Wassim LaCorchy

New York, NY ◇ 571-603-8538 ◇ wassimlacorchy@gmail.com ◇ linkedin.com/in/wassimlacorchy ◇ github.com/LLwassim

EXPERIENCE

Lennar New York, NY

Principal Software Engineer - Applied AI ***February 2025 - July 2025***

- Co-led architecture for 'The Lennar Machine' dynamic AI pricing platform (Next.js, Prisma, AWS); CI/CD (Azure, GitHub Actions); 1% uplift (~$36M/yr). Cited in Q2 Earnings Report as 'Crown Jewel' of portfolio.
- Stood up Lennar Connect's GenAI platform (LangGraph multi-agent, GPT-4o/Azure) automating Tier-1 HR (PTO, ServiceNow, Workday) via governed RAG; led 9-engineer team; owned eval+prod telemetry; partnered w/ CTO on AI roadmap (use-case priority, success metrics, rollout).
- Built an eval framework for agent workflows (golden task suites + rubric scoring + regression gating) and production telemetry (quality + latency + cost) enabling iterative improvements without regressions.

Amazon Web Services (AWS) New York, NY

Senior Solutions Architect (ProServe) ***July 2021 - February 2025***

- Contributed to Amazon Compliance Screening's AI-powered investigation tier for sanctions screening at Amazon scale (~2B transactions/day), implementing multi-agent workflow patterns and tool interfaces.
- Implemented deterministic orchestration / dependency sequencing and exception handling (graph-style multi-agent execution) and enforced SOP compliance with confidence-based human escalation paths.
- Built/standardized MCP servers + agent tool integrations (data aggregation, geospatial/address validation, open-source intelligence lookups) enabling agents to gather evidence across internal/external systems.
- Led 16+ ProServe deliveries (discovery→prod) w/ VP/Dir stakeholders; produced reusable deployment playbooks/templates used across engagements.

Lasaria New York, NY

Co-Founder ***April 2024 - Present***

- Delivered production LLM agents (LangGraph, Azure OpenAI/AISearch) w/ eval loop (goldens, LLM-judge, regression gates) tracking quality, p50/p95, and cost across client deployments.
- React Native + Node/AWS, event-driven pipelines + real-time rankings; improved startup latency ~7.5× and added AI search (Pinecone/Bedrock).

Capital One Mclean, VA

Full Stack Engineer ***June 2018 – June 2021***

- Led end-to-end delivery of ML decisioning systems (fraud, credit risk, workflow automation) in regulated environments; owned production rollout, incident response, and performance optimization.
- Built MLOps: dataset/model versioning, CI/CD for model artifacts, drift monitoring, retrain + rollback; operationalized eval/risk tradeoffs w/ release gates.

SKILLS

Languages: TypeScript, JavaScript, Python, Java, SQL

Frontend: React, Next.js, React Native

Backend: Node.js, FastAPI, REST/GraphQL, microservices

AI/ML: tool calling, agents, eval harnesses, RAG, MCP; LangChain/LangGraph, Bedrock, Pinecone, AUC/PR

Cloud: AWS (Lambda, S3, DynamoDB, SQS, SNS, EC2, EKS, Fargate, IAM, API Gateway, Cloudwatch)
Azure (OpenAI, AI Search, Functions, Azure Pipelines)

DevOps: Docker, Kubernetes, Cloudformation, Terraform, OpenTelemetry, CloudWatch, Datadog

Data: Kafka, Redis, NoSQL, event-driven systems

Certifications: Scrum.org Professional Scrum Master I, PMI Certified PM, AWS Solutions Architect