# 1_Defensive-Publication_Der_Prior-Art_Schutz_NWM_v3.0

# Defensive Publication: Dezentrale KI-Systemarchitektur für souveräne Organisationen & Einzelanwender

**Projekt-Code:** #NWM (Netz-Werkzeug.Macherei)
**Status:** Öffentlich – v3.0
**Datum:** 260415 (15.04.2026)
**Verfasser:** Projektteam Netz-Werkzeug.Macherei

**Motto:** *"Das sind keine Chats, das sind Schätze!"*

---

## 1. Technisches Feld

Die vorliegende Offenlegung betrifft ein informationstechnisches System zur Integration von Künstlicher Intelligenz (KI) in die Infrastruktur von souveränen Organisationen (KMUs, Vereine, NGOs, Gemeinden, Interessensgemeinschaften) sowie datenschutzbewussten Einzelanwendern (Heim-Server-Betreiber). Insbesondere beschreibt sie ein hybrides Architekturmodell, das lokale Datenverarbeitung (On-Premise) mit cloud-basierten KI-Diensten unter Wahrung der Datensouveränität verbindet.

**Zweck dieser Veröffentlichung:** Sicherung des Standes der Technik/Wissenschaft (Prior Art) zur Verhinderung von Patentansprüchen Dritter auf das beschriebene Systemdesign.

Hinweis: Diese Veröffentlichung beschreibt ein Architekturkonzept. Die tatsächliche Sicherheit im Betrieb hängt von der korrekten Implementierung und der Einhaltung von Sicherheitsrichtlinien durch den Betreiber ab.

---

## 2. Hintergrund

Bestehende Lösungen zur KI-Integration neigen dazu, entweder vollständig cloud-basiert zu sein (mit Risiken hinsichtlich Datenschutz, Datenabfluss und Vendor-Lock-in) oder vollständig lokal zu operieren (mit Einschränkungen hinsichtlich Rechenleistung und Modellkapazität). Für Organisationen und Einzelanwender mit hohen Datenschutzanforderungen (z.B. aufgrund von DSGVO, Berufsgeheimnissen oder persönlicher Präferenz) besteht ein technisches Problem darin, leistungsfähige KI-Modelle zu nutzen, ohne sensible Daten unkontrolliert an externe Anbieter zu übermitteln.

Es besteht ein Bedarf an einem System, das diese Nachteile überwindet, indem es eine hybride Orchestrierung mit mehrschichtigen Sicherheitsmechanismen bietet.

---

# 3. Systemarchitektur

Das beschriebene System umfasst folgende Hauptkomponenten:

## 3.1 Zentrale Wissensplattform

Als zentrale Ablage- und Kollaborationsplattform dient eine selbstgehostete oder managed Cloud-Instanz (beispielsweise basierend auf Nextcloud-Technologie). Diese fungiert als Single Source of Truth für Dokumente, Kalender, Kontakte und strukturiertes Wissen (Wiki-Funktionalität). Die Datenhoheit verbleibt ausschließlich beim Betreiber der Instanz.

- **Struktur:** Trennung von "Bühne" (geteiltes Wissen via Collectives) und "Küche" (internes Denken via lokaler Clients wie Obsidian).
- **Synchronisation:** Verschlüsselte Synchronisation zwischen lokalem Client und zentraler Plattform.
- **Projektmanagement:** Nutzung von Kanban-Boards (z.B. Deck) für Aufgabensteuerung.
- **Hosting:** Eigener Server (Linux, Windows, Proxmox, VMware, etc.) oder Managed Hosting. Empfehlung: Proxmox/Linux für maximale Kontrolle.

## 3.2 Hybrider KI-Orchestrator

Ein zentrales Merkmal des Systems ist eine Orchestrierungsschicht, die eingehende Anfragen analysiert und routet. Das System ist **Gateway-agnostisch**, nutzt jedoch referenziell Multi-Model-Gateways (z.B. OpenRouter.ai) zur Implementierung.

- **Lokale Verarbeitungsebene (iKi - interne KI):**
  - Anfragen, die sensible personenbezogene Daten (PII) oder vertrauliche Informationen enthalten, werden primär an lokale KI-Modelle weitergeleitet.
  - Diese laufen auf einer lokalen Hardwareeinheit (z.B. Workstation mit GPU-Unterstützung) innerhalb des lokalen Netzwerks.
  - Genutzte Software umfasst Open-Source-Lösungen zur Modellausführung (z.B. Ollama, LM Studio, Msty Studio).
  - Modelle: Lokale LLMs im Bereich 8B bis 13B Parameter (z.B. Llama-3-8B, Mistral).
  - Einsatz: Datenschutzkritische Anfragen, erste Filterung aus persönlichen Quellen, PII-Anonymisierung.
- **Cloud-Verarbeitungsebene (cKi - cloudbasierte KI):**
  - Anfragen, die keine sensiblen Daten enthalten und/oder eine höhere Modellkapazität erfordern, werden an externe KI-Dienste über ein Gateway weitergeleitet.
  - Modelle: Größere Modelle bei Bedarf (z.B. Qwen, Gemini, Grok, DeepSeek).

    - Datenschutz: Einstellung "EU-Only" wo möglich; Nutzung eines lokalen PII-Filters & Gateway-PII-Filters (Guardrails) zur Anonymisierung sensibler Daten vor dem Cloud-Versand. Spezielle Einstellungen können separaten API-Schlüsseln zugeordnet werden.
- **Routing-Logik (Orchestrator-gesteuert):**
    - Der KI-Orchestrator übernimmt das Routing basierend auf konfigurierten Regeln (Sensibilität, Komplexität).
    - Der Nutzer bestätigt die Cloud-Nutzung bei sensiblen Kontexten (Human-in-the-Loop).
    - Eine manuelle Überschreibung der Modellwahl pro Frage ist möglich, aber nicht Standard.

| Fall | Sensibilität | Leistung | Lösung |
|---|---|---|---|
| **A** | - | - | → iKi/cKi (Lokale Ressourcen bevorzugt) |
| **B** | - | + | → cKi |
| **C** | + | - | → iKi |
| **D** | + | + | → L1 / L2 / L3 (siehe unten) |

**Lösungsoptionen für Fall D (Sensibel UND Hochleistung):**

- **L1:** Lokale Anonymisierung (iKi) → Dann Cloud (cKi); Standard-Fall.
- **L2:** In mehrere Fragen aufteilen, nur über iKi (Verzicht auf Hochleistung).
- **L3:** Direkt mit cKi (Akzeptanz des minimalen Restrisikos via Human-in-the-Loop).
- **Wissensanbindung (RAG):**
    - Technologie: Retrieval-Augmented Generation (RAG) (z.B. via LlamaIndex, ChromaDB oder Vektor-Datenbanken).
    - Funktion: Die iKi zieht relevante Dokumente aus der Nextcloud/Wissensdatenbank heran (Token-Effizienz).
    - Orchestrierung: Router Service (Python/FastAPI + litellm) oder Skripte (n8n/Node-RED) für Workflows.

## 3.3 Hardware- und Software-Referenz

Zur Umsetzung des Verfahrens wird eine modulare Hardwarestruktur beschrieben:

- **Server-Komponente:** Hostet die Wissensplattform und Vektor-Datenbanken. Betrieb auf eigener Hardware (Linux, Windows, Proxmox, VMware) oder Managed Hosting.
    - Speicher: Nextcloud-Instanz mit ausreichendem Platz für Dokumente und Vektor-Datenbanken; zusätzliche SSD-Speicher an Server (Datenmengen/Backup).
- **Client-Komponente (iKi-Maschine):** Verfügt über lokale Rechenressourcen (GPU mit ausreichendem VRAM, beispielhaft ≥16 GB, z.B. RTX 4060 Ti 16GB) zur Ausführung kleinerer bis mittlerer KI-Modelle (beispielsweise 8B bis 13B Parameter).

- **Betriebssysteme:** Windows, Linux, macOS – alle drei unterstützbar.
- **Software-Stack:** Nutzung von Open-Source-Software für die Plattform, Containerisierung (z.B. Docker) für die Dienste und skriptbasierte Automatisierung für die Orchestrierung.
- **Clients:** Obsidian, LM Studio, Msty Studio, Browser.

---

# 4. Verfahren zur Datensicherheit (Defense-in-Depth)

Das Verfahren umfasst spezifische, mehrschichtige Sicherheitsebenen zur Gewährleistung der Souveränität:

## 4.1 Lokale Vorfilterung (iKi-Ebene)

Bevor Daten das lokale Netzwerk verlassen, durchlaufen sie eine interne KI-Schicht (iKi).

- **Funktion:** a) Automatische Erkennung und Anonymisierung von personenbezogenen Daten (PII) sowie vertraulichen Informationen. b) Automatische Filterung der internen Daten aus den bereitgestellten Datenbanken.
- **Mechanismus:** Die iKi formuliert Anfragen so um, dass der semantische Gehalt erhalten bleibt, aber der Bezug zur ursprünglichen Identität oder zum Geheimnis getrennt wird.
- **Ziel:** a) Reduktion des Datenrisikos vor dem Cloud-Transfer. b) Reduktion der hochgeladenen Tokenanzahl, sowie heranziehen, aber dennoch Verschleiern des eigenen Winnsens des Wiki-Kontexts. [bitte g'scheid umformulieren]

## 4.2 Human-in-the-Loop (Hoheitliche Firewall)

Der Nutzer behält die finale Entscheidungsgewalt über jeden Datenfluss.

- **Funktion:** Validierung der anonymisierten Anfrage vor dem Absenden an die Cloud-Ebene (cKi).
- **Mechanismus:** Keine automatische Weiterleitung ohne Nutzerbestätigung bei sensiblen Kontexten.
- **Ziel:** Der Mensch ist die letzte Instanz („Hoheitliche Firewall") und trägt die Verantwortung für die Freigabe.

## 4.3 Anbieter-Souveränität & Routing-Strategie

Das System vermeidet Abhängigkeiten von einzelnen Anbietern durch aktive Diversifizierung und strategische Auswahl.

- **Modell-Wahl:** Der Nutzer definiert die Regeln für Modell-Hersteller (z.B. Google, Microsoft, Europäische/Asiatische Anbieter) und spezifische Modelle. Der Orchestrator wendet diese Regeln an.

- **Infrastruktur-Randomisierung:** Nutzung von Gateways, die im Hintergrund Rechen-Dienstleister dynamisch verteilen.
    - **Effekt:** Durch die Verteilung der Anfragen auf verschiedene Rechenstandorte und Anbieter wird die Profilbildung durch Metadaten-Analyse (Traffic-Mustering) erschwert.
- **ZDR-Auswahlkriterium:** Die Auswahl der genutzten Infrastruktur erfolgt primär unter Zuhilfenahme von **Zero-Data-Retention (ZDR)**-Richtlinien.
    - **Kombinationseffekt:** Die Randomisierung schützt die Metadaten (Wer fragt wann?), während die ZDR-Richtlinien den Inhalt schützen (Was wird gespeichert?). Beides zusammen minimiert das Risiko der Nachverfolgung.

## 4.4 Multi-Level-Guardrails (Richtlinien-Ebenen)

Zur Durchsetzung der Sicherheitspolitik werden vier Ebenen von Guardrails kombiniert:

1. **Ebene 1 (ZDR-Policy):** Sicherstellung von „Zero Data Retention" durch den Gateway-Anbieter (keine Speicherung der Anfragen zu Trainingszwecken).
2. **Ebene 2 (Regionale Compliance):** Erzwingung von Rechenstandorten innerhalb bestimmter Jurisdiktionen (z.B. „EU-Only"), um datenschutzrechtlichen Anforderungen zu genügen.
3. **Ebene 3 (Inhaltsfilter):** Zusätzliche PII-Kontrolle und Moderation auf Gateway-Ebene vor der Verarbeitung durch das Zielmodell.
4. **Ebene 4 (API-Schlüssel-Spezifika & Obfuskation):**
    - **Funktion:** Möglichkeit, pro API-Schlüssel individuelle Guardrails zu definieren (z.B. unterschiedliche Retention-Policies für verschiedene Themenbereiche oder Teams).
    - **Pooling-Effekt:** Bei Nutzung eines gemeinsamen API-Schlüssels durch mehrere Nutzer oder für stark variierende Eingaben (Stil, Ton, Inhalt) wird die Rückverfolgbarkeit einzelner Anfragen auf einen spezifischen Endnutzer statistisch erschwert (Identitäts-Obfuskation).
    - **Ziel:** Erhöhung der Anonymität im Datenstrom durch Vermischung von Metadaten.

## 4.5 Netzwerk-Sicherheit (Perimeter)

Die Infrastruktur folgt dem „Zero Trust"-Ansatz.

- **Keine offenen Ports:** Es werden keine Ports für den allgemeinen Zugriff aus dem öffentlichen Internet geöffnet.
- **Zugriff:** Ausschließlich über verschlüsselte VPN-Tunnel (z.B. Tailscale, WireGuard).
- **Segmentierung:** VLAN-Trennung zwischen KI-Systemen, Nutzer-Clients und Server-Infrastruktur, wo möglich.

**Fazit des Sicherheitskonzepts:**
Ein mehrschichtiges Sicherheitskonzept durch Kombination aus technischer Filterung,

menschlicher Kontrolle und Anbieter-Vielfalt. Das Risiko eines Datenlecks wird nach aktuellem Stand der Technik auf ein Minimum reduziert.

---

## 5. Lizenzierung & Patentschutz

- **Lizenz für Dokumentation: CC BY-NC-SA 4.0**.
  - *Wirkung:* Dritte müssen den Urheber nennen, dürfen nichts kommerziell damit verdienen (Doku-Text), müssen Abwandlungen gleich lizenzieren.
- **Patentschutz (Prior Art):**
  - **Methode: Defensive Publication**.
  - **Begründung:** Die **Veröffentlichung selbst** (Zeitstempel) etabliert den "Stand der Technik/Wissenschaft". Dies verhindert, dass Dritte (z.B. BigTech) die Idee später patentieren können.
  - **Wichtig:** Nicht die Lizenz schützt vor Patenten, sondern die **Veröffentlichung selbst** (Zeitstempel).

---

## 6. Schlussfolgerung

Durch diese Veröffentlichung wird ein technischer Stand der Technik/Wissenschaft etabliert, der eine dezentrale, datensouveräne KI-Integration für souveräne Organisationen und Einzelanwender beschreibt. Das spezifische Zusammenspiel aus lokaler Orchestrierung, RAG-Anbindung an eine selbstgehostete Cloud-Plattform und hybrider Modellnutzung unter Beachtung mehrschichtiger Sicherheitsguardrails stellt eine neuartige Lösung für das Problem der Datensouveränität bei gleichzeitiger Nutzung moderner KI-Leistung dar.

Dieses Dokument dient als Prior Art, um die freie Nutzbarkeit dieser Architektur für alle Marktteilnehmer zu sichern und eine Monopolisierung durch Patentansprüche Dritter zu verhindern.

**Leitsatz:** *"Wir glauben an ein Ökosystem, in dem Datensouveränität ein Grundrecht und kein Luxusgut ist."*

---

*"... mit dem Sinn für Verbindungen und Zusammenhänge."*

*--- Ende der Defensive Publication v3.0 (Öffentlich) ---*