Vladyslav Dykyi

AI Engineer | RAG · LLMOps · Agentic Systems · MLOps

Linz, Austria | vladdikiy17@gmail.com | +43 688 64059508
linkedin.com/in/vladyslav-dykyi-4321a9283 | github.com/dykyivladk1

SUMMARY

AI Engineer focused on production-oriented GenAI systems, RAG pipelines, agentic workflows and LLMOps. Experienced in building and operating Python-based AI applications, stateful LangGraph workflows, evaluation pipelines, model deployment, CI/CD, observability and cloud infrastructure. Hands-on experience with MLflow for experiment tracking and model lifecycle workflows, as well as practical self-directed experience with Databricks for data, ML and GenAI workflows. Comfortable owning AI systems end to end, from retrieval and evaluation to deployment, monitoring and reliability.

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript, JavaScript

GenAI & LLMOps: RAG, LangChain, LangGraph, agentic workflows, tool/function calling, prompt engineering, evaluation pipelines, LLM-as-a-judge concepts, evaluation datasets, guardrails, human-in-the-loop workflows, structured outputs, model/prompt testing

ML & MLOps: MLflow, model experiment tracking, model deployment, inference pipelines, LoRA fine-tuning, ONNX Runtime, TensorFlow Lite, model quantization, automated testing

Data & Retrieval: PostgreSQL, pgvector, Qdrant, FAISS, MongoDB, Elasticsearch, vector search, data pipelines

Cloud & DevOps: Azure, Databricks (hands-on/self-directed), Docker, Kubernetes, Terraform, GitHub Actions, CI/CD, Linux

Backend: FastAPI, Flask, NestJS, REST APIs, SSE, API integrations

Observability & Reliability: Logging, tracing, Kibana, Grafana, monitoring, root-cause analysis, retries, checkpointing, failure recovery

PROFESSIONAL EXPERIENCE

Tractive
Feb 2026 – Present

Production Engineer | Linz, Austria | Full-time

- Build and maintain Python automation tools, diagnostic services and APIs supporting production systems and manufacturing data workflows.
- Analyze production data using MongoDB, Elasticsearch and Kibana to detect anomalies, investigate failures and support engineering decisions.
- Lead incident investigation and root-cause analysis for production and factory test issues.
- Improve reliability and maintainability of diagnostic workflows used across global manufacturing operations.
- Collaborate with engineering, production and external manufacturing partners to resolve complex technical issues.

LORAtech e.U. – LORA Cube
Sep 2024 – Sep 2025

AI Developer | Full-time

- First AI engineer on the team; developed the initial production-oriented MVP of LORA Cube, combining edge AI models with cloud-hosted LLM services.
- Used MLflow for model experiment tracking, comparison and development workflows during ML model training and optimization.
- Deployed ML models for server-side inference and optimized edge models using ONNX Runtime, TensorFlow Lite and INT8/INT4 quantization.
- Fine-tuned LLMs using LoRA and evaluated model behaviour across different configurations.
- Built backend and data services with NestJS, PostgreSQL and REST APIs.
- Deployed containerized workloads with Docker and Kubernetes on Azure using Terraform and GitHub Actions CI/CD.
- Implemented logging, tracing, guardrails and PII detection using spaCy and rule-based validation.
- Worked across model development, backend engineering, cloud infrastructure and production reliability.

Johannes Kepler University Linz
Mar 2024 – Dec 2025

AI Engineer | Linz, Austria | Part-time

- Built a multi-agent AI platform providing students with specialized AI assistants for domains including biology, physics and mathematics.
- Designed RAG pipelines with dedicated knowledge bases, Qdrant vector retrieval and PostgreSQL-based conversation memory.
- Developed stateful agent workflows using LangChain and LangGraph, including checkpointing, retries and failure recovery.
- Integrated multiple cloud and local LLM providers through a common orchestration layer.
- Built tracing and evaluation pipelines for testing retrieval and agent behaviour.
- Developed structured agent routing and tool-use workflows designed for predictable and recoverable execution.
- Deployed and operated the platform on Azure.

Freelance 2022 – Mar 2024

Backend Developer & AI Engineer | Remote

- Built custom RAG pipelines, AI agents, backend APIs and automation solutions for different clients.
- Developed Python backend systems using FastAPI and Flask.
- Built SOMA, a Telegram automation platform for a Ukrainian clothing business with 80k+ social media followers.
- Built WIZA, an AI-powered travel assistant covering 39 countries and 462 cities using external APIs and automated data retrieval.
- Delivered web scraping, data-processing and API-integration solutions.

SELECTED PROJECTS

Adify AI

Python · Flask · Qdrant · OpenAI · Docker

- Built a RAG application that converts natural-language mood requests into Spotify playlist recommendations.
- Implemented vector retrieval, MMR diversification and constrained generation to prevent hallucinated tracks.
- Added deterministic fallback logic when LLM calls fail.
- Added automated tests with pytest and containerized the application using Docker Compose.

Universe

Python · Flask · LangChain · LangGraph · SSE

- Built a platform for creating and running custom AI agents using OpenAI and Anthropic models.
- Implemented multi-agent orchestration with a LangGraph supervisor and structured routing outputs.
- Added tool selection, agent templates and persistent chat history.
- Streamed token generation, tool calls and agent hand-offs to the browser using SSE.

Databricks & GenAI Experimentation

Databricks · MLflow · Python · Vector Search

- Hands-on self-directed experimentation with the Databricks platform for ML and GenAI workflows.
- Explored notebooks, data processing, MLflow integration, model lifecycle concepts and vector-search-based GenAI architectures.
- Studied deployment patterns for RAG applications, model serving and LLMOps workflows on Databricks.
- Comfortable transferring existing production experience with Python, MLflow, Azure, RAG and MLOps into Databricks-based environments.

EDUCATION

BSc Artificial Intelligence
Johannes Kepler University Linz

BSc Economics
Odesa National Economic University

CERTIFICATIONS

Python Programming Certificate
Odesa Law Academy

Responsible AI
Coursera | Issued Jun 2024 · Expires Jun 2027

Generative Ai
Coursera | Issued Jun 2024 · Expires Jun 2027

Build a Machine Learning Model with Python Skill Path
Codecademy | Issued Sep 2023

LANGUAGES

Ukrainian — Native | Russian — Fluent | English — C1 | German — A2