Aditya Rathi

Machine Learning Engineer
(215) 930-6153 | aditya.rathi2000@gmail.com | linkedin.com/in/aditya-rathi | adityarathi.com | New York, NY

EDUCATION

University of Pennsylvania, Philadelphia, PA *May 2023*
M.S.E., Data Science; B.S.E., Electrical Engineering, Summa Cum Laude
Key Coursework: Principles of Deep Learning, Artificial Intelligence, Graph Neural Networks, Applied Machine Learning

PROFESSIONAL EXPERIENCE

Burning Glass Institute, *Machine Learning Engineer* | New York, NY *June 2023 – Present*
First engineering hire at a labor-market research institute; tech lead for matching pipelines and multiple models in production

- Lead the upstream ML layer of the **Credential Value Index** (credentialvalueindex.org), one of the nation's most comprehensive public indices of non-degree credentials that states like Idaho have used to assess programs for LAUNCH grant funding. Designed entity resolution across ~20 sources (blocking, FAISS embedding similarity, LLM pairwise adjudication, graph clustering), growing canonical providers **~35% YoY**, and an LLM cascade that generated **~4.3M** provider descriptions, recovering ~77% of the hardest residuals with grounded search.
- Built a multi-domain classifier for **600M** unlabeled education records (university, K-12, international taxonomies); domain-specialized models achieving **93% accuracy at 94% coverage** on the 165M-record US slice; **+19% F1** from task-adaptive fine-tuning over a frozen encoder; shipped with CI/CD and drift-triggered retraining.
- Ran domain-adaptive pretraining across **8 encoder architectures** using an agent-driven experiment loop with automated HPO, comparing models on bits-per-byte as a tokenizer-agnostic metric. The selected ModernBERT checkpoint is the backbone of the postings extraction stack, served with FP8 quantization, FlashAttention-2, and minimal-padding batching on RTX 5090s.
- Designed a tiered extraction cascade for job postings. A fine-tuned ModernBERT NER model segments postings into 8+ section types (education, experience, skills, certifications, responsibilities, etc.), feeding several downstream classifier models. NER labels distilled from an LLM over a weak-supervision-stratified sample with intern audits (~88% F1 vs. gold). **440M** education predictions shipped, validated by LLM-as-judge.

Sophos AI Research, *Data Science Intern* | Philadelphia, PA *June 2022 – December 2022*

- Extended a cybersecurity alert-prioritization system limited to numerical/temporal signals with learned text representations of command lines and file paths (Sentence-BERT vs. ELMo A/B, domain-specific tokenization). Early fusion with temporal behavioral features gave a **16% AUC lift (XGBoost)** over the text-free baseline across 282K alerts from 3.2K sensors.
- Shipped to production, where the system cut analyst queue time by **23%** and suppressed **54%** of false positives at 95.1% detection.

PROJECTS

Manifold, *Construction Image Classification API* | app.manifold.ae *2026*

- Shipped a hosted API classifying construction-site photos by trade and phase (structural, MEP, finishing, sitework) for a paying property-developer client in Dubai. Encoded the construction taxonomy into prompting for a Qwen3.5-9B; served on AWS SageMaker with API-key auth, usage-based pricing, published docs, and sub-2-second inference.

Kalshi Weather Trading Bot, *Live, trading real money* *2026*

- Trades Kalshi's CFTC-regulated daily high-temperature contracts across 20 U.S. cities: Proprietary model trained on CRPS blends multiple national weather prediction and AI weather models into calibrated bracket probabilities, traded against market odds with 1/8-Kelly sizing and correlated-cluster exposure caps. Live ECE < 6.8%.
- Pre-registered, leakage-safe walk-forward backtests with block-bootstrap CIs decide model changes; a statistical gate governs shadow-to-live promotion; fail-closed execution reconciles orders with the exchange (4.5K+ tests).

BrainAI, *EEG/ECG Biometric Identification* | Course Project, University of Pennsylvania *2022*

- Closed-set identification of 200+ subjects: bandpass, FFT, wavelet, PSD features; LR/FNN/LSTM, stratified k-fold CV (77% F1).

SKILLS

AI/ML: PyTorch, Hugging Face Transformers, supervised fine-tuning, domain-adaptive pretraining, NER, text classification, entity resolution, XGBoost/LightGBM, scikit-learn
Method Selection: regex/rules → classical ML →fine-tuned encoders → LLMs, chosen by accuracy, cost, and scale
LLMs & Retrieval: OpenAI, Gemini (batch + search grounding), Claude, embeddings, FAISS, LLM-as-judge evaluation
Inference & MLOps: FP8 quantization, FlashAttention-2, FastAPI, Docker, CI/CD, drift monitoring, AWS SageMaker
Data & Tooling: Python (expert), SQL, PySpark, Snowflake, AWS, GCP Vertex AI, Linux, Claude Code, Codex, Cursor