THANOOJ LINGAMPALLY

New York, NY · Open to relocation (San Francisco preferred)
(716) 259-7927 · thanoojlingampally@gmail.com · linkedin.com/in/thanooj-lingampally · thanooj.com

SUMMARY

Founding backend and AI engineer who owns an agentic contract-intelligence product in production: the Claude agent loop, its honesty guards and evals, signature-field detection, and the AWS infrastructure under it. M.S. in AI. Previously built a RAG platform for 10K+ enterprise users at TCS.

TECHNICAL SKILLS

AI systems: Claude / Amazon Bedrock, agent loops and tool calling, evaluation harnesses, prompt caching, RAG, LangGraph
Backend / infra: Python, FastAPI, SQLAlchemy, PostgreSQL, DynamoDB, SQS, Docker, ECS Fargate, Terraform, GitHub Actions
Also: Java, Spring Boot, React / Next.js, PyTorch, MLflow, Apache Spark

EXPERIENCE

Founding AI Engineer | Dottr (VC-backed contract intelligence, Leo Capital) **Jan 2026 – Present**

Joined as the first engineer. Own the backend and agent for a contract-intelligence product in production with paying customers. Author of over half the repo's commits across **7** deployed services.

- Built and operate the Claude-on-Bedrock agent behind chat and email: tool dispatch across **40+** tools, prompt-cache-aware context assembly, a five-layer honesty guard that checks every claimed action against owned records, and an eval harness of recorded production flows replayed against every prompt and model change. Cut perceived chat latency from a **23s** median to under **2s** time-to-first-token by streaming replies while the guard audits concurrently and retracts on failure.
- Rebuilt signature-field detection as a propose-then-select pipeline: a **37** MB int8 ONNX object detector proposes candidates and Claude selects. Built the eval harness first (**793** documents, **6,567** human-reviewed boxes) and used it to catch a coordinate-transform bug that had put **30%** of candidates off-page.
- Audited the e-signing pipeline end to end and found **7** critical defects, including one that could complete a contract with a required signer never contacted. Led the fixes to prod and gated deploys on the test suite so a red build can no longer ship.
- Shipped multilingual document ingestion: benchmarked four OCR vendors on real scanned Indian legal filings, built a Textract-first router that detects garbled output and re-routes to Sarvam, and swapped the embedding model to Cohere multilingual after measuring that Titan could not bridge English queries to Telugu text. Backfilled **11k** chunks in prod with zero downtime.
- Rebuilt alert triage to fingerprint log signatures, collapsing **222** open GitHub issues into **42** root causes and cutting auto-filed noise by over **80%.** Ran the AWS cost audit that found **$2.1k/month** of spend hidden behind credits and removed roughly **$7k/year** of orphaned resources.
- Benchmarked every vendor decision head to head: document parsing (Docling beat Reducto on table fidelity, **2.2s** vs **3.4s** median latency), signature detection (Extend, Reducto), and placement models (Claude vs Gemini, type agreement **1.00**).

ML Engineer | Tata Consultancy Services **Dec 2022 – Aug 2024**

- Scoped and built a customer-facing RAG platform for **10K+** enterprise users, working directly with business stakeholders and meeting corporate security and compliance requirements.
- Built Spring Boot microservices with secure REST endpoints that connected legacy enterprise databases to ML pipelines.
- Led the MLflow-based MLOps pipeline for model versioning, tracking, and promotion to production.

Software Engineering Intern | Counselit **Apr 2021 – Aug 2021**

- Built Node.js/Express REST APIs and React interfaces for client pilots, including JWT authentication and role-based access control.

PROJECTS

Goal-Conditioned Quadruped Navigation with Mixture-of-Experts RL

- Built a custom MuJoCo environment and a Mixture-of-Experts policy in PyTorch, reaching a **92%** success rate on target-reaching tasks within **200** timesteps.

EDUCATION

M.S., Computer Science (Artificial Intelligence) | University at Buffalo, SUNY **Aug 2024 – Dec 2025**

Coursework: Natural Language Processing, Computer Vision, Reinforcement Learning, Data-Intensive Systems

B.Tech, Mechanical Engineering | National Institute of Technology Durgapur **2019 – 2023**

Work authorization: F-1 OPT, STEM extension eligible through Feb 2029.