

Patryk Kozak

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

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

 [Github](#)

PROFIL

KI-Ingenieur und MLOps-Berater mit über 4 Jahren Erfahrung in der Umsetzung von Enterprise-KI-, Datenplattform- und Cloud-Native-Lösungen in der DACH-Region. Erfahrung in Kubernetes-basierter KI-Infrastruktur, Observability, Bereitstellung von Generativer KI sowie produktionsreifen ML-Systemen für Automobilindustrie, öffentlichen Sektor und Unternehmensumgebungen. Ausgeprägte Fähigkeit, technische Implementierung mit geschäftsorientierter Beratung und Kommunikation mit Stakeholdern zu verbinden.

TECHNISCHE KOMPETENZEN

GENERATIVE AI & INTELLIGENTE AUTOMATISIERUNG: LLMs · OpenAI API · Anthropic Claude API · Retrieval-Augmented Generation (RAG) · AI Agents & Orchestrierung · Multimodale AI · Vector Databases (Milvus, ChromaDB) · Semantische Suche · LLM Observability · Prompt Engineering · AI Workflow-Automatisierung · Intelligent Document Processing (IDP)

AI-INFRASTRUKTUR & 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: HPE Ezmeral REST API, Data Pipelines, Apache Kafka · PostgreSQL · Neo4j · Cypher Query Language · Linux · Git · CI/CD

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

BERUFLICHER WERDEGANG

Data Scientist · Technology Consultant, Hewlett Packard Enterprise

Vienna, Austria (Hybrid)

02/2023 – 06/2026

Advisory & Professional Services — DACH Delivery

Durchführung von Enterprise-Consulting-Projekten in den Bereichen Automobilindustrie, öffentlicher Sektor, Bauwesen, industrielle Distribution und Sportorganisationen mit Schwerpunkt auf KI-Infrastruktur, MLOps, Cloud-Plattformen, Data Engineering, Workflow-Automatisierung, API-Integrationen und Modernisierung von Enterprise-Anwendungen. Strategische und technische Beratung in den Bereichen Enterprise AI, Edge-to-Cloud-Architekturen, Datenplattformen und Workflow-Integrationslösungen zur Modernisierung digitaler Fähigkeiten und Beschleunigung der KI-Adoption.

Full-Stack KI-Infrastruktur & MLOps — Digitaler KI-Avatar, staatliches Immobilienunternehmen

- Entwicklung und Bereitstellung eines End-to-End-Observability-Stacks auf einem Multi-Tenant-Kubernetes-Cluster zur Aggregation von Metriken und Logs über Namespaces hinweg sowie zur Überwachung von GPU-Auslastung, LLM-Inference-Performance (Latenz, Token-Durchsatz), RAG-Pipelines, Vector Stores und verteilten AI-Systemkomponenten.
- Implementierung von Monitoring für NVIDIA NIM Microservices, vLLM-bereitgestellte Modelle und Milvus Vector Databases; Identifikation von Performance-Bottlenecks, Kapazitätsengpässen und Infrastrukturproblemen im gesamten AI-Inference-Stack.

- Aufbau produktiver Dashboards und Alerting-Workflows mit Prometheus, Grafana, Grafana Loki und Alertmanager zur Unterstützung von Reliability, Performance Monitoring und SLO-Tracking für kundenorientierte Generative-AI-Workloads.
- Entwicklung und Unterstützung von Agentic-AI-Workflows durch Integration von LLMs, Retrieval-Systemen, Vector Databases und externen Tools zur Ermöglichung autonomer Aufgabenbearbeitung, Informationsbeschaffung und AI-gestützter Entscheidungsunterstützung.
- Arbeit mit AI-Agent-Deployment-Patterns einschließlich Agent-Orchestrierung, Tool-Integration, Prompt-Workflows, RAG-basiertem Knowledge Grounding sowie skalierbarer Bereitstellung von LLM-basierten Anwendungen auf Enterprise-AI-Infrastrukturen.
- Unterstützung bei der Operationalisierung von Generative-AI- und Agentic-AI-Anwendungen durch Integration von AI-Services, APIs, Datenquellen und Infrastrukturkomponenten in zuverlässige Produktions-Workflows.

MLOps & Time-Series Forecasting — Internes HPE-Produktprojekt

- Mitwirkung an einem Forecasting-Microservice über den gesamten MLOps-Lifecycle hinweg, einschließlich Data Ingestion, Model Training, Deployment sowie Streaming- und Batch-Verarbeitung.
- Bereitstellung von Forecasting-Funktionalitäten über REST APIs zur Integration mit HPE Private Cloud AI, LLM-basierten Anwendungen und Enterprise-Plattformen.
- Entwicklung API-basierter Integrationsarchitekturen zwischen AI-Services und Enterprise-Anwendungen zur Ermöglichung automatisierter Workflows, Datenaustausch und skalierbarer Service-Integration.
- Unterstützung bei ML-Lifecycle-Orchestrierung und Time-Series-Architekturdesign einschließlich Stream-Processing-Konzepten mit Apache Flink.

Data Science & Graph Analytics — Globaler Automobilhersteller

- Co-Design und Entwicklung eines Python-basierten Similarity-Search-Algorithmus mit kundenspezifischen Evaluierungsmetriken (Jaccard Similarity für Strings sowie Pearson- und Spearman-Korrelation für numerische Features) zur Verbesserung der Enterprise-Metadaten-Retrieval-Qualität.
- Anwendung von Neo4j Graph Data Science Algorithmen einschließlich Similarity Analysis, Centrality Measures und Path-Finding mittels Cypher Queries zur Analyse hochvernetzter Enterprise-Daten.
- Entwicklung einer Python Data Transformation Engine zur Integration von Daten über Apache Kafka in strukturierte Analytics-Pipelines; Implementierung von Supervised-Classification-Workflows auf Metadaten.
- Verwaltung von PostgreSQL- und Neo4j-Datenbanken sowie Containerisierung von Services mittels Docker für produktive Deployments.
- Durchführung von Exploratory Data Analysis (EDA), Validierung analytischer Hypothesen und Evaluation von Machine-Learning-Pipeline-Ergebnissen.

Data Engineering für Professionelle Sportorganisation

- Entwicklung und Scheduling von Apache Airflow DAGs zur Orchestrierung von Datenpipelines für eine Lottery-basierte Fan-Engagement-Plattform sowie Sicherstellung zuverlässiger und auditierbarer Datenverarbeitung.

Platform Engineering für Globaler Premium-Automobilhersteller

- Entwicklung und Validierung von Zero-Downtime-Zertifikatsrotationsstrategien für HPE Ezmeral Runtime Environment zur Durchführung nahtloser Zertifikatsaktualisierungen ohne Service-Unterbrechungen.
- Durchführung von Post-Installation-Konfigurationen und Stabilisierungstätigkeiten zur Unterstützung zuverlässiger Produktionsbereitstellungen.

Python Automation & Infrastructure Support — Internationales Handels- und Distributionsunternehmen

- Entwicklung von Python-Automatisierungs-Workflows für File Retention, Log Rotation und serverseitiges Data Lifecycle Management.
- Automatisierung operativer Prozesse mittels Scheduled Jobs und Scripting sowie Unterstützung bei Server-Updates, Zertifikatserneuerungen und Runtime-Migrationen.

Data-Centric Architecture & Strategy Consulting — Nationales Bauunternehmen

- Moderation und Dokumentation von Stakeholder-Workshops unter Verwendung des HPE Edge-to-Cloud Adoption Frameworks (E2CAF), einschließlich Mapping von Datenfähigkeiten, Zugriffskonzepten und RBAC-Modellen.
- Unterstützung beim Aufbau eines Transformation Business Office zur Verbesserung von Governance, Koordination und Entscheidungsfindung innerhalb des Edge-to-Cloud-Transformationsprogramms.

Training, Community & Thought Leadership

- Teilnahme an einem gemeinsamen Entwicklungsprogramm von HPE und NVIDIA mit Fokus auf AI-Infrastruktur-Betrieb, Generative AI, Large Language Models und multimodale AI.
- Aufbau praktischer Erfahrung mit dem NVIDIA AI Ecosystem und Vorbereitung auf NVIDIA-Zertifizierungen.

- Abschluss technischer Trainings für HPE Private Cloud AI (PCAI) mit Fokus auf AI-Infrastruktur, Deployment-Patterns und Enterprise-AI-Workloads.
- Abschluss eines Microsoft Azure ML Trainings entsprechend dem Azure Data Scientist Associate Track mit Fokus auf ML Lifecycle Management, AutoML und MLOps Deployment Patterns.
- Mitwirkung an einem LLM Hackathon durch Design und Entwicklung eines autonomen AI Agents mit LangChain.
- Teilnahme an HPE DeliverAI Sessions zur Weitergabe von AI-Delivery-Praktiken und Förderung des internen AI-Wissensaustauschs.
- Durchführung technischer Präsentationen zu AI-basierter Luftbildanalyse (Anomalieerkennung und Objekterkennung) sowie zum HPE Generative-AI-Portfolio einschließlich GreenLake- und Aleph-Alpha-Partnerschaftslösungen für Enterprise-Kunden.

Google Ads Global Care Associate, Google (via Accenture)

Warsaw, Poland (Remote, Dublin team)

08/2021 – 10/2022

- Betreuung von Google Ads-Kampagnen für Werbetreibende im DACH-Markt sowie proaktive Nutzung interner Google-Ads-Tools zur Diagnose und Behebung von Kampagnenproblemen.
- Bereitstellung strategischer und analytischer Empfehlungen für unterschiedliche Werbetreibenden-Portfolios zur Optimierung der Kampagnenleistung im Einklang mit den Geschäfts- und Marketingzielen der Kunden.

Data Analyst, Accenture

Warsaw, Poland

05/2020 – 07/2021

- Durchführung von Project Delivery Operations für eine globale Social-Media-Plattform im DACH-Markt.
- Analyse von Social-Media-Ads-Kampagnendaten zur Identifikation von Verkaufstrends mittels Korrelationsanalysen und Datenvisualisierung.
- Verarbeitung und Strukturierung großer Social-Media-Datensätze unter Einhaltung plattformspezifischer Richtlinien und Qualitätsanforderungen.
- Prüfung eingehender Daten auf Vollständigkeit und Genauigkeit sowie Identifikation und Korrektur von Anomalien zur Vermeidung von Duplikaten und Qualitätsverlusten.

B2B Sales Specialist, Colligo GmbH

Vienna, Austria

09/2018 – 07/2019

- Unterstützung von B2B-Vertriebs- und Digital-Marketing-Services für Kunden in der DACH-Region.
- Durchführung von Kundenansprache, Lead-Qualifizierung und Kundeninteraktion in deutschsprachigen Märkten.

AUSBILDUNG

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

Warsaw, Poland

09/2020 – 07/2022

Zweijähriges Data-Science-Programm mit Schwerpunkt auf Python, Statistik, Machine Learning, SQL, Datenvisualisierung und praxisorientierten Projektarbeiten.

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

Certificate ID: DD-D9D0C2C28B — Data Scientist Career Track mit Schwerpunkt auf

Wahrscheinlichkeitsrechnung, Statistik, Python, SQL, Machine Learning und Deep Learning sowie praxisorientierten Projektbewertungen.

Dr. Roland AHS, Matura

Vienna, Austria

06/2016

SPRACHEN

German

Zweisprachig

English

Verhandlungssicher

Polish

Muttersprache

BERUFLICHE ZERTIFIZIERUNGEN & ZUSÄTZLICHE WEITERBILDUNGEN

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*