Aditya Mittal

New York, NY, USA | 1-347-206-6415 | theadityamittal@gmail.com | LinkedIn | GitHub

Summary

Software engineer specializing in machine learning and cloud infrastructure, with hands-on experience driving end-to-end AI solutions. Led the design of a multi-stage ML data pipeline and LLM evaluation harness on SageMaker and Bedrock, improving coaching quality by 42% and cutting system latency by 41%. Delivered an automated onboarding agent on AWS Lambda and Pinecone RAG, significantly reducing intern ramp-up time and manual workloads. Ready to leverage a track record in building robust, scalable AI systems to enhance next-generation product performance.

EXPERIENCE

NovumAI | *Software Engineer* **Nov 2025 - Present**

- Constructed a 6-pass ML data pipeline on SageMaker converting 13K BPO transcripts into 30K SFT training pairs via Bedrock Batch diarization, stage labeling, and concurrent NEPQ rewriting; quality gates held a 98% pass rate.
- Designed an 8-dimension LLM-as-Judge evaluation harness and QLoRA fine-tuned Llama 3.1 8B on 30K domain pairs via SageMaker; combined with prompt optimization, composite coaching quality improved 42% over the production baseline.
- Deployed fine-tuned Llama 3.1 8B via Bedrock CMI and replaced the monolithic coaching prompt with a dynamic stage-aware architecture validated against the 8-dimension eval harness; p50 suggestion latency fell 41% with zero idle cost.
- Architected a background reasoning system with SQS-triggered Lambda pre-computing emotional arc, objection tracking, and call stage into Redis; suggestions read cached state at sub-1ms overhead, lifting composite quality +0.5 points.

Changing The Present | *Software Engineer* **Jul 2025 - Oct 2025**

- Developed a Slack onboarding agent that automatically creates personalized plans, assigns channels, schedules meetings, and tracks progress, reducing intern ramp time from about 5 to 2 days for 15 interns.
- Architected the onboarding agent on AWS Lambda + SQS FIFO with dual-LLM routing and hierarchical Pinecone RAG with 4-factor confidence scoring; grounded answer accuracy reached 85%.
- Built an event-triggered pipeline where Jotform webhooks are validated, cleaned, and transformed before loading into RDS PostgreSQL, eliminating manual Google Sheets tracking for over 10,000 applicant records

Engagebud | *Product Engineer* **Nov 2022 - Mar 2023**

- Engineered Thompson Sampling bandit over K-means visitor segments for real-time campaign routing, containerized inference with Docker, and deployed via GitLab CI with New Relic monitoring; clients hit 80% engagement.
- Launched the gamified widget runtime handling concurrent game sessions across Shopify storefronts with lazy loading and code splitting; platform campaigns delivered 17–35% CTR and up to 375% conversion lift for game participants.
- Partnered with founders and enterprise clients to translate requirements into scoped roadmaps, then led demos and retros across a 5-month engagement; landed 3 reference customers within 1 quarter and strengthened the renewal pipeline.

Juniper Networks | *Software Engineer Intern* **Jan 2022 - Jun 2022**

- Strengthened Mist AI's predictive failure pipeline by building PySpark telemetry ETL and feature extraction jobs with schema validation and data quality checks; early device failure recall improved 15% across production deployments.
- Operationalized Mist AI's LSTM by implementing PSI-based feature drift monitoring and automated retraining triggers on threshold breach; false positive alerts in predictive failure detection fell 20% across firmware release cycles.
- Developed a dual-mode Junos configuration tool in Perl and LISP: a validator that audited configs and flagged conflicting knobs before test runs, and a debugger that parsed failures and surfaced root causes; test failures dropped 30%.

EDUCATION

New York University | *Master of Science, Computer Science* **(GPA: 3.90)** **Sep 2023 - May 2025**

Manipal University Jaipur | *Bachelor of Technology, Computer Science* **(GPA: 8.86)** **Jul 2018 - May 2022**

SKILLS

- **Programming Languages:** Python, TypeScript, JavaScript, SQL, Bash
- **AI / Machine Learning:** Amazon Bedrock, SageMaker, PyTorch, Hugging Face, Pinecone, QLoRA fine-tuning, SFT, LLM-as-Judge evaluation, RAG pipelines, ReAct Framework, Prompt Engineering
- **Backend & APIs:** FastAPI, Node.js, WebSockets, OAuth 2.0
- **Cloud & Infrastructure:** AWS Lambda, API Gateway, SQS FIFO, ECS Fargate, EC2 Auto Scaling, KMS, AWS SAM, CloudFormation, Amazon S3
- **Databases:** PostgreSQL, DynamoDB, Redis
- **DevOps & Observability:** Docker, GitHub Actions, CloudWatch, New Relic

PROJECTS

Claude Professor | GitHub

- Built a developer-workflow Claude plugin preventing knowledge atrophy via state machine with FSRS-5 spaced-revisit scheduling; covers 400+ SWE concepts across 18 domains and 19 research-backed architectural concerns.
- Engineered a two-stage semantic retrieval pipeline over software architecture concepts with zero external npm dependencies; shipped 88 automated tests spanning contract, integration, and migration suites.

Sherpa | GitHub

- Generalized a bespoke onboarding bot into a distributable multi-tenant Slack platform; workspace isolation via DynamoDB and Pinecone namespaces enables any nonprofit to self-install via Slack OAuth with zero configuration overlap.
- Hardened with 4-layer cost protection and KMS field-level token encryption; 7-layer middleware split for Slack 3s timeout compliance with 90%+ coverage across 63 test files.