

WHITEPAPER · NOVEMBER 2026 · 1. AUFLAGE

Das KI-Mesh.

Wie aus den Rechnern, die du schon hast, eine KI wird, die in Deutschland bleibt.

von Matthias Eger
matthiaseger.de

Member of GemeinsamKI.de

Stand: Weiden, 08.11.2026

Matthias Eger
Pfarräckerstr. 14 · 92637 Weiden i.d.OPf.
Tel. +49 (0) 961 634 32 61 · Mobil +49 (0) 175 585 73 05
mail@matthiaseger.de · www.matthiaseger.de

VORWORT

Die Rechner stehen schon da.

Ich richte seit 2004 Rechner ein. Erst im IT-Systemhaus, dann vierzehn Jahre im Konzern, seit 2008 für Betriebe in Weiden und der Oberpfalz. Dabei habe ich etwas gesehen, das mir lange nicht aufgefallen ist, weil es so normal wirkt: **Die Rechner stehen die meiste Zeit nur herum.**

Der Büro-PC wartet, bis jemand tippt. Die Workstation mit der guten Grafikkarte rendert zweimal pro Woche. Der Mac des Chefs schläft ab 18 Uhr. Rechenleistung, die bezahlt ist, abgeschrieben wird und im Leerlauf Strom zieht – während derselbe Betrieb jeden Monat Geld an ein Rechenzentrum in den USA überweist, damit dort eine KI seine Texte liest.

Mein Kollege Deniz Fritsch und ich haben uns irgendwann gefragt: warum eigentlich? Wir führen gemeinsam die ResultOS UG. Deniz betreibt dort unser Rechenzentrum, ich baue die Software. Aus der Frage ist **GemeinsamKI.de** geworden: ein KI-Mesh, in dem Macs, Windows-PCs und Linux-Rechner ihre freie Rechenleistung teilen und zusammen eine KI bilden, die in Deutschland läuft. Jedes Gerät, das mitmacht, wird zur KI-Rechenmaschine. Der Grundsatz ist einfach: Wer teilt, nutzt mit.

Dieses Whitepaper erklärt, wie das funktioniert, was es einem Betrieb bringt und wo die Grenzen liegen. Es ist kein Prospekt. GemeinsamKI ist in der Beta; manches ist fertig, manches im Bau, und ich schreibe beides auf. Was hier steht, ist der Stand vom November 2026. Die Einzelheiten stehen öffentlich auf gemeinsamki.de und ändern sich mit jeder Version.

Für mich ist das Mesh die wörtlichste Form meines Satzes „Ich verbinde die echte Welt mit der KI“: Die Rechner, die ich bei dir aufstelle, sind die echte Welt. Das Mesh macht aus ihnen die KI-Welt. Dazwischen steht ein Netzwerk, das sauber gebaut sein muss – und einer, der dafür geradesteht.

Matthias Eger

ZUSAMMENFASSUNG

Zum Mitnehmen.

- In fast jedem Betrieb steht Rechenleistung herum, die bezahlt ist und nichts tut. Gleichzeitig wandern Daten und Geld in fremde Rechenzentren.
- GemeinsamKI verbindet diese Rechner über einen Hub zu einem Mesh. Jeder Knoten teilt, wann und wie viel er will: Play, Stop, Zeitplan, Speicherlimit.
- Im Mesh laufen starke offene Modelle, zum Beispiel gpt-oss-20b, Qwen3 32B, Gemma 3 27B oder Mistral Small 24B. Genutzt wird es per Chat im Browser, in VS Code und über eine OpenAI-kompatible API.
- Datenschutz steckt in der Architektur: Eine Clearingstelle ersetzt Personendaten durch Platzhalter, bevor ein fremdes Gerät den Text sieht. Ein Wächter prüft nach. Drei Vertrauensstufen regeln, welches Gerät was bekommt.
- Cluster machen aus dem großen Mesh einen eigenen Kreis: Nur die Geräte deines Betriebs rechnen für deinen Betrieb.
- Für Betriebe heißt das: vorhandene Hardware statt Abo pro Kopf, Daten in Deutschland, Rechenzeit nachts, offene Modelle ohne Lock-in – und eine Schnittstelle, an die sich eigene Werkzeuge anschließen lassen.
- Ehrlich gesagt: Die Beta braucht eine Freischaltung, die Clearingstelle pseudonymisiert und anonymisiert nicht, Bilder lassen sich nicht schützen, und die Modelle im Mesh sind kleiner als die größten Cloud-Modelle. Für die meisten Aufgaben im Betrieb reicht das. Für manche nicht.

Inhalt

1 Das Problem: Rechenleistung hier, Daten dort
2 Die Idee: Viele Rechner, ein Netz
3 Was im Mesh läuft
4 Datenschutz by Design: Die Clearingstelle
5 Cluster: Dein Betrieb als eigener Kreis
6 Die Mehrwerte für deinen Betrieb
7 Was das Mesh (noch) nicht ist
8 So fängst du an
9 Wer dahintersteht
Glossar

KAPITEL 1

Das Problem: Rechenleistung hier, Daten dort.

1.1 Was im Betrieb steht

Rechne einmal durch, was in deinem Betrieb an Rechnern steht. Ein Büro mit acht Arbeitsplätzen hat acht PCs oder Laptops, oft einen Server oder ein NAS, manchmal eine Workstation für CAD, Bildbearbeitung oder Videoschnitt. Jeder dieser Rechner ist gekauft, abgeschrieben und wird gewartet. Und jeder läuft im Mittel vielleicht ein Zehntel seiner Zeit unter Last. Den Rest wartet er: auf Eingaben, auf den nächsten Morgen, auf das Wochenende.

Bei einem Gerät fällt das nicht auf. Über einen ganzen Betrieb gerechnet ist es ein kleines Rechenzentrum, das die meiste Zeit nichts tut. Über ein Land gerechnet ist es ein riesiges. Genau so steht es in der Mission von GemeinsamKI: Millionen Rechner in Deutschland langweilen sich die meiste Zeit. Zusammen wären sie ein Rechenzentrum, für das niemand ein Gebäude bauen muss.

1.2 Wohin Geld und Daten gehen

Gleichzeitig läuft KI im Mittelstand heute fast immer so: Abo pro Kopf bei einem US-Anbieter, bezahlt pro Monat, berechnet in einem Rechenzentrum, dessen Adresse du nicht kennst. Das funktioniert, und ich sage niemandem, er solle es lassen. Aber es hat drei Haken.

- **Die Kosten.** Pro Kopf und Monat klingt klein. Über ein Team und ein Jahr ist es ein fester Posten, der nie kleiner wird – egal, ob jemand das Werkzeug nutzt oder nicht.
- **Die Daten.** Jede Anfrage verlässt den Betrieb. Für einen Blogtext ist das egal. Für ein Angebot mit Kundenname, eine Mahnung mit IBAN oder einen Arztbrief ist es ein Problem – rechtlich, und im Zweifel auch vor dem Kunden.
- **Die Abhängigkeit.** Wer seine Abläufe auf ein Modell baut, das morgen teurer, anders oder abgeschaltet sein kann, baut auf Sand. Im Whitepaper „Der Fluch der KI“ habe ich das ausführlich beschrieben.

1.3 Der dritte Weg

Zwischen „alles in die Cloud“ und „gar keine KI“ gibt es einen dritten Weg: KI, die auf eigener oder vertrauter Hardware läuft, in Deutschland, mit offenen Modellen. Bisher scheiterte er oft am Aufwand. Ein einzelner Rechner ist zu klein für große Modelle, ein eigener KI-Server zu teuer für einen kleinen Betrieb. Das Mesh löst genau das: **Es macht aus vielen kleinen Rechnern einen großen.**

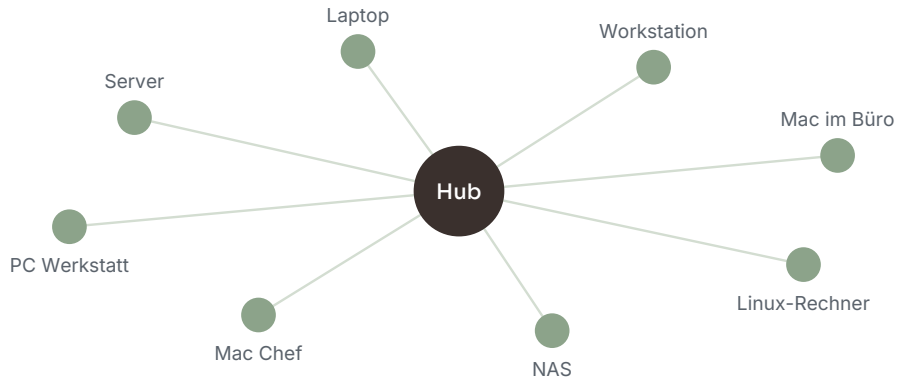
> WORUM ES GEHT
>
> Nicht um eine weitere KI. Sondern um die Frage, **wo** gerechnet wird, **wem** die Rechner gehören und **was** die Daten auf dem Weg dorthin erleben.

KAPITEL 2

Die Idee: Viele Rechner, ein Netz.

2.1 Die Bausteine

Begriff	Was gemeint ist
Mesh	Das Netz aus allen verbundenen Rechnern.
Hub	Die Zentrale, die Anfragen verteilt. Sie läuft auf Servern in Deutschland.
Knoten	Ein Rechner, der gerade Leistung teilt – Mac, Windows-PC, Linux-Rechner oder Server.
Modell	Die offene KI, die auf einem Knoten läuft, zum Beispiel gpt-oss-20b.
Clearingstelle	Das Regelwerk, das Personendaten durch Platzhalter ersetzt, bevor ein Knoten den Text sieht.
Cluster	Ein eigener kleiner Verbund im Mesh – etwa für ein Team, einen Betrieb oder einen Verein.
Agent	Die App auf dem Rechner, die ihn ans Mesh anschließt: Play, Stop, Zeitplan.



Jeder Knoten teilt, wann er will: Play, Stop, Zeitplan. Der Hub verteilt nur – rechnen tun die Knoten.

2.2 Drei Schritte: Anschließen. Freigeben. Mitnutzen.

Anschließen

Der GemeinsamKI-Agent wird installiert: als Setup für Windows, als Paket für Linux, mit einem Befehl auf dem Mac. Ein Einrichtungsassistent verbindet Gerät und Konto – ohne Terminal, ohne Konfigurationsdateien. Auf dem Rechner läuft dazu eine bewährte lokale KI-Laufzeit: LM Studio oder Ollama auf dem Mac, llama.cpp unter Windows und Linux. Der Agent bringt die Modelle, die dort geladen sind, in den Verbund.

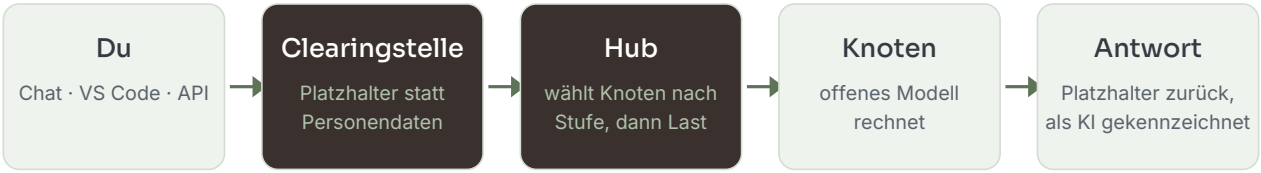
Freigeben

Play startet die Freigabe, Stop beendet sie sofort. Oder du setzt einen Zeitplan: nur nachts, nach Feierabend, am Wochenende, eigene Zeiten. Du legst fest, wie viel Speicher der Rechner teilt. Und Stop heißt Stop: Laufende Anfragen enden, der Rechner ist raus.

Mitnutzen

Im Browser unter chat.gemeinsamki.de, in VS Code und im Terminal mit „GemeinsamKI Code“ oder aus eigenen Programmen über die API unter api.gemeinsamki.de. Zur Wahl stehen alle Modelle, die gerade im Mesh geladen sind. Wer selbst keinen Rechner teilt, kann trotzdem mitnutzen; wer teilt, lässt das Mesh wachsen.

2.3 Der Weg einer Anfrage



TLS-verschlüsselt auf dem ganzen Weg · kein Knoten erfährt, wer gefragt hat

1. **Du fragst** – im Chat, in VS Code oder per API. Die Verbindung ist per TLS verschlüsselt.
2. **Die Clearingstelle** ersetzt Personendaten durch Platzhalter. Im Chat und über die API passiert das im Hub, in VS Code schon auf deinem Rechner, bevor etwas den Rechner verlässt.
3. **Der Hub** wählt einen Knoten mit passendem Modell: zuerst nach Vertrauensstufe, dann nach Last. Kein Knoten erfährt, wer gefragt hat – keine Kontokennung, keine IP-Adresse, keine E-Mail.
4. **Der Knoten** rechnet und schickt die Antwort verschlüsselt zurück.
5. **Bei dir** werden die Platzhalter wieder durch die echten Angaben ersetzt. Die Antwort ist als KI-Antwort gekennzeichnet.

KAPITEL 3

Was im Mesh läuft.

3.1 Offene Modelle

Im Mesh laufen offene Modelle, die jeder lokal betreiben kann: zum Beispiel gpt-oss-20b von OpenAI, Qwen3 32B, Gemma 3 27B oder Mistral Small 24B. Welche Modelle gerade verfügbar sind, hängt davon ab, was die Knoten geladen haben. Die Auswahl wächst mit jedem Rechner.

Das ist ein Unterschied zu den großen Cloud-Modellen, und ich will ihn nicht kleinreden: Die größten Modelle der Cloud-Anbieter sind stärker als ein Modell mit 20 oder 30 Milliarden Parametern. Für die Aufgaben, die im Betrieb am häufigsten vorkommen – Texte zusammenfassen, Mails formulieren, Dokumente ordnen, Daten aus Belegen ziehen, Code schreiben – reicht ein gutes offenes Modell heute aus. Für sehr lange, sehr anspruchsvolle Aufgaben nicht immer. Das ist der ehrliche Stand, und er verschiebt sich jedes Jahr zugunsten der offenen Modelle.

3.2 Was dein Rechner leisten kann

Der Geräte-Check auf gemeinsamki.de zeigt, was ein Gerät im Mesh bewegt. Die Zahlen sind grobe Schätzungen für 4-Bit-quantisierte Modelle:

Gerät	Speicher	Modelle bis etwa	Beispiele
Mac mit Apple Silicon (Pro-Chip)	48 GB gemeinsamer Speicher	60 Mrd. Parameter	Qwen3 32B, Gemma 3 27B
PC mit Grafikkarte	16 GB VRAM	25 Mrd. Parameter	Mistral Small 24B, gpt-oss-20b
PC mit Grafikkarte	8 GB VRAM	kleine Modelle	Einstieg, einfache Aufgaben

Macs mit Apple Silicon teilen sich Arbeits- und Grafikspeicher und eignen sich darum besonders für große Modelle; sie sind der Schwerpunkt der ersten Testphase. Windows- und Linux-Rechner mit leistungsfähiger Grafikkarte (ab etwa 8 GB VRAM) werden schrittweise angebunden. Unter Linux läuft der Agent auch auf Servern ohne Desktop.

3.3 Die Werkzeuge

Dienst	Wofür	Stand
Chat · chat.gemeinsamki.de	Das Mesh im Browser fragen – mit Dateien, Bildern und Spracheingabe.	läuft
GemeinsamKI Code · VS Code & Terminal	Programmieren mit den Modellen des Verbunds: Chat, Aufträge, Änderungen – geplant, umgesetzt, geprüft.	läuft
API · api.gemeinsamki.de	OpenAI-kompatible Schnittstelle: eigene Programme mit den Modellen verbinden.	läuft
Hub · hub.gemeinsamki.de	Die Zentrale für Geräte, Cluster, Freigaben und API-Schlüssel.	läuft
Agent · macOS, Windows, Linux	Schließt den Rechner ans Mesh an.	Beta
Suche · search.gemeinsamki.de	Suchmaschine für Deutschland und Europa. Websites anmelden geht schon, die Suche kommt mit dem Index.	im Aufbau
Forum	Austausch mit anderen Teilnehmenden, Fragen, Feedback.	läuft

KAPITEL 4

Datenschutz by Design: Die Clearingstelle.

Datenschutz ist bei GemeinsamKI kein Anhang, sondern Teil der Architektur. Das Mesh ist nach Art. 25 DSGVO ausgerichtet: Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen. Herzstück ist die Clearingstelle.

4.1 Erst schützen, dann rechnen

Die Clearingstelle ist ein Regelwerk – ohne Cloud, ohne KI-Modell, offline. Es arbeitet im Hub und in GemeinsamKI Code gleich. Im Chat und über die API läuft es im Hub; in VS Code läuft es auf deinem Rechner, bevor etwas den Rechner verlässt. Die Erfahrung dafür stammt aus Igakuro, unserer lokalen KI, die personenbezogene Daten in deutschen Texten erkennt und unkenntlich macht.

4.2 Fünf Schritte

1. **Erkennen.** Regeln mit Prüfsummen und Kontext finden 21 Arten personenbezogener Angaben: Namen mit Kontext, E-Mail-Adressen, Telefonnummern, IBAN, Kartennummern, Steuer-IDs, Sozialversicherungsnummern, Ausweisnummern, Krankenversichertennummern, Arztnummern (LANR), Betriebsstättennummern, Institutionskennzeichen, Rezeptnummern, Kfz-Kennzeichen, Straßen mit Hausnummer, Postleitzahl und Ort, Geburtsdaten, IP- und MAC-Adressen, Zugangsdaten und Geheimnisse, Benutzerpfade.
2. **Ersetzen.** Jede Angabe wird durch einen Platzhalter wie [[PER_1]] ersetzt. Im Code-Modus sind die Ersatzwerte formatgleich, damit Programme und Tests weiterlaufen.
3. **Prüfen.** Ein Wächter prüft den schon ersetzten Text unabhängig ein zweites Mal. Findet er einen Rest, entscheidet die Vertrauensstufe des Geräts, ob der Text überhaupt geht.
4. **Rechnen.** Nur die bereinigte Fassung geht an ein Gerät im Mesh – bevorzugt an ein vertrautes.
5. **Zurückführen.** Die Platzhalter in der Antwort werden mit dem „Tresor“ wieder durch die echten Angaben ersetzt – in VS Code auf deinem Rechner, im Chat im Hub, nur für dich.

> BEISPIEL MIT ERFUNDENEN DATEN
>
> **Deine Eingabe:** „Formuliere eine freundliche Zahlungserinnerung an Petra Wagner, Lindenstraße 12, 95028 Hof, IBAN DE12 3456 7890 1234 5678 90.“
>
> **Was das Mesh sieht:** „Formuliere eine freundliche Zahlungserinnerung an [[PER_1]], [[ADR_1]], [[ORT_1]], IBAN [[IBAN_1]].“
>
> **Antwort bei dir:** „Hallo Petra Wagner, sicher ist es im Alltag untergegangen: Die Rechnung für Lindenstraße 12 ist noch offen …“ – die Platzhalter werden erst bei dir wieder ersetzt.

4.3 Pseudonymisieren ist nicht Anonymisieren

Die Clearingstelle **pseudonymisiert** (Art. 4 Nr. 5 DSGVO): Personendaten werden durch Platzhalter ersetzt, die Zuordnung bleibt getrennt und geschützt im Tresor. So lässt sich trotzdem mit Briefen, Akten oder Code arbeiten. **Anonymisieren** wäre: Daten so entfernen, dass sich niemand mehr bestimmen lässt. Das verspricht die Clearingstelle nicht. Bleibt ein Rest stehen, kann er Personenbezug haben. Deshalb der Wächter, deshalb die Vertrauensstufen. Dass GemeinsamKI diesen Unterschied selbst so deutlich aufschreibt, ist für mich einer der wichtigsten Punkte am ganzen Projekt.

4.4 Drei Vertrauensstufen

Stufe	Welche Geräte	Was sie bekommen
A – vertraut	Geräte von GemeinsamKI und von Geschäftskunden mit Auftragsverarbeitungsvertrag, eigenem Geräteschlüssel und von GemeinsamKI als vertraut eingestuft.	Klartext und Bilder. Werden bevorzugt gewählt.
B – Verschwiegenheit	Geräte, deren Betreiber den Knotenvertrag angenommen und eine Verschwiegenheitsverpflichtung abgegeben haben.	Nur pseudonymisierten Text – auch dann, wenn der Wächter noch einen Rest meldet.
C – offen	Alle übrigen Geräte der Community.	Nur pseudonymisierten Text ohne erkannten Rest. Sonst wird die Anfrage gesperrt oder an eine höhere Stufe gegeben.

4.5 Was sonst noch gilt

- **Verschlüsselt:** Alle Verbindungen zu Website, Hub, Chat und den Rechnern im Verbund laufen per TLS.
- **Server in Deutschland:** Website, Hub und Chat laufen auf Servern in Deutschland. Geräte im Verbund stehen in der EU.
- **Signierte Updates:** Agent, Engine und VS-Code-Erweiterung installieren nur Updates, deren Signatur sie prüfen können.
- **Kein Protokoll ohne Grund:** Das Prompt-Protokoll des Hubs ist standardmäßig aus. Ist es zur Fehlersuche an, zeigt der Chat das an.
- **Anonym gegenüber dem Knoten:** Felder wie user und metadata aus API-Anfragen entfernt der Hub. Kein Gerät erfährt, wer gefragt hat.
- **Zwei-Faktor-Anmeldung** für Konten; für Administratoren Pflicht.

4.6 Die Grenzen, ehrlich

Ein Name ohne erkennbaren Kontext im Fließtext („Sabine hat angerufen“) kann dem Regelwerk entgehen. Dafür gibt es den Wächter und die Stufen, aber keine Garantie. Bilder lassen sich nicht pseudonymisieren; sie gehen nur an vertraute Geräte oder nach deiner ausdrücklichen Freigabe. Und: Gesundheitsdaten, Patienten- oder Mandantendaten und andere brisante Daten gehören nicht in den Chat. Praxen und Kanzleien arbeiten über eine Anbindung mit Igakuro und einen Schutzcluster auf vertrauten Geräten der Stufe A (siehe Kapitel 5).

KAPITEL 5

Cluster: Dein Betrieb als eigener Kreis.

Ein Cluster ist ein eigener kleiner Verbund in GemeinsamKI – für ein Team, einen Betrieb, einen Verein. Mitglieder kommen per Einladung, Geräte per Geräte-Code. Anfragen aus dem Cluster rechnen nur auf seinen Geräten.

5.1 Wie ein Cluster entsteht

1. **Cluster anlegen** im Hub. In der Beta legt das vorerst das GemeinsamKI-Team an.
2. **Mitglieder einladen** – per Einladungslink (sieben Tage gültig) oder über den Benutzernamen. Wer annimmt, sieht vorher Name, Besitzer und Regeln des Clusters.
3. **Geräte einladen.** Der Geräte-Code kommt in den Agenten. Der Betreiber liest den kurzen Cluster-Zusatz zum Knotenvertrag und entscheidet, wo das Gerät rechnet.
4. **Im Chat die Welt wählen:** GemeinsamKI allgemein oder einer deiner Cluster.

5.2 Drei Welten je Gerät

Einstellung	Das Gerät rechnet …
Allgemein	für alle im Mesh – wie bisher.
Nur Cluster	nur für die Mitglieder des Clusters. Für einen Betrieb die naheliegende Wahl.
Beides	allgemein und im Cluster, je nach Anfrage.

Das lässt sich jederzeit umschalten, und austreten geht genauso. Die Clearingstelle bleibt auch im Cluster aktiv; die Vertrauensstufe eines Geräts hebt eine Welt nie. Mitglieder sehen die Geräte des Clusters mit Anzeigename, Stufe und Modellen – nie das Konto des Besitzers, den Rechnernamen oder die IP-Adresse. Besitzer und Verwalter sehen die Nutzung als Zahl, nie die Inhalte.

5.3 Der Schutzcluster für Praxen und Kanzleien

Für Berufsgeheimnisträger gibt es eine strengere Form: Die Software der Praxis oder Kanzlei ersetzt Namen und Kennnummern schon vor dem Senden über Igakuro durch Platzhalter, der Hub prüft nach, und gerechnet wird nur im Schutzcluster auf vertrauten Geräten der Stufe A. Dort rechnet ein Gerät nur, wenn sein Agent meldet, dass es sicher eingestellt ist: keine Protokolle mit Inhalten, nicht aus dem Netz erreichbar, Festplatte verschlüsselt. Unterhaltungen löscht der Hub dort 30 Tage nach der letzten Nachricht.

> KLEINER KREIS, EHRLICH GESAGT
>
> In einem kleinen Cluster kann ein Betreiber ahnen, wer fragt – aus dem Zusammenhang, auch ohne Namen. Platzhalter ersetzen Namen, Adressen und Kontodaten, nicht den Inhalt einer Frage. Frag im Cluster deshalb nichts, was der Kreis nicht wissen soll. Ein Cluster läuft im Hub von GemeinsamKI; einen eigenen Server braucht er nicht, einen eigenen Firmen-Hub ersetzt er aber auch nicht.

KAPITEL 6

Die Mehrwerte für deinen Betrieb.

Was bringt das Mesh einem Betrieb mit zehn, zwanzig oder fünfzig Leuten? Neun Punkte, nüchtern aufgeschrieben – mit dem, was dahinter technisch passiert.

1 · Vorhandene Hardware wird zur KI

Die Rechner stehen schon da und sind bezahlt. Mit dem Agenten werden sie zu Knoten – ohne neue Server, ohne Rechenzentrum, ohne Abo pro Kopf. Ein Mac mit 48 GB bewegt Modelle bis etwa 60 Milliarden Parameter. Drei, vier solcher Geräte in einem Cluster sind für einen Betrieb mehr KI-Leistung, als er an einem normalen Tag braucht.

2 · Daten bleiben in Deutschland – Personendaten bleiben bei dir

Hub und Chat laufen auf Servern in Deutschland, die Knoten stehen in der EU oder gleich in deinem Haus. Vor dem Rechnen ersetzt die Clearingstelle Personendaten durch Platzhalter. Wer eine Mahnung, ein Angebot oder einen Schriftsatz formulieren lässt, muss den Kundennamen nicht mehr aus dem Text streichen, bevor die KI ihn sieht.

3 · Rechenzeit, wenn der Betrieb schläft

Der Zeitplan ist der unterschätzte Teil. Rechner teilen nachts, nach Feierabend oder am Wochenende – genau dann, wenn sie sonst nichts tun. Tagsüber bleibt die volle Leistung beim Mitarbeiter. Der einzige laufende Posten ist der Strom, und den zahlt, wer seinen Rechner bereitstellt.

RECHENBEISPIEL STROM · ANNAHMEN, KEINE MESSWERTE

Ein Mac mini zieht unter KI-Last grob 40 Watt. Teilt er 9 Stunden pro Nacht an 22 Arbeitstagen, sind das rund 8 kWh im Monat – bei 0,35 €/kWh etwa 3 Euro.

Ein PC mit großer Grafikkarte zieht unter Last grob 300 Watt. Gleicher Zeitplan: rund 60 kWh, etwa 21 Euro im Monat. Beides nur für die Stunden, in denen wirklich gerechnet wird; im Leerlauf fällt deutlich weniger an.

Zum Vergleich kostet ein KI-Abo pro Kopf bei den großen Anbietern typischerweise einen zweistelligen Euro-Betrag im Monat – pro Person, nicht pro Gerät. Rechne mit deinen Zahlen, nicht mit meinen.

4 · Offene Modelle ohne Lock-in

Die Modelle im Mesh sind offen und lokal lauffähig. Wird eines abgelöst, lädt der Knoten das nächste. Niemand kann dir das Modell abschalten, das deine Abläufe tragen, und niemand kann über Nacht den Preis pro Anfrage ändern. Du bleibst Eigentümer deiner Werkzeuge – ein Prinzip, nach dem ich ohnehin arbeite.

5 · Eine Schnittstelle für eigene Werkzeuge

Die API ist OpenAI-kompatibel. Das heißt: Alles, was heute gegen eine Cloud-API programmiert ist, lässt sich mit einer anderen Adresse und einem anderen Schlüssel auf das Mesh umstellen. Für die Werkzeuge, die ich für Betriebe baue – Dokumentenmanagement mit KI, automatische Angebotserstellung, der KI-Assistent auf eigenen Daten – ist das Mesh damit ein Rechenwerk, das ich anschließen kann, ohne sie umzubauen.

6 · Für Entwickler: KI im Editor, berechnet in Deutschland

GemeinsamKI Code bringt die Modelle des Verbunds nach VS Code und ins Terminal: Chat, Aufträge und Änderungen am Code – geplant, umgesetzt, geprüft. Die Clearingstelle läuft dabei auf dem eigenen Rechner, Zugangsdaten und Benutzerpfade verlassen ihn nicht. Ich baue meine KI-Werkzeuge inzwischen so. Nicht aus Prinzip, sondern weil es funktioniert.

7 · Ein eigener Kreis: der Cluster

Mit einem Cluster rechnen nur die Geräte deines Betriebs für deinen Betrieb. Die Werkstatt-Workstation, der Mac der Geschäftsführung und der Server im Keller bilden zusammen eine KI, die niemand außer deinen Leuten nutzt – verwaltet im Hub, ohne eigenen Server.

8 · Kontrolle, die man sehen kann

Play, Stop, Zeitplan, Speicherlimit – je Gerät, jederzeit. Der Live-Status zeigt, welche Geräte gerade rechnen. Die Vertrauensstufen stehen offen beschrieben, ebenso die Grenzen. Das ist für mich der Kern von Datenschutz by Design: nicht versprechen, sondern zeigen.

9 · Ein Ansprechpartner vor Ort

Ein Mesh ist nur so gut wie das Netzwerk darunter. Ich baue Netzwerke auf, richte PCs und Server ein, schließe sie als Knoten an, lege den Cluster an und verbinde die Werkzeuge, die der Betrieb braucht. Bei größeren Projekten gemeinsam mit Deniz Fritsch. Den Auftrag und den kurzen Draht hast du trotzdem nur zu einer Person. Genau das meine ich mit: Ich verbinde die echte Welt mit der KI.

KAPITEL 7

Was das Mesh (noch) nicht ist.

Ein Whitepaper, das nur Vorteile aufzählt, ist ein Prospekt. Deshalb hier die Liste, die du vor einer Entscheidung kennen solltest.

- **Beta.** Mitmachen heißt: Konto erstellen, für die Beta anmelden, Freischaltung abwarten. Einen Anspruch auf die Freischaltung gibt es nicht. Die erste Testphase dreht sich ums Coding.
- **Kostenlos – in der Beta.** Der Grundsatz lautet „Wer teilt, nutzt mit“. Wie es danach weitergeht, wird offen mit der Community besprochen. Plane nicht mit „für immer kostenlos“.
- **Kleinere Modelle.** Die größten Cloud-Modelle sind stärker als die offenen Modelle im Mesh. Für die meisten Betriebsaufgaben reicht es; für manche nicht.
- **Antwortzeit.** Sie hängt vom Knoten ab, der gerade frei ist. Nachts im eigenen Cluster ist das planbar, im allgemeinen Mesh schwankt es.
- **Pseudonymisieren ist nicht Anonymisieren.** Ein Name ohne Kontext kann durchrutschen, Bilder lassen sich nicht schützen. Gesundheits- und Mandantendaten gehören nicht in den Chat.
- **Strom.** Zahlt, wer teilt. Das ist fair, aber es ist nicht null.
- **Cluster sind kein Firmen-Hub.** Ein Cluster läuft im Hub von GemeinsamKI. Einen Gruppenchat zwischen Mitgliedern gibt es nicht; „im Cluster chatten“ heißt: mit der KI auf den Geräten des Clusters.
- **Windows und Linux** werden nach dem Mac-Start schrittweise angebunden. Anmelden geht jetzt, rechnen vielleicht erst in ein paar Wochen.

> MEIN URTEIL
>
> Für Texte, Dokumente, Auswertungen und Code im Betrieb ist das Mesh heute eine ernsthafte Alternative zum Abo pro Kopf – vor allem dort, wo Kundendaten im Spiel sind. Für Forschung an der Grenze des Machbaren ist es das nicht. Beides stimmt gleichzeitig.

KAPITEL 8

So fängst du an.

Weg 1: Selbst

1. Konto auf gemeinsamki.de erstellen. Damit bist du im Forum.
2. Im Konto für die Beta anmelden: mit welchem Gerät, wofür.
3. Freischaltung abwarten – kommt per E-Mail.
4. Agent laden, Assistent durchklicken, Zeitplan setzen. Dann chatten, in VS Code arbeiten oder die API anbinden.

Weg 2: Mit mir

1. **Bestandsaufnahme.** Welche Rechner stehen da, was können sie, wann sind sie frei? Der Geräte-Check liefert die erste Einschätzung.
2. **Netzwerk sauber machen.** Eigenes Segment für KI-Knoten, Firewall, VPN fürs Homeoffice, Backup nach der 3-2-1-Regel. Ohne Fundament kein Mesh.
3. **Knoten anschließen.** Agent auf den Geräten, Zeitplan nach Betriebszeiten, Speicherlimit so, dass tagsüber niemand etwas merkt.
4. **Cluster anlegen.** Mitglieder einladen, Geräte per Code aufnehmen, „nur Cluster“ einstellen.
5. **Werkzeuge anbinden.** Dokumentenmanagement, Angebots-KI oder Assistent mit eigenen Daten über die API ans Mesh hängen.
6. **Team schulen.** Ein KI-Tag im Haus bringt Regeln und Können ins Team und erfüllt die Schulungspflicht nach Art. 4 der EU-KI-Verordnung.
7. **Danach.** Wenn du willst, bleibe ich dein Ansprechpartner – als laufender IT-Support vor Ort und per Fernwartung.

Acht Fragen vor dem Start

- Welche Aufgaben sollen KI übernehmen – und welche ausdrücklich nicht?
- Welche Daten stecken in diesen Aufgaben: Kundennamen, Kontodaten, Gesundheitsdaten?
- Welche Rechner stehen frei – und wann?
- Reicht das allgemeine Mesh, oder braucht der Betrieb einen eigenen Cluster?
- Welche bestehenden Werkzeuge sollen ans Mesh – und sind sie OpenAI-kompatibel angebunden?
- Wer im Betrieb verantwortet Play, Stop und Zeitplan?
- Ist das Netzwerk dokumentiert: Netzplan, Geräteliste, Zugänge?
- Wer schult das Team, und wann?

KAPITEL 9

Wer dahintersteht.

GemeinsamKI ist ein Projekt der ResultOS UG (haftungsbeschränkt), Lohgasse 5, 95679 Waldershof. Geschäftsführer sind Deniz Fritsch und Matthias Eger. Projektstart 2026, Domains gemeinsamki.de und gemeinsamki.org. Die ResultOS UG ist ein privates Unternehmen, kein Verein; wer das Projekt unterstützen will, tut das als Partner, Investor oder mit Rechenleistung – nicht per Spende.

Deniz Fritsch	Ubiquiti Certified Partner, hat das Fach im Dienst bei der Bundeswehr gelernt und betreibt das eigene Rechenzentrum der ResultOS UG. Netzwerk, Hardware, Infrastruktur.
Matthias Eger	Fachinformatiker Systemintegration (IHK), Manager für angewandte KI-Transformation (IHK). Seit 2008 selbstständig, 314 Projekte für Handwerk, Gesundheitswesen, Handel, Kanzleien und Agenturen. Software, KI-Werkzeuge, Prozesse.
Igakuro	Lokale KI aus demselben Haus, die personenbezogene Daten in deutschen Texten erkennt und unkenntlich macht. Dieses Wissen fließt direkt in die Clearingstelle des Mesh.

Kontakt

GemeinsamKI: hallo@gemeinsamki.de · +49 9231 6682270 · gemeinsamki.de
Matthias Eger: mail@matthiaseger.de · 0961 – 634 32 61 · matthiaseger.de

> HINWEIS
>
> Alle Angaben zu GemeinsamKI geben den Stand vom 08.11.2026 nach gemeinsamki.de wieder. Das Projekt ist in der Beta; Funktionen, Modelle und Bedingungen ändern sich. Dieses Whitepaper ist keine Rechtsberatung. Ob ein Einsatz im eigenen Betrieb datenschutzrechtlich zulässig ist, klärt im Zweifel der Datenschutzbeauftragte – ich helfe gern bei der technischen Seite der Antwort.

ANHANG

Glossar.

Begriff	Erklärung
Agent	Die GemeinsamKI-App auf dem Rechner. Schließt ihn ans Mesh an, steuert Play, Stop und Zeitplan.
API (OpenAI-kompatibel)	Programmierschnittstelle, die sich wie die von OpenAI anspricht. Programme lassen sich ohne Umbau umstellen.
Clearingstelle	Regelwerk, das Personendaten durch Platzhalter ersetzt, bevor ein Knoten den Text sieht.
Cluster	Eigener kleiner Verbund im Mesh. Anfragen rechnen nur auf seinen Geräten.
Hub	Zentrale, die Anfragen an Knoten verteilt. Läuft auf Servern in Deutschland.
Knoten	Ein Rechner, der gerade Leistung teilt.
Mesh	Das Netz aus allen verbundenen Rechnern.
Modell	Die KI selbst, etwa gpt-oss-20b. „Offen“ heißt: frei verfügbar, lokal lauffähig.
Parameter	Grobes Maß für die Größe eines Modells. 20B = 20 Milliarden Parameter.
Quantisierung (4-Bit)	Verfahren, das ein Modell kleiner macht, damit es in den Speicher eines normalen Rechners passt.
Schutzcluster	Cluster nur mit vertrauten Geräten der Stufe A für Praxen und Kanzleien, Anbindung über Igakuro.
Tresor	Die getrennt gespeicherte Zuordnung von Platzhaltern zu echten Angaben.
Vertrauensstufe	A vertraut, B Verschwiegenheit, C offen. Regelt, welches Gerät welchen Text bekommt.
VRAM	Grafikspeicher. Bestimmt bei PCs, wie groß ein Modell sein darf.
Wächter	Zweite, unabhängige Prüfung des Texts vor dem Versand an einen Knoten.