

AHMED WAGDY MOHY

ahmedwagdymohy@gmail.com | +20 10032 07939 | [linkedin/ahmedwagdymohy](https://linkedin.com/in/ahmedwagdymohy)
github.com/Ahmedwagdymohy

EDUCATION

- **Faculty of Engineering | Cairo University 2021 – 2026** Bachelor of Electronics and Computer Engineering (EECE) senior with GPA: 3.1
- **Certificates:** [Google Professional Cloud Architect](#), [AWS Solutions Architect – Associate](#), [Google Associate Cloud Engineer](#), [AWS CloudOps engineer](#), [AWS Certified Cloud Practitioner](#), [KCNA \(Kubernetes and cloud essentials\)](#), MLOps-Kodecloud, AZ-900
- **Achievements:** 3rd Place ITIDA **DevOpsDays Hackathon** 2026, Top 10 Global Finalist – **Qt World** Competition

WORK EXPERIENCE

Incorta | New Cairo, Egypt

Aug 2025 – Present

DevOps / Platform / MLOps Engineer

AI Platform & MLOps:

- Deployed **LiteLLM** as the company-wide **AI gateway** for production and staging on GCP, backed by Cloud SQL (PostgreSQL), with centralized access control, per-team quota management, and usage tracking.
- Fixed LiteLLM dashboard crashes caused by high-volume usage logs by building a central LLM observability layer on **Langfuse**. It ingests **OpenTelemetry** traces from the dev and prod gateways across GCP projects over Private Service Connect (PSC).
- Built an **AI on-call agent** using **OpenClaw** and Claude Sonnet, integrated with **Slack** and **Telegram**, providing proactive incident detection, daily build monitoring, and conversational root-cause analysis with automated remediation.
- Designed an **AIops** API layer on **Cloud Run** functions that exposes cluster data from the core Incorta application to AI agents.
- Hardened an internal build of **OpenCode** (an open-source AI coding agent) by applying company security controls before rolling it out across the company.
- Provisioned **NVIDIA DGX** Spark machines as **GPU** nodes in a **K3s** cluster and set up **GPU** and node monitoring with **Prometheus** and **Grafana**.
- Configured dedicated **CPU** and **GPU** node pools on GKE for ML workloads, including **GoodMem** and **Reranker** services.
- Developed a full-stack internal **MLOps portal** (React, Python) with real-time API monitoring dashboards and automated quota request workflows.

Platform & DevOps:

- Built a browser-based, **Cloud Shell terminal** with multi-session support and a remote file browser for managing remote machines.
- **Automated Jira** issue reporting directly from remote machines, cutting the manual work of triaging and logging issues.
- **Reduced** AWS infrastructure **costs by 14%** through resource usage analysis and rightsizing.
- Operated production **Kubernetes** clusters running business-critical workloads, keeping production and testing environments reliable, scalable, and available.
- Built reusable **Terraform** modules for networking, Kubernetes, databases, and compute provisioning.
- Extended the internal DevOps portal with Python **AWS Lambda** functions and **step function** and **SSM** Agent integration, automating Incorta installation and configuration, daily build imports, and on-demand pipeline triggers.
- Set up observability with Prometheus and Grafana for proactive monitoring and alerting across CI/CD and infrastructure platforms.

DevSecOps:

- Designed a container **image security scanning** pipeline (Jenkins, **Trivy**) that generates CycloneDX and SPDX SBOMs for production images, laying the groundwork for supply-chain compliance and CVE traceability.
- Packaged the scanner as a reusable Jenkins shared library step, so any of ~25 delivery pipelines can add a security gate with one line, with severity gating (NONE/CRITICAL/HIGH) that supports report-only and blocking modes.

All In Functional Health LLC | Remote USA

July 2026 – Oct 2026

AI Automation Engineer & Full-Stack (Contract)

- Built the company's multi-tenant genomic evidence platform end to end using **Python, Flask, and Supabase (PostgreSQL and object storage)**, deployed on **Railway**. For any gene or disease, it returns ranked variants, related research papers, and per-study evidence profiles.
- Sped up bulk loading of large biomedical datasets using streaming parsers, chunked **PostgreSQL COPY** with per-chunk commits, and deferred index builds. A 644k-row load dropped from 2.5 hours to 25 minutes (6x faster), and release-based refreshes rebuild only the data that changed.
- Built **automated data pipelines** that pull from 9 public biomedical APIs and datasets (ClinVar, dbSNP, PubMed, OpenAlex, HGNC, UniProt, MANE, Rhea, GENCODE). They include per-variant risk scoring, full-text paper retrieval (**XML, JSON, PDF**), and strict data provenance rules.
- Built an **LLM** stage that classifies each study, with caching keyed on model and prompt version, so re-runs give the same results at no extra LLM cost. Pipeline run state survives crashes and page reloads mid-job.
- Protected sensitive health data with PostgreSQL row-level security (RLS), service-role-only access, private storage buckets with insert-only upload policies, and time-limited signed URLs.
- Shipped three patient intake web apps (Next.js, TypeScript) with multi-step forms, Zod validation across ~90 fields, private lab-file uploads, and email notifications via Resend, deployed on Vercel.

Inpro.sa | Remote KSA

Jun 2025 – Sep 2025

DevOps Engineer

- Designed and maintained CI/CD pipelines using Azure DevOps, implementing Gitflow workflows, branch policies, and automated deployment processes.
- Deployed ArgoCD with umbrella chart, resolving application OutOfSync issues to ensure seamless deployment processes.
- Developed observability dashboards and alerting solutions for Kubernetes clusters using Grafana, improving operational visibility and incident response.

EFG-Holding (EFG Hermes)

Oct 2024 – Dec 2024

SRE /DevOps Intern

- Implemented CI/CD pipelines using Azure DevOps to automate deployment and release processes.
- Configured and managed Nginx as a reverse proxy and load balancer.
- Integrated Prometheus, Grafana, and Jaeger to enhance observability.

Alpinum Consulting | UK (Remote)

Jun 2026 – Aug 2026

Software Engineering Instructor (Part-time)

- Taught real-time operating system fundamentals with FreeRTOS, including task scheduling, inter-task communication, and resource management for embedded devices.
- Delivered training on Zephyr RTOS for embedded and IoT applications.
- Built hands-on demos and lab projects hosted on Linux servers, giving students working environments for practical exercises.

PROJECTS

Sandbox Siege — AI Agent Security Testing Platform | DevOpsDays Cairo 2026

- built a chaos engineering framework that tests an AI agent's judgment before it gets real cloud access. Agents run normal tasks in simulated AWS environments (LocalStack Pro with IAM enforcement) seeded with realistic resources, permissions, secrets, and security traps.
- Designed a Python gateway (FastAPI, boto3) that intercepts every agent tool call, records it, forwards it to LocalStack, and runs security and behavior detectors. Each run gets two verdicts: whether IAM allowed the action and whether the action was wise.
- Turned agent scores into a pass/fail gate in GitHub Actions CI, with a real-time React/Vite dashboard that streams results over Server-Sent Events (SSE).

Serverless Kubernetes with Knative & Kourier

- Deployed serverless container workloads on Kubernetes using Knative Serving
- Implemented scale-to-zero autoscaling, reducing idle resource cost to zero

Event-Driven Autoscaling with KEDA on GKE

- Built an event-driven serverless platform using KEDA + Google Kubernetes Engine (GKE)
- Implemented Pub/Sub-based autoscaling, scaling deployments from 0 → 10 pods in seconds
- Provisioned infrastructure using Terraform (GKE cluster, VPC, IAM, Workload Identity)

3-Tier Web App on AWS EKS

- Frontend built with React, backend services using Flask, and PostgreSQL on RDS for persistent storage
Deployed on AWS EKS

Scythe Studio

- Qt Factory - C++ Desktop Application
- User authentication and role-based access
- Real-time sensor data dashboards with configurable thresholds
- Live graphing and weather data integration

SKILLS

Container & GitOps Orchestration: Kubernetes (AWS EKS, K3s, GKE), GitOps Workflows (FluxCD, ArgoCD), Cluster Operations, Bare-Metal & Edge Deployments, Docker Infrastructure as Code & Automation: Terraform, CloudFormation, Ansible, CI/CD Pipeline Design, Configuration Management, Automation Observability & Monitoring: Prometheus, Grafana, Loki, Tempo, OpenTelemetry, Alerting Infrastructure, Incident Response Security & Compliance Automation: Secrets Management (SOPS, KMS), Encryption, Access Controls, Audit Logging, Security Hardening, Linux Systems Administration ML Infrastructure Support