

Patryk Kozak

Senior AI Platform Engineer · Data Scientist · AI & Data Consultant

 patrykpiotrkozak@gmail.com  +436602656336  Vienna, Austria  [LinkedIn Profile](#)

 [Github](#)

PROFILE

AI Engineer and MLOps Consultant with 4+ years of experience delivering enterprise AI, data platform, and cloud-native solutions across the DACH region. Experienced in Kubernetes-based AI infrastructure, observability, Generative AI deployments, and production-grade ML systems for automotive, public sector, and enterprise environments. Strong background in bridging technical implementation with business-oriented consulting and stakeholder communication.

SKILLS

GENERATIVE AI & INTELLIGENT AUTOMATION: LLMs · OpenAI API · Anthropic Claude API · Retrieval-Augmented Generation (RAG) · AI Agents & Orchestration · Multimodal AI · Vector Databases (Milvus, ChromaDB) · Semantic Search · LLM Observability · Prompt Engineering · AI Workflow Automation · Intelligent Document Processing (IDP)

AI INFRASTRUCTURE & MLOPS: Kubernetes · Docker · Prometheus · Grafana · Loki · Alertmanager · Promtail · Grafana Alloy · GPU & System Monitoring · NVIDIA CUDA · Apache Airflow · Apache Flink · Kubeflow · HPE Private Cloud AI · HPE Ezmeral

DATA SCIENCE & MACHINE LEARNING: Python · TensorFlow · PyTorch · NVIDIA RAPIDS (cuDF, cuML) · Time-Series Forecasting · Neo4j Graph Data Science · Graph Similarity Algorithms · Exploratory Data Analysis (EDA) Pandas, Numpy, Matplotlib

DATA ENGINEERING & PLATFORM: REST API, Apache Kafka · PostgreSQL · Neo4j · Cypher Query Language · Linux · Git · CI/CD

AGENTIC AI & ORCHESTRATION: LangChain · LangGraph · Model Context Protocol (MCP) · OpenAI Agents SDK · LlamaIndex · Anthropic Claude API

EMPLOYMENT HISTORY

Data Scientist · Technology Consultant, Hewlett Packard Enterprise

Vienna, Austria (Hybrid)

02/2023 – 06/2026

Advisory and Professional Services — DACH Region Delivery

Delivered enterprise consulting engagements across automotive, public sector, construction, industrial distribution, and sports organizations, specializing in AI infrastructure, MLOps, cloud platforms, data engineering, workflow automation, API integrations, and enterprise application modernization. Provided strategic and technical advisory across Enterprise AI, Edge-to-Cloud architectures, data platforms, and workflow integration solutions, enabling organizations to modernize digital capabilities and accelerate AI adoption.

Full-Stack AI Infrastructure & MLOps — Digital AI Avatar, State-Owned Real Estate Enterprise

- Designed and deployed an end-to-end observability stack on a multi-tenant Kubernetes cluster, aggregating metrics and logs across namespaces to monitor GPU utilization, LLM inference performance (latency, token throughput), RAG pipelines, vector stores, and distributed AI system components.
- Implemented monitoring for NVIDIA NIM microservices, vLLM-served models, and Milvus vector databases; identified performance bottlenecks, capacity constraints, and infrastructure issues across the AI inference stack.
- Built production dashboards and alerting workflows using Prometheus, Grafana, Grafana Loki, and Alertmanager to support reliability, performance monitoring, and SLO tracking for customer-facing Generative AI workloads.

- Designed and supported Agentic AI workflows integrating LLMs, retrieval systems, vector databases, and external tools to enable autonomous task execution, information retrieval, and AI-driven decision support.
- Worked with AI agent deployment patterns, including agent orchestration, tool integration, prompt workflows, RAG-based knowledge grounding, and scalable deployment of LLM-powered applications on enterprise AI infrastructure.
- Supported the operationalization of Generative AI and Agentic AI applications by connecting AI services, APIs, data sources, and infrastructure components into reliable production workflows.

MLOps & Time-Series Forecasting — Internal HPE Product Initiative

- Contributed to a forecasting microservice covering the complete MLOps lifecycle, including data ingestion, model training, deployment, and both streaming and batch processing.
- Exposed forecasting capabilities through REST APIs for integration with HPE Private Cloud AI, LLM-based applications, and enterprise platforms.
- Designed API-driven integration patterns between AI services and enterprise applications, enabling automated workflows, data exchange, and scalable service integration.
- Supported ML lifecycle orchestration and time-series architecture design, including stream-processing concepts using Apache Flink.

Data Science & Graph Analytics — Global Automotive Manufacturer

- Co-designed a Python-based similarity search algorithm with custom evaluation metrics (Jaccard similarity for strings, Pearson and Spearman correlation for numerical features) to improve enterprise metadata retrieval quality.
- Applied Neo4j Graph Data Science algorithms, including similarity analysis, centrality measures, and path-finding, using Cypher queries to analyze highly connected enterprise data.
- Developed a Python data transformation engine integrating data through Apache Kafka into structured analytics pipelines; implemented supervised classification workflows on metadata.
- Managed PostgreSQL and Neo4j databases and containerized services using Docker for production deployment.
- Performed exploratory data analysis (EDA), validated analytical hypotheses, and evaluated machine learning pipeline results.

Data Engineering — Professional Sports Organization

- Developed and scheduled Apache Airflow DAGs to orchestrate data pipelines supporting a lottery-based fan engagement platform, ensuring reliable and auditable data processing.

Platform Engineering — Global Premium Automotive Manufacturer

- Designed and validated zero-downtime certificate rotation strategies for HPE Ezmeral Runtime Environment, enabling seamless certificate updates without service interruption.
- Executed post-installation configuration and stabilization activities to support reliable production deployments.

Python Automation & Infrastructure Support — International Trade & Distribution Company

- Developed Python automation workflows for file retention, log rotation, and server-side data lifecycle management.
- Automated operational processes using scheduled jobs and scripting while supporting server upgrades, certificate renewals, and runtime migrations.

Data-Centric Architecture & Strategy Consulting — National Construction Company

- Facilitated and documented stakeholder workshops using the HPE Edge-to-Cloud Adoption Framework (E2CAF), mapping data capabilities, access policies, and RBAC models.
- Supported the creation of a Transformation Business Office to improve governance, coordination, and decision-making across the edge-to-cloud transformation program.

Training, Community & Thought Leadership

- Participated in an HPE and NVIDIA collaborative development program focused on AI infrastructure operations, Generative AI, large language models, and multimodal AI.
- Built hands-on experience with the NVIDIA AI ecosystem and pursued NVIDIA certifications.
- Completed technical training for HPE Private Cloud AI (PCAI), covering AI infrastructure, deployment patterns, and enterprise AI workloads.
- Completed Microsoft Azure ML training aligned with the Azure Data Scientist Associate track, including ML lifecycle management, AutoML, and MLOps deployment patterns.
- Contributed to an LLM Hackathon by designing and developing an autonomous AI agent using LangChain.

- Participated in HPE DeliverAI sessions to share AI delivery practices and contribute to internal AI knowledge exchange.
- Delivered technical presentations on AI-based aerial image analysis (anomaly detection and object recognition) and HPE's Generative AI portfolio, including GreenLake and Aleph Alpha partnership solutions, for enterprise customers.

Google Ads Global Care Associate, *Google (via Accenture)*

Warsaw, Poland (Remote, Dublin team)

08/2021 – 10/2022

Managed Google Ads campaigns for DACH-market advertisers, proactively using Google Ads' internal tooling to diagnose and resolve campaign issues. Delivered strategic and analytical recommendations across diverse advertiser portfolios to optimize campaign performance against client business goals.

Data Analyst, *Accenture*

Warsaw, Poland

05/2020 – 07/2021

Project Delivery Operations for a global social-media platform, DACH market. Analyzed social-media ads campaign data, identifying sales trends through correlation analysis and visualization. Processed and organized large-scale social-media datasets in compliance with platform-specific guidelines. Reviewed incoming data for completeness and accuracy; flagged and corrected anomalies to prevent duplication and quality degradation.

B2B Sales Specialist, *ColligoGmbH*

Vienna, Austria

09/2018 – 07/2019

Supported B2B sales and digital marketing services for clients across the DACH region. Conducted outreach, lead qualification, and client engagement across German-speaking markets.

EDUCATION

Software Development Academy , *Certified Data Scientist — Professional Diploma*

Warsaw, Poland

09/2020 – 07/2022

Two-year data science program covering Python, statistics, machine learning, SQL, data visualization, and applied project work.

365 Data Science , *Certified Data Scientist — Professional Career Diploma*

Certificate ID: DD-D9D0C2C28B Data Scientist Career Track covering probability, statistics, Python, SQL, machine learning, and deep learning with applied project assessments.

Dr. Roland AHS, *High School Diploma (Matura)*

Vienna, Austria

06/2016

LANGUAGES

German	English	Polish
Professional working proficiency	Professional working proficiency	Native

PROFESSIONAL CERTIFICATIONS & SUPPLEMENTARY EDUCATION

Machine Learning Specialization, *DeepLearning.AI*

Deep Learning Specialization, *DeepLearning.AI*

Microsoft Azure Data Science Track, *Microsoft*

Docker Kubernetes, *TheLinuxFoundation*
Self-paced

Python Developer, *ZTMAcademy*

HPE NVIDIA AI Development Initiative, *HPE & NVIDIA*

NVIDIA-Certified Associate: AI Infrastructure & Operations (NCA-AIIO), *NVIDIA*

NVIDIA-Certified Associate: Generative AI with LLMs (NCA-GENL), *NVIDIA*

NVIDIA-Certified Associate: Generative AI Multimodal (NCA-GENM), *NVIDIA*

NVIDIA DLI — AI Infrastructure & Operations Fundamentals, *NVIDIA Deep Learning Institute*

NVIDIA DLI — Introduction to Transformer-Based Natural Language Processing,
NVIDIA Deep Learning Institute

LFS207: Linux System Administration Essentials, *TheLinuxFoundation*

NVIDIA DLI — Building AI Agents with Multimodal Models, *NVIDIA Deep Learning Institute*

NVIDIA DLI — Building Conversational AI Applications, *NVIDIA Deep Learning Institute*

NVIDIA DLI — Accelerating End-to-End Data Science Workflows, *NVIDIA Deep Learning Institute*

Neo4j Graph Data Science, *Neo4j*

Neo4j Cypher Query Language, *Neo4j*

Software Engineering for Data Scientists in Python, *DataCamp*

Time Series Modeling, *Skillsoft*

Agile Project Management, *Accenture*

Google Ads Associate, *Google Career Certificates*