



# Philip Vana

## AI Engineer

LLM-Evaluation · Retrieval-Engineering · RAG-Systeme · EU-souveräne Architekturen

## PROFIL

AI Engineer mit End-to-End-Ownership einer RAG-Plattform im öffentlichen Sektor — von der Ingestion-Pipeline über Retrieval-Optimierung bis zu einem evaluierten AWS-Bedrock-Deployment. Mein Hintergrund als Data Engineer (ETL, Data Warehousing, Large-scale Data) prägt einen system-first, datengetriebenen Ansatz — optimiert für Reproduzierbarkeit, Messbarkeit und Betrieb. Ich denke in Systemen, Datenflüssen und Failure Modes.

## ERFAHRUNG

### Data Engineer · GenAI- / RAG-Fokus

2024–Heute

Deutsche Rentenversicherung Bund, Berlin

- Seit 10/2025: die End-to-End-RAG-Pipeline eines internen GenAI-Assistenten entworfen und gebaut — eine AWS Bedrock Knowledge Base über ~800 Dokumentationsseiten, entwickelt im Pilot-Scrum-Team, eng abgestimmt mit Product Ownern und Architekten. Deployt und im User-Acceptance-Testing vor dem Produktions-Rollout (Aug. 2026).
- Die Hypothese aufgestellt, dass Bedrocks Standard-Chunking für den Content nicht trägt — in frühen Messungen bestätigt bei ~30 % Recall@5 — und es durch eine content- und source-aware Chunking-Strategie ersetzt: 78 % Recall@5 (MRR 0,66) auf dem experten-validierten Gold-Set (n=56) erreicht, jeder Hebel isoliert gemessen gegen ein eigens gebautes Evaluations-Harness — drei Test-Suiten, ein 5-Prozentpunkt-Recall-Regression-Gate (CI-blockierend), Bootstrap-Konfidenzintervalle und automatisiertes Stakeholder-Reporting.
- Eine EU-datenresidente, agentische KI-Entwicklungsumgebung in AWS SageMaker gebaut (opensource → AWS Bedrock via LiteLLM-Proxy) — credential-frei über IAM-Execution-Roles, automatisiertes Single-File-Bootstrapping, selbstheilender Lifecycle.
- Data-Engineering-Fundament im BI-Team (SAP HANA / Data Warehouse): End-to-End-Ownership von produktivem Batch-ETL im Milliarden-Datensatz-Maßstab — Monitoring, Metadaten, Root-Cause-Analyse, Produktions-Deployments.

### Software Engineer (Praktikum)

Jun. 2023–Aug. 2023

Vaultree Ltd., Dublin

- Objektorientierte Backend-Entwicklung (Java, Python) in einem internationalen, produktorientierten Team; verschlüsselungsfokussierte API-Anwendungen (PostgreSQL, REST-APIs, Vaultree Encryption API).

Frühere Erfahrung (2014–2021): Projekt- & Eventmanagement (Wien) und Fahrzeugprüfungs-Engineering (KÜS), vor dem Wechsel in die Tech-Branche.

Ausgewählte eigene Systeme: ein öffentlicher EU-souveräner RAG-Service über den EU AI Act + die DSGVO, live auf philip-vana.com (FastAPI, Qdrant, EU-only-Inferenz) · eine spec-getriebene SAP-HANA-Warehouse-Engine mit verlustfreien, byte-stabilen Schema-Roundtrips (453 Tests) · eine agentische Coding-Umgebung mit Tier-Routing der Modelle hinter einem zentralen Gateway-Guardrail.

## AUSBILDUNG

### Fachinformatiker, Daten- & Prozessanalyse (duale Ausbildung)

2021–2024

Deutsche Rentenversicherung Bund, Berlin · Abschlussnote: 1,7

### Mittlere Reife

2007–2012

Wilhelm-Busch-Realschule, München · Abschlussnote: 2,0

## KONTAKT

Berlin, Deutschland

+49 163 5217112

philip@philip-vana.com

philip-vana.com

linkedin.com/in/philip-vana

## SKILLS

### Generative AI / RAG

RAG · AWS Bedrock (Knowledge Bases, Titan, Rerank) · Vektorsuche · Embeddings · Chunking · Query-Expansion · LLM-Evaluation (Recall@k, MRR, nDCG) · LLMops

### Cloud & Infrastruktur

AWS (Bedrock, SageMaker, S3, Lambda, DynamoDB) · IAM · AWS CDK · CodeBuild · LiteLLM

### Data Engineering

ETL / ELT · Data Warehousing · Data Modeling · Batch- & Chunk-Processing · SAP HANA XSA · Oracle DB · PostgreSQL

### Sprachen & Tools

Python · SQL · FastAPI · Pytest · Git

## SPRACHEN

Deutsch Muttersprache

Englisch C1, fließend