 2022

Cloud & IT Analyst Co-op

- Resolved 3,000+ ServiceNow tickets and 600+ Active Directory incidents across a university IT estate.
- Deployed 1,000+ machines by writing PowerShell automation.

SELECTED PROJECTS

KickCast · World Cup match prediction, ensemble ML + Monte Carlo [kickcast-dashboard.vercel.app](#)

- Calibrated a tuned XGBoost match model, cutting 2022 World Cup holdout log-loss 1.347 to 1.093 and 10-bin ECE 0.157 to 0.120 (full test: ECE 0.098 to 0.017, n=3,552), via per-class isotonic regression fit on the validation split, holdout untouched.
- Trained 7 classifiers plus an Optuna-tuned PyTorch net on a leakage-safe 21,371-match x 38-feature matrix with chronological splits, optimizing log-loss; attributed signal with SHAP.
- Simulated the 48-team, 104-game 2026 World Cup with 10,000 Monte Carlo iterations over calibrated probabilities, served as a live Next.js dashboard.

Bastion · AI security-assessment SaaS live, invite-only

- Lowered LLM token spend about 80% versus full-context analysis by designing a multi-stage Claude pipeline: local section filtering, Haiku relevance triage, Sonnet deep gap analysis, Haiku self-verification, and Pinecone RAG.
- Automated security gap analysis across 5 frameworks (NIST CSF 2.0, CIS v8, ISO 27001, SOC 2, CMMC 2.0); every finding cites its source evidence and requires human sign-off before reaching a client report.
- Enforced consultant/client tenant isolation at the database layer with Supabase row-level security.

INTELIPS · LLM-labeled email-priority NLP github.com/karimsemaan/PAEPS

- Reached 72.85% macro-F1 (weighted ensemble) against a 72.18% XGBoost baseline by benchmarking five architectures (TF-IDF XGBoost, context MLP, BERT + attention, fine-tuned BERT, user-embedding net) in PyTorch / scikit-learn.
- Generated training labels for 25,640 Enron emails (3 priority classes) with an LLM served via the Groq API; explained predictions with SHAP and LIME.

PayStream · enterprise payroll ETL in production (BDO clients)

- Engineered the deterministic ETL core: FastAPI + pandas/openpyxl pipelines convert vendor payroll exports into bank payment files, GL/variance reports, and accruals, with multi-currency support.
- Hardened the platform with JWT/bcrypt auth, RBAC, rate limiting, audit logging, and multi-database support; shipped with Docker.

EDUCATION

Northeastern University · Boston, MA Sep 2025 – May 2027 (expected)

M.S. Artificial Intelligence, Machine Learning concentration

- Graduate coursework: Deep Learning, Natural Language Processing, Reinforcement Learning.
- **B.S. Computer Science & Mathematics**, GPA 3.70, Dean's List, NU Partners Scholarship.

CERTIFICATIONS

AWS Certified Machine Learning – Specialty January 2025

AWS Certified Solutions Architect – Associate July 2024

SKILLS

ML / AI: Python, PyTorch, scikit-learn, XGBoost, LightGBM, BERT / Transformers, isotonic calibration, SHAP, Optuna, pandas, NumPy, SageMaker

LLM & Full-Stack: Claude API, RAG (Pinecone), Next.js, React, TypeScript, FastAPI, Supabase, PostgreSQL, REST APIs

Cloud & DevOps: AWS (EC2, S3, Lambda, EKS, CloudFormation, CDK), Docker, Kubernetes, Terraform, GitHub Actions, Linux, Git