

Jonathan Wrede

Website: jonathanwrede.de
E-Mail : kontakt@jonathanwrede.de
GitHub : github.com/Jwrede
LinkedIn : linkedin.com/in/jonathan-wrede
Mobil : +49 151 - 29499386

Platform & Observability Engineer. Ich betreibe produktive Kubernetes-, Observability- (Prometheus, Grafana,

OpenTelemetry) und Daten-/ML-Plattformen im Flottenmaßstab, aktuell für 100.000+ IoT-Geräte, und bringe dieselbe Produktionsreife in LLM-Inferenz-Infrastruktur ein. 10+ gemergte OSS-Infrastruktur-PRs in llm-d, OpenTelemetry und Feast; öffentliche Tools mit aipreflight, llmprobe und einem Live-LLM-Benchmark. Zuvor habe ich Cloud-Datenplattformen für 5 Unternehmenskunden umgesetzt.

BERUFSERFAHRUNG

- Vilisto** Hamburg
- *Machine Learning Engineer* Jan. 2025 - Heute
- o **Batterie-Analytics-Plattform (alleiniger Entwickler):** End-to-end Batterie-Analytics-Plattform für 100.000+ IoT-Thermostate (~400 Kundenstandorte) entworfen und gebaut: getestete, typgeprüfte Bibliothek zur Kapazitäts-/Laufzeitschätzung (**Polars**, veröffentlicht im internen **PyPI**), **Dagster**-Ingestion in partitioniertes Parquet/MinIO, **FastAPI**-Services, **Streamlit**-Dashboards und Health-Monitoring mit Alerting; in das firmeneigene **Vue**-Dashboard integriert. Batteriewechselkosten um EUR 50-100K im ersten Jahr gesenkt.
 - o **Plattform & Observability:** Betrieb und Monitoring von Daten-/ML-Services auf **Kubernetes (Helm, Probes, Logs/Metriken)**; **Grafana**-Dashboards und **Python**-Alerting für 100.000+ Geräte gebaut, wodurch tagelanges manuelles Prüfen durch automatische Erkennung von Offline-Thermostaten und Gateway-Ausfällen ersetzt wurde.
 - o **Datenplattform (ELT):** Firmen-Analytics-Warehouse gebaut und produktiv gesetzt: **dbt**-Modelle (Staging und Marts) über IoT-Gerätelemetrie, PII-sichere Pipelines, automatisierte Tests/Lineage via **Dagster**, stufenweise **ArgoCD**-Rollouts in Produktion.
- synvert (vormals Saracus)** Münster
- *Data Platform Consultant* Jan. 2023 - Dez. 2024
- o **Cloud-Datenplattformen:** Data Warehouses für 5 Unternehmenskunden in Versicherung, E-Commerce und Energie auf **Snowflake (AWS)**, **Google BigQuery** und **Azure Synapse** entworfen und deployed; automatisierte **Python**-ETL-Pipelines für Millionen von Datensätzen.
 - o **Apps & Governance:** **Streamlit**-Apps mit OAuth-SSO und Snowflake-Integration gebaut (deployed via **Azure Container Apps**); **Master-Data-Management** mit **Ataccama** + **Azure** implementiert.
- Vilisto** Hamburg
- *Werkstudent - Data Scientist* Sep. 2021 - Okt. 2022
- o **ML, Data & Software:** Präsenzerkennungsmodelle mit **Python/scikit-learn** erweitert; **PostgreSQL/TimescaleDB**-Pipelines, Backend-/Frontend-Features (**Docker, Vue.js**) und ein **PyQt**-Labeling-Tool gebaut.

OPEN SOURCE & INFRASTRUKTUR-Projekte

- **Inferenz-Plattform OSS-PRs:** 10+ gemergte Infrastruktur-PRs in llm-d (Router, Inference-Sim), OpenTelemetry und Feast: Router-CI-Hardening und Flow-Control-Konfiguration, OpenAI-Telemetrie, BigQuery/DynamoDB/Trino-Feature-Store-Fixes; mit aktiven offenen Beiträgen in llm-d-kv-cache und llm-d-benchmark.
- **aipreflight / llmprobe:** Open-Source Readiness-Gate für LLM-/RAG-Services: Eval-Checks, Kostenbudgets, externe Probes, SLA-Gates, Prometheus/vLLM-Korrelation, Kubernetes-Manifeste, Runbooks, Tests und CI.
- **tokenoll:** Infracost-ähnliches Tool für LLM-Kosten: statische Analyse von OpenAI/Anthropic/Google/LiteLLM/LangChain-Call-Sites, Pricing-Cache, GitHub-PR-Kommentare und Kostenwirkungs-Reporting für Modelländerungen; in CI von öffentlichen AI-Repositories inklusive gpt-researcher genutzt.

BILDUNG

- Universität Münster** Münster
- **Master Informatik** Schwerpunkt KI / ML / Statistik Okt. 2023 - Okt. 2026
Thesis: Memory profiling and deployability of privacy-preserving transformer inference systems
- TU Dortmund** Dortmund
- **Bachelor Informatik** mit Nebenfach Statistik Okt. 2019 - Apr. 2023
Thesis: Phase transitions in sparse principal component analysis

TECHNISCHE KENNTNISSE

- **Infrastruktur, Cloud & Observability:** Kubernetes, Helm, ArgoCD, Docker, Terraform, Linux, AWS, Azure, GCP, GitHub Actions, Jenkins, Prometheus/Grafana, OpenTelemetry, CI/CD
- **Inferenz & ML:** llm-d, vLLM, SGLang, KServe, Readiness-Gates, TTFT/TPOT-Benchmarking, LLM-APIs, scikit-learn, Feature Stores/Feast
- **Daten- & ML-Plattformen:** Python, Go, SQL, dbt, Dagster, MLflow, PostgreSQL/TimescaleDB, Snowflake, BigQuery, Azure Synapse, Pandas/Polars
- **Backend, Tests & Apps:** FastAPI, pytest, TypeScript, Streamlit, Vue.js, OAuth SSO, Azure Container Apps, MinIO