Peter Atef

AI Engineer | Generative AI, LLM Applications, RAG, NLP, Agentic Workflow Automation

Giza, Egypt | +20 121 277 3495 | peter.atef2000@gmail.com

engpeteratef.github.io | linkedin.com/in/peter-atef-b31651231 | github.com/EngPeterAtef

SUMMARY

AI Engineer with two years of experience building LLM, RAG and agentic automation systems end to end, from requirements and architecture through deployment and stakeholder delivery. At Ejada Systems, built a code-understanding RAG system over a legacy COBOL codebase, a reinforcement learning planning system for a water desalination plant and a multi-stage LLM pipeline that generates test cases from business requirements; at Wego, delivered agentic workflow automations on n8n across HR, Engineering and Operations. Cairo University Computer Engineering graduate (AI specialization) with an Arabic NLP focus; native Arabic, fluent English.

EXPERIENCE

Wego | AI and Automation Engineer December 2025 to October 2026

Cairo, Egypt (Hybrid). MENA's leading travel metasearch platform with teams in Dubai, Singapore, Cairo and Bangalore.

- Built an agentic declined-refunds automation (n8n) that pulls refund records from web portals into BigQuery, posts a status summary to Slack and emails customers their current refund status; each step is exposed to the agent as a tool and the process is triggered through chat.
- Built an agentic HR contract renewal system (n8n, chat-triggered by HR) that queries the BambooHR API for contracts ending within 60 days, emails each manager a renewal decision form, reads decisions from Google Sheets and fills the contract template with employee data; 30 contracts per quarter, saving about 10 hours per month.
- Built an n8n pull request review pipeline that listens for PR creation on GitHub, sends the diff to three LLMs to summarize the change, flag bugs and draft review comments, then posts the consolidated review on the PR.
- Built a Confluence documentation agent (chatbot with GitHub access) that updates the module's Confluence page whenever a pull request is created, keeping technical documentation in sync with the codebase.
- Owned requirements, scoping and delivery with HR, Engineering and Operations stakeholders without a dedicated product manager, and reported progress, risks and outcomes to management across a multinational organization; used Claude Code and an OpenClaw agent instance, with prompt and context engineering, to accelerate delivery.

Technologies: Python, n8n, Robot Framework, BigQuery, FastAPI, REST APIs, GitHub, Confluence, Claude Code, OpenClaw, MCP, GitHub Actions, Git.

Ejada Systems | AI Engineer October 2024 to November 2025

Cairo, Egypt (Hybrid). Saudi Arabia headquartered enterprise systems integrator serving banking, government and industry.

- Built a reinforcement learning planning system (R&D) for a water desalination plant that learns the optimal times to open and close process valves and switch chemical treatments, maximizing water output at minimum operating cost and chemical consumption; implemented the deep RL approach proposed in a published study on this problem (reference paper).
- Built a code-understanding RAG system that lets software architects explore a legacy COBOL codebase spanning multiple applications: a question-answering chatbot over the code and similarity search between scripts, using a code-specific embedding model and a FAISS vector store self-hosted on the company's server.
- Designed its context-aware chunking strategy: each function and code block became its own chunk, and an LLM pass captured how every chunk relates to its parent script; complemented with a graph database of code relationships, auto-generated system documentation and an MVP re-implementation in Java and Python.
- Built a closed-box test generation framework that turns SRS and BRD documents into test cases before implementation: NER to extract variables, a RAG stage that derives each variable's constraints (length, range, format), and a second RAG stage that generates positive and negative test paths as step graphs.
- Completed the pipeline with an LLM data-generation module that produces valid and constraint-violating data for each path, and an LLM-based coverage evaluator that checks the generated cases against the business documents to reach at least 95% coverage; extended after implementation to generate unit test code for system modules.
- Built and maintained the FastAPI services and REST APIs behind these agent workflows, containerized with Docker; ran exploratory data analysis on numerical and text data to inform model, retrieval and prompt design.

Technologies: Python, LangChain, OpenAI API, Hugging Face, FAISS, ChromaDB, NER, FastAPI, Streamlit, Docker.

VNCR Media Group | Generative AI Engineer (Internship) July 2023 to October 2023

United Kingdom (Remote).

- Delivered a Blog Writer Agent end to end within the three-month internship: LangChain orchestration, OpenAI API generation, ChromaDB and FAISS retrieval over user-provided sources and a Streamlit interface; one request (title, target length, sources) returns a retrieval-grounded draft in minutes (GitHub).
- Worked fully remotely with a UK team across time zones; the delivery earned a formal recommendation letter from the company.

PROJECTS

- **Arabic Text Diacritization** (PyTorch, Python). Restored Arabic diacritical marks using two character-level sequence models: a Bi-LSTM with an embedding layer and the CBHG architecture (convolution bank, highway network, bidirectional GRU); reduces ambiguity for Arabic text-to-speech, machine translation and search (GitHub).
- **Vector Database Indexer** (Python). Built an index that returns the most relevant vectors for a query while minimizing memory and query time: IVF partitioning to narrow the search space, LSH for approximate nearest neighbors and fully vectorized operations; the retrieval layer that RAG systems depend on (GitHub).
- **KidAI, Graduation Project** (Unity, C#, Python). Educational platform that teaches the machine learning pipeline (dataset creation, training, validation, testing) to ages 8 to 18 through games; in the image classification game, children capture or import images, train a model and then play against it to see how it behaves. Unity and C# games with Python for the machine learning, built in an Agile team (GitHub).
- **Transaction Prediction** (scikit-learn, Python). Predicted whether a Santander customer will make a specific future transaction, regardless of amount; compared Linear Regression, AdaBoost, SVM, Random Forest and XGBoost (GitHub).
- **SBA Loan Approval Prediction** (PySpark, Python). Predicted approval or denial of U.S. Small Business Administration loan applications: Spark preprocessing, exploratory data analysis and association rule visualization, then KNN, logistic regression, random forest and SVM models (GitHub).
- **Hand Gesture Recognition** (OpenCV, Python). Recognized hand-drawn digits with HOG, LBP and PCA features and an SVM classifier; owned hand extraction with GMM and K-means and the LBP feature extractor (GitHub).
- **Virtual Calculator** (OpenCV, Python). Real-time detection of hand-shown digits and arithmetic operators with image processing and an SVM classifier at 90% accuracy in a multithreaded pipeline; owned image enhancement and accuracy evaluation (GitHub).

TECHNICAL SKILLS

AI Engineering: LLM application development, Retrieval-Augmented Generation (RAG), agentic workflows and tool use, prompt and context engineering, LLM evaluation, embeddings and vector search (FAISS, ChromaDB, IVF, LSH), Model Context Protocol (MCP), LangChain, OpenAI API, Hugging Face; LangGraph and CrewAI (working knowledge)

NLP and Machine Learning: named entity recognition, text classification, tokenization and embeddings, sequence models (LSTM, Bi-LSTM, CBHG), Arabic NLP and diacritization, semantic search, reinforcement learning, supervised learning, ensemble methods (Random Forest, XGBoost, AdaBoost), SVM, clustering (K-means, GMM), PCA, feature engineering, model evaluation, exploratory data analysis; PyTorch, scikit-learn, NumPy, Pandas, OpenCV, PySpark

Automation and AI Tooling: n8n (advanced), Robot Framework, Robotic Process Automation (RPA), Claude Code, OpenClaw, GitHub Actions, Microsoft Copilot Studio (basics)

Backend and Data: Python (expert), FastAPI, Flask, REST APIs, Docker, BigQuery, SQL, Git, CI/CD; integrations with BambooHR, Slack, Gmail, Google Sheets, GitHub and Confluence APIs

Professional: requirements analysis, stakeholder management, cross-functional delivery, technical documentation

EDUCATION

Cairo University, Faculty of Engineering | Bachelor of Computer Engineering (AI specialization) 2019 to 2024

GPA 3.7 / 4.0 (Excellent) | Coursework: Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Big Data | Military status: Exempted

Languages: Arabic (Native), English (Fluent)

CERTIFICATIONS AND TRAINING

Sprints and Microsoft Summer Camp, AI and ML (2025) | Udacity, Advanced NLP (2025) | NVIDIA DLI, Fundamentals of Deep Learning (2022) | ITI Egypt, AI Summer Training (2022) | NTI Egypt, AI Summer Training (2021)