Oleksandr B.
Senior AI Solutions Architect

SUMMARY

- AI Solutions Architect with 5+ years of experience designing and delivering enterprise AI solutions, from stakeholder discovery to production deployment.
- Builds production multi-agent systems and RAG solutions using Python, LangGraph, FastAPI, PostgreSQL, Qdrant, pgvector, and LLMs.
- Experienced in integrations via REST APIs, webhooks, and OAuth, with role-based access, data privacy controls, and secure deployment.
- Designs agent orchestration, reusable modules, evaluation, tracing, and monitoring for quality, latency, and cost.
- Works with AWS, Azure, Docker, Kubernetes, Terraform, and serverless infrastructure; holds Microsoft certifications in AI solutions and cloud security.

TECHNICAL SKILLS

Main Technical Skills	Python, LangGraph, AI-agents, RAG, FastAPI, Azure
Programming Languages	Python, TypeScript
Android Libraries and Tools	ADB
AI & Machine Learning	AI-agents, AI Governance, Azure OpenAI, GenAI, LangChain, LangGraph, Large Language Models, LiteLLM, Microsoft Foundry, n8n, Prompt Engineering, RAG, Traceloop, Vertex AI, Voice AI Agents
Python Frameworks	FastAPI
Python Libraries and Tools	LiteLLM
JavaScript Frameworks	React
Security	AI Governance, Data Governance, OpenID, RBAC
Data Analysis and Visualization Technologies	Data Governance
Databases & Management Systems / ORM	PostgreSQL
Cloud Platforms, Services & Computing	AWS, Azure, GCP
Amazon Web Services	AWS Bedrock
Azure Cloud Services	Azure OpenAI, Microsoft Foundry

Google Cloud Platform	Cloud run, Google Workspace
SDK / API and Integrations	API, FastAPI, MCP Servers, OAuth, RESTful API, Webhooks
QA, Test Automation, Security	Authentication, Playwright
Scripting and Command Line Interfaces	Bash
Deployment, CI/CD & Administration	CI/CD
Platforms	CRM
Virtualization, Containers and Orchestration	Docker, Docker Compose, Kubernetes, Terraform
UI/UX/Wireframing	Figma
Mail / Network Protocols / Data transfer	Firewalls, OpenID, SSH
Version Control	GitHub
Logging and Monitoring	Grafana, Prometheus, Traceloop
Methodologies, Paradigms and Patterns	IaC, microservices
Operating Systems	Linux
Other Technical Skills	AI Orchestration, Anthropic Claude, Langfuse, OpenClaw, PGVector, Qdrant, Ragas, Raspberry Pi 5

WORK EXPERIENCE

AI Agents & Multi-Agent Systems - WorkHold, Senior AI Engineer

Feb 2026 - Aug 2026

AI startup building AI agents and digital employees that perform work inside companies.

- Owns the technical path from client requirements to running deployment, including requirements translation, solution architecture, implementation, rollout and operation of production AI agents.
- Develops LangGraph multi-agent workflows on asynchronous Python, FastAPI and PostgreSQL, using Temporal for durable orchestration, checkpointing and fault recovery on long-running jobs.
- Built the core modular agent engine from reusable modules assembled through per-client configuration. Each client runs a tailored agent without a codebase fork, turning custom engineering into configuration.
- Built a deployment orchestrator that delivers a fully configured agent to a new client in one click, standardizing rollout across the client base.
- Delivered Ragas evaluations and Traceloop tracing to track cost, quality and latency per agent and detect silent failure modes before customers encounter them.
- Implemented retrieval with Qdrant and pgvector and multi-model routing through LiteLLM, with an architecture portable to Azure OpenAI.

- Maintained strict module boundaries in a modular-monolith monorepo using uv workspaces and Turborepo, with documentation enabling another engineer to continue the work.
- Works directly with founders and product owners to turn ambiguous ideas into designs and explain architectural trade-offs. The CV describes the rapid R&D loop as 24 ideas → 6 strong products → 1 scalable winner.

Python, async programming, FastAPI, PostgreSQL, LangGraph, tool calling, Temporal, Qdrant, pgvector, LiteLLM, Ragas, Traceloop, uv workspaces, Turborepo. Additional role-listed technologies: Generative AI, RAG, AI orchestration, Cloud Security, Grafana, Prometheus, LLMOps, CI/CD, Infrastructure as Code, Terraform, Kubernetes, Anthropic Claude, LangChain, MCP, Google Cloud Platform, distributed systems, MLOps, microservices, REST APIs, vector databases, prompt engineering.

OptiAI.pro

Jan 2023 - Feb 2026

AI Solutions Architect

Jan 2023 - Feb 2026

Founded and ran a B2B AI automation consultancy end to end, covering audits, architecture, delivery and training for more than 10 paying clients across fintech, IT, retail and services.

- Delivered an enterprise RAG platform for a fintech with 1,300+ employees, combining a shared knowledge core with departmental sections and role-based access levels. The platform had 200+ monthly active users, integrated with two CRMs and PostgreSQL, and included automated API-spend monitoring and auditing.
- Ran structured discovery audits with business stakeholders, translated ambiguous needs into prototypes, identified high-ROI automation opportunities, and scoped and quoted solutions against measurable operational hours saved. Reported a 70%+ automation-rate baseline.
- Connected AI agents to CRMs, helpdesks and internal databases through REST APIs, webhooks and OAuth.
- Developed customer-facing support Q&A and sales-qualification agents using vector-database retrieval, explicit retries, JSON Schema validation and human escalation. Reported a 70% reduction in support workload for a mid-sized client and a 100% revenue increase in the first month after automated lead qualification with CRM synchronization replaced the manual step.
- Developed inbound and outbound voice AI agents for reminders, review collection and first-touch sales calls, with fully automated call flows handling up to 1,000 users per month per client.
- Worked as a senior agent developer on an agent platform, implementing deep GitHub, Figma, Telegram and Google Workspace integrations behind a unified agent interface, as well as Playwright and ADB automation. Made proof-of-concept and MVP architecture decisions with the technical lead.
- Architected pipelines using n8n and Make.com alongside custom asynchronous Python, including moving low-code systems onto services.
- For regulated clients and clients whose policies prohibited public LLM APIs, deployed fine-tuned models locally on client-owned servers against their own databases, with zero external data egress and full client data control.
- Ran workshops and training so in-house teams could own and extend solutions after handover.

Python, async programming, PostgreSQL, REST APIs, webhooks, OAuth, vector databases, JSON Schema validation, GitHub, Figma, Telegram, Google Workspace, Playwright, ADB, n8n, Make.com, fine-tuned models. Additional role-listed technologies: Generative AI, AI orchestration, Cloud Security, Microsoft Defender for Cloud, Microsoft Azure, Azure OpenAI Service, Amazon Bedrock, Data Governance, Anthropic Claude, LangGraph, LangChain, MCP, Google Cloud Platform, AWS, Qdrant, prompt engineering, FastAPI, Docker.

AI Solutions Architect - OpenClaw

Built and continues to operate OpenClaw, an eight-agent self-hosted LLM orchestration system on a bare-metal Raspberry Pi 5. A central project-management router drives an explicit state machine, validates worker outputs and escalates to a human. Approximately 15 production tasks run continuously at zero cloud cost.

OpenClaw, LLM orchestration, multi-agent systems, Raspberry Pi 5, state machine.

Afina, Team Lead

Sep 2024 - Nov 2025

Led support operations and backend infrastructure at a Web3 infrastructure company, acting as the link between users and engineering.

- Reduced median ticket response time from 15 minutes to 2 minutes at up to 200 tickets per day across a seven-person team by architecting data synchronization between distributed backend systems and the support interface.
- Operated distributed backend infrastructure across cloud environments and automated provisioning, dependency management and deployment using Docker and custom Python.
- Built knowledge bases and automated support workflows across Telegram and Discord.
- Turned vague user reports into reproducible engineering tickets and wrote the guides, documentation and tutorials used by the team.

Docker, Python, distributed backend systems, Telegram, Discord. Additional role-listed technologies: RAG, Anthropic Claude, LangGraph, LangChain, AI agents, MCP, AWS, multi-agent orchestration, Linux.

Blockchain & Automation - Infrastructure Engineer

Jun 2021 - Oct 2025

- Operated production validator and RPC node infrastructure for multiple L1/L2 networks, including Ethereum and Cosmos, with monitoring, alerting, security hardening, fault tolerance and automatic recovery.
- Operated up to 30 rented Linux servers simultaneously, including VPS and dedicated servers across major hosting providers. Automated monitoring and log parsing using Bash and Python.
- Managed Docker containers and network security, including SSH hardening and firewalls.
- Built Python browser-automation tooling for high-volume Web3 workflows.

Linux, Docker, Python, Bash, Ethereum, Cosmos blockchain, validator nodes, RPC nodes, SSH, firewalls. Additional role-listed technologies: Cloud Security, Anthropic Claude, LangGraph, LangChain, Google Cloud Platform, AWS, multi-agent orchestration, distributed systems, REST APIs.

Sales-enablement RAG platform

Company-wide answer engine using hybrid vector and keyword retrieval with reciprocal rank fusion on PostgreSQL and pgvector. Includes inline source citations and a numeric verifier that refuses unsupported figures and logs a knowledge gap.

An LLM gateway provides a provider fallback chain, golden-set evaluation with an 80% CI evaluation gate, and Langfuse tracing.

A curator console includes TOTP two-factor authentication, RBAC and an audit log. Docker Compose keeps development and production on the same stack.

PostgreSQL, pgvector, hybrid retrieval, reciprocal rank fusion, LLM gateway, golden-set evaluation, CI, Langfuse, TOTP 2FA, RBAC, audit logging, Docker Compose.

Multi-provider cloud-spend audit agent

PRODUCTION SINCE MAY 2026

Production agent that collects weekly spending data from eight cloud and LLM-provider APIs into a finance-owned spreadsheet.

Runs as a serverless job protected by an OIDC check on the scheduler's identity, with secrets from Secret Manager, per-adapter failure isolation and idempotent upserts. Infrastructure is provisioned using Terraform, with structured JSON logging and a log-based alert.

Cloud and LLM-provider APIs, serverless computing, OIDC, Secret Manager, idempotent upserts, Terraform, structured JSON logging, log-based alerting.

EDUCATION

- Kyiv National Economics University
  BSc Economics & Finance. Basis for ROI and cost modelling of solution options.
  2015 - 2019

CERTIFICATION

- **Microsoft Applied Skills: Integrate MCP tools with agents in Microsoft Foundry (Azure OpenAI)**
  2026
- **Microsoft Applied Skills: Secure AI solutions in the cloud**
  2026
- **AWS Agentic AI Demonstrated**
  2026
- **Anthropic: MCP Advanced Topics**
  2026