

Alexey Berezka

BERATER KI & DIGITALE
TRANSFORMATION · KI / LLM
ENGINEER

KONTAKT

@ almberezka@gmail.com
☎ +49 176 4295 0906
✚ Gernersheim · bundesweit
umzugsbereit
in linkedin.com/in/alexberezka
🌐 berezka.me

VERFÜGBARKEIT

Ab Oktober 2026 · Vollzeit · deutschlandweit oder remote

EU-Arbeitserlaubnis · kein Sponsoring erforderlich

SPRACHEN

Russisch	Muttersprache
Englisch	fließend
Tschechisch	fließend
Deutsch	B2

KERNKOMPETENZEN

- KI & Generative KI
- LLMs (Claude, GPT-4o, Gemini) · RAG · Agentic / Multi-Agent-Systeme · LLMOps · Prompt Engineering · MCP · Embeddings / Semantische Suche · Computer Vision
- KI-Strategie & Transformation
- KI-Adoption · Use-Case-Identifikation · BPMN · ROI / Business Case · Stakeholder-Management · MCDM / AHP
- Programmierung
- Python · TypeScript · C/C++ · SQL · Bash
- Infrastruktur & MLOps
- Docker · Linux · nginx · CI/CD (GitHub Actions) · FastAPI · Git · n8n · Power BI
- Governance & Branchen
- EU AI Act · DSGVO · ISO/IEC 42001 · ISO 9001 / IATF 16949 · TISAX · Automotive · Fertigungs-KMU

PROFIL

Ich **baue und betreibe produktive KI** — und mache sie fürs Geschäft nutzbar. Über 5 Jahre Engineering in Automotive, Luftfahrt und Analytics; heute betreibe ich täglich **LLM-Systeme**: RAG-Pipelines, einen **Orchestrator mit 200+ Agenten** und LLMOps mit einem Eval-Gate, das die **Modellkosten um ~95 % senkte**. Aktuell verantworte ich das Programm **KI & Digitale Transformation** bei **Kyocera Automotive**; M.Sc. am KIT (Masterarbeit: Entscheidungsmodelle für KI-Einführung in der Fertigung). Zuhause im Vorstand wie in der Codebase — ROI, Stakeholder, **EU AI Act / DSGVO** — auf DE / EN / CZ / RU.

BERUFSERFAHRUNG

KI und Digitale Transformation (Werkstudent)

Mai 2026 – Sep 2026

- Kyocera Automotive and Industrial Solutions GmbH · Dietzenbach · Studentische Supportkraft (befristet)
- Greenfield-Programm für **KI & Digitalisierung** im europäischen B2B-Vertrieb — zweisprachige (DE/EN) Koordination zwischen Fachbereichen, IT und Muttergesellschaft.
 - Prozessaufnahme**: Stakeholder-Interviews und Workshops, **BPMN**-Modelle für Qualität, Customer Service und Vertrieb; **Use-Case-Identifikation**.
 - Toolbewertung & Pilot**: **RAG**, **Copilot Studio**, **Azure OpenAI** (EU Data Zone), **Claude API** — Lieferantenbewertung und ehrlicher ROI-Case unter **EU AI Act / DSGVO**.

Embedded Software Developer

Dez 2019 – Mär 2023

- ZF Engineering · Pilsen, Tschechische Republik · Automotive-Steuergeräteentwicklung
- Automatisierte HiL/SiL-Validierungspipelines für sicherheitskritische ECUs in vier OEM-Programmen — **25 % kürzere Validierungszyklen**.
 - Release-Reviews unter ASIL-D; quantitative Go/No-Go-Kriterien — die Logik, die meine KIT-Masterarbeit heute formalisiert.
 - Strukturierte 8D-Root-Cause-Analyse; Fehlerwiederholrate sank Quartal für Quartal.

Ingenieur — Luftfahrtmesstechnik

Jul 2017 – Jun 2018

- Forschungs- und Produktionsverband „Typhoon“ · Russland
- Architekturentwurf für Messsysteme zur Wettererfassung in der Luftfahrt unter strengen regulatorischen Vorgaben; Übersetzung von Zertifizierungsanforderungen in technische Spezifikationen — dieselbe Machbarkeits-Disziplin, die ich heute auf KI-Use-Cases anwende.

Data Analytics Engineer

Mai 2016 – Jun 2017

- Stream · Moskau, Russland · Marketing-Analytics
- Aufbau von Kundensegmentierungsmodellen (SAS + Python) und ETL-Pipelines — **40 % weniger manuelles Reporting**; Wochenreports in Self-Service-Dashboards überführt.

AUSGEWÄHLTE KI-SYSTEME — IN PRODUKTION GEBAUT & BETRIEBEN

JARVIS — Multi-Agent-KI-Assistent. Self-hosted 24/7: ein Meta-Orchestrator koordiniert **200+ Spezialagenten** über eigene MCP-Server, mit Sprache (Whisper) und E-Mail-/Kalender-Tools. *Python · Claude API · MCP · Docker*

DeepSignal — autonome LLM-Pipeline. Verarbeitet Markt- und Nachrichtendaten und publiziert Analysen planmäßig, ohne manuelle Schritte; anbieterneutraler Modell-Adapter (Anthropic / OpenAI / DeepInfra / Ollama) plus Shadow-Eval-Gate — **Modellkosten ~95 % gesenkt.** *LLMOps · Eval-Gating · Backtesting*

Hybrides RAG-Wissenssystem. Fundierte Suche unter einer Sekunde über eine Basis von 4.500 Notizen — dichte Embeddings (bge-m3) + Sparse-FTS5 + Reranker, ~2.000 Dokumente im Index, nächtlicher Reindex. *RAG · Hybrid-Retrieval*

Eigene KI-Infrastruktur. Produktions-Stack auf dediziertem Server — Docker, nginx, systemd, CI/CD, FastAPI, eigene MCP-Server; anbieterportabel by design. *Docker · CI/CD · MCP*

AUSBILDUNG & FORSCHUNG

M.Sc. Information Systems Engineering & Management — Karlsruher Institut für Technologie (KIT) · HECTOR School · **Sep 2024 – Okt 2026**

Masterarbeit: Multi-Criteria Decision Model for Evaluating AI-Based Image Processing Feasibility in Manufacturing SMEs (Abgabe Sep 2026).

M.Sc. Elektronik & Angewandte Informatik — Westböhmische Universität, Pilsen · **Sep 2019 – Aug 2021**

Abschlussarbeit: Entwicklung elektronischer Systeme für eine Stratosphärenballon-Mission. Schwerpunkt: Embedded Systems, Hardware-Software-Co-Design, C/C++.

B.Sc. Funktechnik (Telekommunikation) — MIREA, Moskau · **Sep 2014 – Aug 2018**

Abschlussarbeit: Weltraumgestütztes mehrskaliges Phasen-Entfernungsmesssystem. Schwerpunkt: Telekommunikationssysteme und Signalverarbeitung.

Angewandte Forschung: *Adversarial Backtesting & Walk-Forward-Validierung von ML-Handelsstrategien; Multi-Strategie-Portfolio (Sharpe 3,78, Walk-Forward-validiert 2013–2025).*