

Yasaswin Palukuri

Toronto, ON | +1 437-433-7488 | palukuriyasaswin@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

Software engineer building full-stack AI applications with Python, React and FastAPI, including a self-hosted multi-agent LLM platform (since Jan 2026) and an LLM document-extraction service. IEEE co-author seeking a junior full-stack AI role.

SKILLS

Languages: Python, JavaScript, SQL, C++, Go, C, Java, Bash

Frontend: React, Vite, Tailwind CSS, server-sent events (SSE) streaming

Backend & Data: FastAPI, REST APIs, PostgreSQL, Snowflake, Redshift, SQLite, ChromaDB, dbt, PySpark

AI / LLM: LangGraph, prompt engineering, structured LLM outputs (Pydantic v2), RAG (hybrid BM25 + vector), Ollama, Ragas, LLM guardrails

Cloud & DevOps: AWS (S3, Glue, Lambda, Redshift), Docker, Kubernetes, GitHub Actions, n8n, Git, Linux (Ubuntu)

Testing & Process: Pytest, GoogleTest, unit/integration testing, Ruff, Bandit, clang-tidy, code review, Agile/Scrum

Currently Learning: Redux, LangChain

EXPERIENCE

Full-Stack Developer Intern — Gaddr

Remote · Oct 2026 – Present

- Working on the AI job recommendation feature for the Gaddr Jobs platform (React, React Router, TypeScript, Better Auth, Stripe, PostgreSQL).
- Delivered the master prompt behind the recommendation engine; it is in validation and being refined against real results.

PROJECTS

Groot — Full-Stack Multi-Agent AI Platform [\[GitHub\]](#)

Python, FastAPI, React, LangGraph, n8n, Docker

- Architected and operate a 6-agent LLM platform running 24/7 on Ubuntu as 4 Docker services (FastAPI backend, ChromaDB, n8n, React dashboard) auto-started by systemd, with streaming chat and live metrics.
- Engineered hybrid retrieval (40% BM25, 60% dense vector) across 6 ChromaDB collections with role-based access filtering; evaluated with Ragas at 0.84 overall and 1.0 faithfulness.
- Automated 8 n8n workflows, including a job-matching pipeline that pulls live listings from an external API, scores each for fit, generates tailored PDFs and delivers them via Telegram, tracking 34 applications to date.
- Designed an asyncio priority-queue scheduler that checks free RAM before each inference and re-queues jobs under a 3 GB buffer, eliminating out-of-memory crashes on CPU-only hardware.
- Gated every merge behind CI (Pytest, Ruff, Bandit), including 20 red-team tests for prompt injection and PII leakage.

Structured Data Extractor — LLM Document Parsing

Python, Pydantic v2, Ollama, PyMuPDF, pdfplumber

- Built a document processing and data extraction service converting free text and PDFs into Pydantic v2-validated JSON for 3 document types: job descriptions (12 fields), resumes and invoices with line items.
- Designed a two-stage classifier: keyword scoring (<10 ms) types each document and calls qwen2.5:7b only below 60% confidence.
- Enforced schema-valid output with temperature-0 inference and automatic JSON-only retries on validation failure; synced results to Notion with SHA-256 deduplication and rate-limited API calls.

DataPulse — Pipeline Monitor with Data Quality

Python, Snowflake, dbt, AWS, Great Expectations, FastAPI, Streamlit

- Built a daily batch pipeline loading NYC taxi data into a Snowflake star schema (1 fact, 5 dimensions via dbt), with Great Expectations flagging bad rows without halting the load and a parallel AWS Glue/PySpark path to Redshift.

RailSim — Railway Signaling Simulator [\[GitHub\]](#)

C++17, CMake, GoogleTest

- Designed an 8-component simulator (Dijkstra route planner, interlocking, PID speed control) and verified its safety invariant with 0 violations across 90 GoogleTest cases and 20,000 randomized requests.

Centralized Electronic Medical Records System [\[Paper\]](#)

IEEE publication

- Collaborated with 3 co-authors (2 students, 1 professor) from design through publication; designed the hashing-based storage and retrieval layer, enabling in-place record updates without full rewrites.

ADDITIONAL EXPERIENCE

Crew Member and Trainer — Shake Shack

Toronto, ON · May 2024 – Sep 2026

- Trained 20+ new team members in cooking, customer engagement, POS and service, including new recruits at New Shack Openings, and ran the shift for two days at one; served 100+ guests per shift on average.

EDUCATION & CERTIFICATIONS

Executive MBA, Data Science — KL University, India · 2025–2026 · CGPA 8.96/10

Post-Graduate Certificate, AI & Machine Learning — Lambton College, Toronto · 2024–2025 · CGPA 8.35/10

B.Tech (Honours), Computer Science — KL University, India · 2019–2023 · CGPA 8.98/10

Certifications: [Designing and Implementing Microsoft DevOps Solutions](#) (Microsoft) · Java J2EE (Wipro TalentNext)