Punit Kumar Mohanty
Nürnberg, Germany

punit@mohantypunit.com
mohantypunit.com
+49 176 58 04 52 74

PROFIL

Senior Solution Architect für KI- und Datenplattformen mit 12+ Jahren Erfahrung in KI-, Daten- und Plattformprojekten in Deutschland und Indien. Spezialisiert auf Architektur, Aufbau und Operationalisierung von KI-/Data-Plattformen, MLOps/DevOps und Automatisierung in regulierten Umfeldern (Steuern, Recht, Wirtschaftsprüfung, Industrie).
Umfangreiche Erfahrung in KI-Strategie, Use-Case-Definition, Stakeholder-Beratung, Workshops und Enablement von Fach- und IT-Teams. Stark in der Übersetzung von Business-Anforderungen in skalierbare, produktive AI-Lösungen. Seit 2026 selbstständig mit eigenem Gewerbe; Schwerpunkt auf KI-Systemen, die im regulierten, souveränen oder abgeschotteten Umfeld (On-Premises) tatsächlich in Betrieb gehen. Durchgehende Arbeitslinie an Wissensgraphen seit 2019.

BERUFSERFAHRUNG

Punit Mohanty IT- und KI-Beratung, Nürnberg — seit Apr 2026

Selbstständiger Solution Architect: KI- und Datenplattformen

- Entwicklung eines Wissensgraphen für das deutsche Recht mit hybrider Suche (Vektor, Graph, BM25) und deterministischer Prüfung von Fundstellen; Betrieb mit quelloffenen Modellen (Neo4j, Qdrant, Python, RAG)
- Auswahl und Bewertung von Sprach-, Einbettungs- und Reranking-Modellen auf öffentlichen deutschen Rechts-Benchmarks und eigenen Prüfsätzen; Auswertung strukturell gegen den Wissensgraphen statt durch ein weiteres Sprachmodell
- Full Paper bei JURIX 2026 angenommen (Toulouse, IOS Press)
- Schulungen zur KI-Kompetenz nach Art. 4 KI-VO für Kanzleien und Unternehmen

Rödl & Partner Global Digital Services GmbH, Nürnberg — Feb 2024 – Nov 2025

Sr. Application Manager / KI Solution Architect

- Aufbau und produktiver Betrieb einer internen Plattform auf quelloffenen und proprietären Modellen (Microsoft AI Foundry, Azure OpenAI), Auswahl je Anwendungsfall nach rechtlicher Zulässigkeit, inkl. Security, Governance und Performance
- Umsetzung von über 20 fachlichen KI-Assistenten für Steuer, Wirtschaftsprüfung, Recht und IT; Entwicklung der mandantenfähigen Authentifizierung
- Automatisierte Risikoanalyse mit OCR und strukturierter Extraktion aus Verträgen, Gesellschafterstrukturen und Jahresabschlüssen mit Quellenverlinkung (Audit Trail); manueller Analyseaufwand je Fall von 6–8 auf 1–1,5 Stunden reduziert, rund 75% weniger. Human-in-the-Loop als Entwurfsprinzip: Das Modell schlägt vor, deterministische Regeln prüfen, und die Risikomatrix bleibt für die Prüfung bearbeitbar.
- Konzeption und Umsetzung eines eigenen Retrieval-Verfahrens (RAG) sowie eines Evaluierungs-Frameworks für die architektonische Bewertung von KI-Tools, Plattformlösungen und Anbietern: Prüfmaßstab aus kuratierten Frage-Antwort-Sätzen, deterministischer Vergleich, Benchmark-Oberfläche; mehrtägige manuelle Vergleiche wurden zu Läufen von unter einer Stunde
- Konzept zur Anonymisierung personenbezogener Daten in KI-Prozessen
- Fachliche und personelle Führung: DevOps, KI-Fachkraft, TypeScript-Entwicklung, Application Management
- Lifecycle-Management von ~300 Anwendungen in den Bereichen Steuern, Recht und Wirtschaftsprüfung inkl. Abstimmung mit Fachbereichen und IT
- Betriebs- und Integrationsmodelle für KI-Anwendungen: Anbindung an Power Automate, Jira und Legacy-ERP-Anwendungen, Architektur der Datentransformation von SAP Concur in das interne ERP-System; Betriebsmodell mit Monitoring über Grafana, Prometheus und Loki
- Konzeption und Durchführung von KI-Workshops, Hackathons und Schulungen für Fach- und IT-Teams (Einsteiger bis Experten), inkl. Use-Case-Definition und Change-Begleitung

Intelizign Engineering Services GmbH, Nürnberg **Okt 2020 – Mai 2023**

Sr. Solution Architect (ML and Data) (Jan'22, Mai'23)

- Aufbau eines Wissensgraphen mit OWL aus vorhandenen XML-Sensordaten; Ergebnis dem eigenen und den weiteren Teams vermittelt
- Einführung und Vergleich von Kafka sowie Elasticsearch, OpenSearch, Loki, Prometheus und Grafana, jeweils mit Schulung des Teams
- Gestaltung und Implementierung von Best Practices für die Entwicklung mit Docker, Kubernetes, OpenSearch, Prometheus, Grafana u. a., um eine kontrollierte und reproduzierbare Entwicklungsumgebung zu standardisieren.
- Entwicklung von Datenpipelines für maschinelles Lernen zur Optimierung von Produktionslinien in enger Zusammenarbeit mit Stakeholdern.
- Sammlung, Transformation und Zusammenführung von Daten aus verschiedenen Quellen und Speicherung in einem Data Lake für maschinelles Lernen.
- Neuerstellung von Pipelines mit lokal einsetzbaren Open-Source-Lösungen zur Standardisierung der lokalen Entwicklung und Tests.
- Automatisierung der Bereitstellung durch Implementierung von Infrastructure-as-Code und CI/CD-Tools.

Machine Learning Engineer for IoT and Digital Manufacturing (Okt'20 – Dez'21)

- Durchführung umfangreicher explorativer Analysen zur Verarbeitung und Aggregation von Daten.
- Training und produktive Bereitstellung von Computer-Vision-Modellen: Bildverarbeitung für die Qualitätsprüfung von Leiterplatten sowie Optimierung von Produktionslinien.
- Entwicklung von Python-Webanwendungen zur Nutzung der trainierten Modelle und Bereitstellung der erforderlichen Ausgaben zur Reduzierung von Produktionszeit und Ressourcen.

Siemens AG, Nürnberg **Jun 2019 – Feb 2020**

Werkstudent / Masterarbeit

- Implementierung der Online-Terminplanung für flexible Fertigungssysteme unter Verwendung von Multiagenten-Deep-Reinforcement-Learning.
- Entwurf von Petri-Netz-Modellen zur Erstellung digitaler Klone von Fertigungssystemen für das Model Training.
- Entwicklung von Belohnungsfunktionen zur effektiven Definition der Ziele der Agenten.
- Implementierung mehrerer Algorithmen für Multi-Agenten Deep Reinforcement Learning.
- Orchestrierung des Trainingsprozesses unter Verwendung von Docker-Containern auf Kubernetes.
- Erstellung von Baselines, Evaluierungsszenarien und Metriken zur Beurteilung der Leistung gelernter Modelle.

DropsLab Technologies GmbH, Aachen **Feb 2019 – Mai 2019**

Freelance Data Scientist

- Datensammlung und Datenaufbereitung für verschiedene Projekte.
- Nutzung von Transfer Learning für die Objekterkennung mit der TensorFlow Object Detection API.
- Nahtlose Implementierung von Objekterkennung und -verfolgung in einer Android-App.

Department of Human Language Technologies and Pattern Recognition, RWTH, Aachen **Mrz 2016 – Jan 2019**

WiHi

- Erstellung von Pipelines für automatische Spracherkennung mithilfe von Bash-Skripten zur einfacheren Organisation und Automatisierung des gesamten Lebenszyklus der Spracherkennungsmethoden.
- Implementierung von Längenmodellen in Spracherkennungswerkzeugen (SPRINT/RASR) unter Verwendung von C++, um invertierte HMMs zu unterstützen.

Tata Consultancy Services, Bengaluru, Indien **Mai 2013 – Okt 2015**

Systems Engineer

- Bereitstellung von SAP-Administration, Linux-Server-Administration und Datenbankadministration, einschließlich kontinuierlicher Loganalyse und Leistungsüberwachung von SAP-Systemen.
- Analyse und Präsentation von Leistungsdaten der SAP-Systeme, die von Kunden gesammelt wurden.


punit@mohantypunit.com


+49 176 58 04 52 74


mohantypunit.com

TECHNISCHE FÄHIGKEITEN

Cloud & Plattformen	AWS, Azure, Microsoft AI Foundry, Docker, Kubernetes, Nginx, Keycloak (IAM), Azure Managed Identities und RBAC, PostgreSQL, Redis, MongoDB, SAP Basis, SAP HANA, SAP Business ByDesign; IT-Infrastruktur On-Premises und in der Cloud, abgeschottete (air-gapped) Umgebungen, souveräner Betrieb
DevOps & Observability	GitHub Actions, Jenkins, GitLab CI, Terraform, Grafana, Prometheus, Logstash, Git, GitOps (Flux), Apache Kafka, Elasticsearch/OpenSearch, Apache Airflow, Apache Spark / PySpark, Celery, Loki, OPC UA
Programmiersprachen	Python, C/C++, CUDA C, MATLAB, Bash, SQL, HTML/CSS, JavaScript, ReactJS, FastAPI, pydantic, ASGI, REST-APIs; früher: Java, .NET (2010–2016)
KI & Data	LLMs, RAG (Retrieval-Augmented Generation), GraphRAG, Fine-Tuning (LoRA/QLoRA), Dokumentenextraktion, OCR, Human-in-the-Loop, agentische KI (MCP, A2A, ACP), LangChain, LangGraph, KI-gestützte Entwicklung (Claude Code, GitHub Copilot), AutoGen, LightRAG, Vercel AI SDK, PyTorch, Hugging Face, Prompt Engineering, LLM-Evaluierung, LLMOps, Wissensgraphen / Knowledge Graphs, Neo4j, OWL/RDF, Vektordatenbanken (Qdrant, PGVector), MLOps, Machine Learning, Computer Vision, Reinforcement Learning, Data Engineering, ETL, Data Lakes, Multi-Tenancy, Datenanonymisierung, DSGVO, EU-KI-Verordnung

WORKSHOPS UND VERÖFFENTLICHUNGEN

Workshops

- Ganztägiger Workshop „Democratizing AI: Empowering the Next Generation“ **Feb 2026**
  United School of Business Management (MCA-Abteilung), Bhubaneswar, Indien
  Gemeinsam mit Avyaktitech; zweiter Referent: Project Lead, TCS

- Gastvortrag im Awareness-Workshop „Responsible and Intelligent Use of Artificial Intelligence“ **Feb 2026**
  Odisha University of Technology and Research, School of Computer Sciences,
  Bhubaneswar, Indien

Veröffentlichung

„Multi-Agent Deep Q-Network Approach for Online Job Shop Scheduling in Flexible Manufacturing“ · ICMSMM 2020, Tokio, Japan

Angenommen

„Per-Substructure Attribution of Knowledge-Graph Value for German Statutory Retrieval“ · Full Paper, JURIX 2026, Toulouse, Frankreich

ZERTIFIZIERUNGEN

- Agentic AI with LangGraph, CrewAI, AutoGen and BeeAI · IBM, Coursera · Okt 2026 · Credential-ID P4EUDL3A0P6C
- Agentic AI with LangChain and LangGraph · IBM, Coursera · Okt 2026 · Credential-ID ODTQFYNKUQLH
- Agentic AI Protocols (MCP, A2A, ACP) · Coursera · Feb 2026 · Credential-ID K2CXL6L47YK3
- IBM Data Science Specialization · Coursera · Mai 2020
- Deep Learning (Neural Networks and Deep Learning; Improving Deep Neural Networks; Structuring Machine Learning Projects) · Coursera · Aug 2017

AUSBILDUNG

RWTH Aachen, Deutschland **Okt 2015 – Mrz 2020**

MSc Software Systems Engineering

- Projekte: Automatische Spracherkennung, Deep Robot Learning (Computer Vision/Reinforcement Learning)
  Präsentationen: Überblick und Vergleich von End-to-End-Recurrent-Neural-Network-Systemen mit Attention Methoden (Basis heutiger LLMs)

punit@mohantypunit.com +49 176 58 04 52 74 mohantypunit.com

Führung: Von internationalen Studierenden zum Vizepräsidenten der Ausländervertretung gewählt; ein Jahr lang im Amt tätig.

- Masterarbeit: „Multi-agent Deep Reinforcement Learning for Production Scheduling in Flexible Manufacturing Systems“- sehr gut (1,3), Prüfer Univ.-Prof. Dr. Gerhard Lakemeyer (Knowledge-Based Systems Group). Gesamtnote gut (2,1).

BPUT - College of Engineering and Technology, Bhubaneswar, Indien (Odisha University of Technology and Research) **Okt 2008 – Aug 2012**

B.Tech. Information Technology

Projekte: Entwicklung einer Web-Plattform, um Busverbindungen in unserer Region einfach zu suchen und Preise zu vergleichen (im Rahmen eines Praktikums bei einem Start-up).

SPRACHKENNTNISSE

Deutsch	Fließend (C1)
Englisch	Fließend

punit@mohantypunit.com +49 176 58 04 52 74 mohantypunit.com