

FÜR KUNDEN · PARTNER · ENTSCHEIDER

LEHRDOKUMENT · AGENTIC AI

Agentic AI, kontrolliert — wie aus Sprachmodellen belastbare Fachabläufe werden

Erst das Fundament: was Sprachmodelle, Agenten, Workflows und Orchestrierung wirklich sind. Dann die Architektur, die daraus ein Arbeitsmittel macht: Belege statt Behauptungen, Prüfung statt Vertrauen, menschliche Freigabe statt Automatik. Reifegrade werden in diesem Dokument offen benannt — Ehrlichkeit ist unser Arbeitsprinzip, nicht unsere Schwäche. Dies ist die Kompakt-Ausgabe — die ausführliche Fassung mit Technik-Teil und Referenzarchitektur senden wir gern auf Anfrage.

Neon Arc × ICO · Kompakt-Ausgabe · Stand 16. Juli 2026

1 · Was wir bauen — und was bewusst nicht

Die wichtigste Klarstellung zuerst: Wir verkaufen weder einen Chatbot noch bloße ‚LLM-Orchestrierung‘. Wir bauen kontrollierte Fachabläufe, in denen KI nur ein Baustein ist.

DER EINE EHRliche SATZ

Wir nehmen einen echten Fachablauf, bauen die verlässlichen Teile als normale Software und setzen ein Sprachmodell nur für Sprache, Einordnung und Entwürfe ein. Der Mensch behält die Entscheidung; Code, Rechte und Protokolle behalten die Kontrolle.

In 30 Sekunden: Die meisten kennen KI als Chat — Frage rein, Antwort raus. Wir gehen weiter: ein abgestimmtes Team spezialisierter Agenten mit einem Koordinator, einer unabhängigen Prüf-Instanz und einem nachprüfbar Ergebnis. Die Entscheidung bleibt immer beim Menschen, und jede Aussage ist belegt.

Vier Begriffe, die nie vermischt werden dürfen

LLM	Das Sprachmodell: ein sehr gutes Muster- und Textsystem. Es liest, formuliert, strukturiert und kann Vorschläge machen. Es ist nicht automatisch ein Agent und besitzt keine Rechte oder Hände.
AGENT	Ein Auftragsbearbeiter: ein Programm mit Ziel, Zustand, erlaubten Werkzeugen und Stopp-Regeln. Er kann ein LLM nutzen, kann aber auch reine, feste Logik sein.
WORKFLOW	Der fachliche Ablauf: Wer gibt was hinein? Welche Prüfung folgt? Was darf automatisch passieren? Was muss freigegeben werden? Der Workflow existiert auch ohne KI.
ORCHESTRIERUNG	Die Verkehrsleitung: Sie startet passende Schritte, gibt nur den nötigen Kontext weiter, wartet auf Ergebnisse, setzt Grenzen und protokolliert alles.

Was ein Kunde tatsächlich bekommt

- 01 Einen klaren Fachablauf.** Eingang, Rollen, Entscheidung, Ausnahme und Rückfallweg werden gemeinsam präzisiert.
- 02 Ein selbst gebautes Softwarepaket.** Oberfläche, Datenmodell, Rechte, Schnittstellen, Prüfregeln, Tests und Audit-Spur sind normaler, versionierter Code.
- 03 Einen gezielten KI-Baustein.** Das Sprachmodell liest unstrukturierte Sprache, extrahiert, klassifiziert oder formuliert einen Entwurf — keine unsichtbare Endentscheidung.

- 04 **Betrieb und Nachweis.** Monitoring, Freigaben, Fehlerpfade, Kostenlimits und eine messbare Pilotfrage machen daraus ein Arbeitsmittel statt einer Demo.

2 · Schwarmintelligenz — und die Linearitätsfalle

Mehr Agenten sind nicht automatisch besser — die Kunst ist die Koordination. Und die häufigste Falle hat mit Agenten gar nichts zu tun.

Die Linearitätsfalle: Warum der Chat allein nicht reicht

Wer eine KI **seriell** befragt — Frage, Antwort, nächste Frage — bekommt flüssige Antworten, aber jede hängt am Gesprächsverlauf davor: Die **Reihenfolge der Fragen färbt das Ergebnis**, und niemand prüft gegen. Das Ergebnis fühlt sich wie Gründlichkeit an, ist aber vom Zufall der Befragungsreihenfolge mitbestimmt. Der Chat ist dabei nicht das falsche Werkzeug — er ist das unstrukturierte: Für eine schnelle Auskunft reicht er, für eine Fach- oder Führungsentscheidung braucht es Struktur, Belege und Gegenprüfung.

AUF DEN PUNKT Gegen die Linearitätsfalle helfen drei Dinge: parallele Perspektiven statt einer Fragenkette, Beleg-Pflicht statt Eloquenz, und eine unabhängige Instanz, die widerlegen darf.

Können Agenten „miteinander reden wie ein Team“?

Technisch **können** mehrere Agenten kommunizieren (Multi-Agenten-Systeme). Aber freies Agent-zu-Agent-Geplauder ist oft **kontraproduktiv**: Der Kontext bläht auf, Fehler pflanzen sich fort, sie drehen sich im Kreis, die Kosten explodieren. Die robuste Lösung ist **strukturierte Koordination**: Ein Koordinator zerlegt, schickt Spezial-Agenten mit eigenem, sauberem Kontext los, jeder liefert ein Ergebnis zurück — koordiniert wird über eine gemeinsame Ablage, nicht über ein Dauergespräch. Ein unabhängiger Prüf-Agent kontrolliert, statt mitzuquatschen.

BILD *Kein Großraumbüro, in dem alle durcheinanderreden — sondern ein Projektleiter, der Aufgaben verteilt und Ergebnisse einsammelt.*

Wann freies „Reden“ doch sinnvoll ist

Für **Vielfalt und Debatte**: mehrere Perspektiven, die bewusst gegeneinander argumentieren — etwa Perspektiv-Agenten, die dasselbe aus verschiedenen Haltungen durchdenken, oder ein Bewertungs-Panel. Da ist der „Streit“ das Produkt, nicht der Störfaktor.

AUF DEN PUNKT Schwarmintelligenz richtig verstanden: nicht „alle reden gleichzeitig“, sondern „viele abgestimmte Perspektiven, geprüft zusammengeführt“. Mehr Agenten sind nicht automatisch besser — die Kunst ist die Koordination.

3 · Datenhoheit, EU-Anbieter & lokale Modelle

Die Frage kommt sicher: „Wo laufen die Daten, welches Modell, ist das DSGVO-sicher?“ — hier die ehrlichen Antworten.

Offenes vs. geschlossenes Modell — und was „lokal“ heißt

Geschlossene Modelle (Claude, GPT) laufen nur beim Anbieter über eine Schnittstelle — top Qualität, aber die Daten gehen dorthin. **Offene Modelle** (die „Gewichte“ sind einsehbar: z. B. Llama, Mistral, Qwen) kann man **selbst betreiben** — auf einem gemieteten oder eigenen GPU-Server. Das ist mit „lokal / self-hosted / on-premise“ gemeint: Das Modell läuft in Ihrer Kontrolle. Das erhöht die technische Kontrolle, ersetzt aber weder Zugriffsrechte, Verschlüsselung, Backups noch eine Datenschutzprüfung.

BILD Geschlossenes Modell = Taxi (bequem, aber ein Fremder fährt und weiß, wohin). Offenes Modell selbst betrieben = eigenes Auto (mehr Aufwand, aber du fährst und niemand sieht mit).

Souveränität in drei Stufen

- 01 **Stufe 1 — EU-Anbieter:** KI-Dienst mit vertraglich und technisch passender EU-Region. Schneller Start, aber Region, Aufbewahrung, Unterauftragsverarbeiter und Produktmodus werden konkret geprüft.
- 02 **Stufe 2 — dedizierter EU-Server:** ein reservierter Server im EU-Rechenzentrum, kein geteilter Dienst.
- 03 **Stufe 3 — self-hosted / on-premise:** offenes Modell im eigenen Haus, bis zum abgeschotteten Netz (air-gapped). Höchster Schutz, höchster Aufwand.

„Ist Cloud-Hosting DSGVO-sicher?“ — die ehrliche Antwort

Ein Cloud-Dienst kann **Teil einer DSGVO-konformen Architektur** sein, aber nicht „automatisch DSGVO-sicher“. EU-Regionen und EU-Datengrenzen gelten nur für passende, korrekt konfigurierte Dienste und sehen begrenzte Ausnahmen vor. Bei US-Anbietern ist die Rechtslage ein zusätzlicher Risikofaktor; das konkrete Risiko hängt von Dienst, Region, Vertrag, Datenart, Schlüsseln und Zugriffskonzept ab. Das ist keine Rechtsberatung — diese Punkte werden mit Datenschutz und IT des Kunden geprüft.

AUF DEN PUNKT Wir behandeln Datenstandort und Zugriff nicht als Marketing-Label: Dienst, Region, Vertrag, Datenfluss und Schlüsselkonzept werden für den konkreten Fall geprüft. Je höher der Schutzbedarf, desto mehr Kontrolle planen wir ein — bis self-hosted.

Was wir nutzen — offen gesagt

Entwickelt wird mit starken Denk-Modellen über moderne Entwicklungswerkzeuge. Für datenschutzkritische Anwendungen wird ein EU-Anbieter wie **Mistral** geprüft und passend konfiguriert. Wir **staffeln** die Modelle nach Aufgabe: starke Modelle für Orchestrierung und Kernlogik, mittlere fürs Prüfen, günstige und schnelle fürs mechanische Lesen und Ausführen. Der Anbieter ist per Konfiguration umschaltbar — kein Programmwechsel.

INFERENCE	Der laufende Betrieb des Modells (eine Anfrage beantworten) — im Gegensatz zum Training.
FINE-TUNING VS. PROMPTING	Fine-Tuning = das Modell nachtrainieren (teuer). Prompting = gutes Anweisen ohne Umbau (unser Weg).
EMBEDDING / VEKTOR-DATENBANK	Texte werden in Zahlen-Fingerabdrücke übersetzt, um Ähnliches zu finden — die Basis von Dokumenten-Suche/RAG.
AVV / SCC	Auftragsverarbeitungs-Vertrag / EU-Standardvertragsklauseln — wichtige Vertragsbausteine, aber kein Ersatz für eine passende technische Architektur.
CLOUD ACT	US-Gesetz, das bei US-Anbietern ein zusätzlicher Rechts- und Risikoaspekt sein kann. Konkrete Bewertung immer mit Datenschutz und Vertrag prüfen.

4 · Der EU AI Act — Checkliste statt Schreckgespenst

Regulierung ist kein Verkaufsargument und kein Schreckgespenst. Sie ist eine Checkliste dafür, wo Verantwortung, Transparenz und Dokumentation hingehören.

DIE KORREKTE HALTUNG

Nicht „wir sind automatisch compliant“, sondern: Wir klassifizieren jeden echten Einsatz, dokumentieren Rolle, Risiko, Daten und menschliche Kontrolle — und holen bei Grenzfällen rechtliche Prüfung dazu.

Die Zeitlinie in Klartext (Stand Juli 2026)

SEIT 02.02.2025	Verbot bestimmter KI-Praktiken und Pflicht zu angemessener KI-Kompetenz: Wer KI einsetzt, muss die beteiligten Menschen passend schulen.
-----------------	--

SEIT 02.08.2025	Governance und Pflichten für Anbieter allgemeiner KI-Modelle (GPAI). Wer ein Modell nur per Schnittstelle nutzt, ist nicht dadurch selbst GPAI-Modellanbieter.
SEIT 02.08.2026	Der AI Act gilt grundsätzlich. Transparenzpflichten können je Einsatz greifen; der genaue Umfang hängt von der Rolle und dem Produkt ab.
HOCHRISIKO	Nicht das Wort „KI“ entscheidet, sondern der konkrete Zweck und die Folgen. Personal, Bildung, Kredit, Leistungen, Biometrie oder kritische Infrastruktur brauchen besondere Prüfung.

Was für jeden Kundenfall schriftlich geklärt wird

- 01 **Rolle und Zweck.** Sind wir Anbieter des Systems, der Kunde Betreiber oder beides? Welche Entscheidung wird vorbereitet — und welche nicht?
- 02 **Daten und Folgen.** Welche personenbezogenen oder vertraulichen Daten gehen hinein? Hat das Ergebnis Einfluss auf Menschen, Geld, Zugang oder Sicherheit?
- 03 **Kontrolle und Transparenz.** Wo sieht der Mensch, dass KI beteiligt war? Wo kann er korrigieren, stoppen und eine Entscheidung begründen?
- 04 **Nachweis.** Modell, Prompt, Quellen, Werkzeuge, Tests, Fehlerfälle, Freigaben und Änderungen werden nachvollziehbar dokumentiert.
- 05 **Rechtliche Schicht.** Der AI Act ersetzt weder DSGVO noch Vertrag, Auftragsverarbeitung, Löschkonzept oder gegebenenfalls Datenschutz-Folgenabschätzung.

Warum Unternehmen trotzdem so stark in KI investieren

Weil in fast jedem Betrieb unstrukturierte Arbeit steckt: E-Mails lesen, Dokumente verstehen, Informationen suchen, Fälle vorsortieren, Texte vorbereiten. Ein gut begrenzter KI-Schritt spart Zeit und macht Wissen schneller verfügbar. Der Wert entsteht aber erst, wenn er in einen echten Prozess, verlässliche Daten und verantwortliche Freigaben eingebettet ist — nicht durch möglichst viele „Agenten“.

Was Selbst-Hosting nicht automatisch löst

Ein eigenes Modell kann Datenwege verkürzen, ersetzt aber weder Risiko-Einstufung noch Transparenz, Rollenrechte, menschliche Aufsicht, Tests oder Datenschutz. Es ist eine Hosting-Entscheidung — kein Compliance-Stempel.

AUF DEN PUNKT Unser Gegenmittel unabhängig von der Risikoklasse: Zweck begrenzen, Daten minimieren, Rollen und Freigaben erzwingen, KI-Ausgaben prüfen, Entscheidungen protokollieren, Tests gegen Fehler fahren und einen Rückfallweg ohne KI vorsehen. Bei Personal-, Kredit-, Zugangs- oder sicherheitsrelevanten Entscheidungen folgt vor Einsatz ein gesondertes Legal-/Risk-Review.

5 · Vom Konzept zum belastbaren System

Der Ablauf hinter einer guten Demo — wie aus einer Idee ein belastbares System wird, ohne so zu tun, als sei KI Magie.

UNSER ARBEITSPRINZIP

Wir starten nicht mit „Welches Modell nehmen wir?“, sondern mit einem konkreten Vorgang: Wer gibt was hinein, welche Entscheidung soll vorbereitet werden, was darf automatisch passieren — und was muss ein Mensch freigeben? KI ist dabei ein Baustein. Die Verlässlichkeit entsteht durch Regeln, Daten, Tests und Verantwortung.

Der Ablauf in fünf klaren Schritten

- 01 **Einen engen, echten Fall wählen.** Zum Beispiel: „Ein Lieferant lädt einen Nachweis hoch; der Prüfer soll nur die kritischen Punkte zuerst sehen.“ Nicht „wir automatisieren alles“.
- 02 **Eingang, Entscheidung und Ergebnis festlegen.** Welche Daten kommen hinein? Welche Regel oder Quelle gilt? Was sieht der Mensch am Ende? Das wird als Ablauf und Akzeptanzkriterium aufgeschrieben.
- 03 **Den sicheren Teil in Code bauen.** Rechte, Datenzugriff, Berechnungen, Fristen, Freigaben und Protokolle sind normale Softwarelogik. Sie muss vorhersehbar bleiben.
- 04 **KI nur für den unstrukturierten Teil einsetzen.** Lesen, Zusammenfassen, Einordnen oder einen Entwurf machen. Quellen und erlaubte Werkzeuge werden eng begrenzt.
- 05 **Beweisen, dann ausweiten.** Mit Testfällen, Gegenbeispielen, Review und einer kleinen Demo. Erst wenn der enge Fall nachvollziehbar funktioniert, kommt der nächste dazu.

BILD Wie beim Hausbau: Erst der Plan, dann das Fundament und die Leitungen; erst danach kommt die schöne Fassade. Das Sprachmodell ist ein sehr schneller Facharbeiter — nicht das Fundament.

Womit wir praktisch arbeiten

FACHLOGIK	Die Regeln des Geschäfts: Fristen, Rollen, Freigaben, Status und Sonderfälle. Sie bestimmen, was das System überhaupt tun darf.
ANWENDUNGSCODE	Die normale Software, die Oberflächen, Datenfluss, Zugriffe und Integrationen verbindet. Versioniert, überprüfbar und testbar.
KI / SPRACHMODELL	Liest und strukturiert Sprache, schlägt nächste Schritte oder Entwürfe vor. Sie erhält nur den nötigen Kontext und keine unbegrenzte Handlungsfreiheit.

QUELLEN & DATEN

Freigegebene Dokumente, Datenbankeinträge und fachliche Regeln. Bei Aussagen mit Risiko zählen Fundstelle und Datenqualität mehr als eloquente Formulierungen.

KONTROLLEN

Rollenrechte, Maskierung, Tests, Gegenprüfungen, Freigabe und Protokolle. Sie machen aus einer KI-Demo ein kontrollierbares Arbeitsmittel.

Reifegrad: das Wort, das Vertrauen schafft

Wir benennen für jedes System offen, wo es steht — mit einem festen Vokabular. Das ist kein Kleingedrucktes, sondern unser Qualitätsversprechen: Sie wissen immer, was ein Nachweis belegt und was nicht.

KONZEPT

Das Problem und der Ablauf sind beschrieben. Noch kein Beweis, dass es technisch oder wirtschaftlich funktioniert.

TECHNISCH GEBAUT

Der Ablauf läuft im Code und auf Beispieldaten. Belegt technische Machbarkeit — nicht automatisch Betriebssicherheit.

IN HÄRTUNG

Fehlerfälle, Rechte, Sicherheit, Bedienung und Wiederholbarkeit werden systematisch geprüft.

KONTROLLIERTER PILOT

Ein klar abgegrenzter echter Fall mit benannten Nutzern, Messgröße und Rückfallebene. Erst hier entsteht Nutznachweis.

BETRIEB

Wiederholbar, mit Verantwortlichen, Support, Monitoring und echten Nutzungsdaten. Nicht mit „fertig programmiert“ verwechseln.

AUF DEN PUNKT Wir bauen keine KI-Show. Wir nehmen einen konkreten Arbeitsablauf, sichern die verlässlichen Teile in Code ab und setzen KI dort ein, wo sie wirklich hilft. Dann belegen wir den Reifegrad offen.

Und ehrlich bleibt ehrlich: Ein grüner Test beweist nicht, dass Kunden zahlen. Eine schöne Demo beweist nicht, dass ein System sicher oder produktionsreif ist. Beides sind unterschiedliche Nachweise — wir verwechseln sie nicht.

Belegtes Wissen und generierter Vorschlag — der wichtige Unterschied

Die Belegpflicht sichert Aussagen **über den Fall**: Jede Zuschreibung braucht ein wörtliches Zitat, sonst wird sie blockiert. Ein **Vorschlag** dagegen — ein Aktionsplan, eine Frist, eine Zielzahl — beschreibt etwas Zukünftiges und ist deshalb *nicht* belegbar: Es gibt nichts zu zitieren, was noch nicht existiert. Ein Sprachmodell formuliert solche Zahlen plausibel, ohne den Fachkontext zu kennen (realistische Laufzeiten, Aufsichtsfristen, Benchmarks). Das ist kein Fehler der Kontrollschicht, sondern die Grenze jeder generierenden KI — und wir benennen sie offen, statt sie zu kaschieren.

Im echten Einsatz wird aus „geraten“ „hergeleitet und geprüft“ — auf drei Ebenen:

- 01 **Der Mensch entscheidet.** Jeder Vorschlagswert wird übernommen, geändert oder verworfen — protokolliert und namentlich zugerechnet. Bereits eingebaut.
- 02 **Fach-Korpus als zweite Quelle.** Der Generator prüft Werte gegen echte Projektdaten, Benchmarks und regulatorische Fristen des Kunden — dann sind sie hergeleitet, nicht geschätzt.
- 03 **Deterministische Guardrails.** Harte Regeln im Code, die kein Modell überreden kann: keine Kennzahl ohne hinterlegte Berechnungsbasis, Fristen nur gegen den echten Projektkalender, definierte Wertebereiche.

Stand der zweiten Ebene: Die erste Ausbaustufe des Fach-Korpus ist eingebaut. Hinterlegt ist destilliertes Methodenwissen mit wörtlichen Quellenbelegen (Fundstelle je Aussage); trägt ein Vorschlag im Aktionsplan eine Methodik-Referenz, ist er aus diesem Korpus hergeleitet — alles ohne Referenz bleibt sichtbar Vorschlag. Zugleich wächst das System zur Plattform: Ein Einordnen-Schritt wählt vorab das Terrain (zwei Fragen, regelbasierter Vorschlag, der Mensch bestätigt), und ein Fall kann künftig mehrere Werkzeuge tragen — die Kraftfeldanalyse liest die Menschen, die EmpowerMap (angekündigt) strukturiert das Thema. Im Pilot wächst der Korpus um die Fachdaten des Kunden — dann werden auch Zahlen und Fristen hergeleitet.

AUF DEN PUNKT Was wir nicht versprechen, ist eine Garantie ohne diese Schichten — sie entstehen erst im Pilot mit den Daten des Kunden. Was wir zusichern: Eine generierte Zahl wird nie unbemerkt zur Entscheidung. Der Vorschlag ist als Vorschlag markiert, die Prüfung liegt beim Menschen und, im Ausbau, bei belegbaren Quellen und festen Regeln.

6 · Die vier Sorgen — und die Architektur-Antworten

Die vier Fragen, die in jedem Unternehmen zuerst kommen — und die Architektur-Antworten darauf.

F **Halluzination — „Erfindet die KI nicht Dinge?“**

A Das Modell kann erfinden — genau deshalb: Zitat-Zwang gegen einen echten Korpus, Unbelegtes wird blockiert und sichtbar als „offen“ markiert, und eine unabhängige zweite Instanz falsifiziert. Erzeuger ist nicht Prüfer.

F **Kontrollverlust — „Macht die KI dann alles allein?“**

A Nein. Der Agent entscheidet nichts, er legt einen belegten Entwurf vor. Verändernde Aktionen brauchen menschliche Freigabe. Dazu mechanische Tore und Kostengrenzen.

F Datenabfluss — „Wo landen unsere Daten?“

- A Personenbezogene Daten werden vor der KI maskiert, europäische Anbieter sind konfigurierbar, Außen-Abrufe sind gehärtet. Ein Pilot läuft in einer Sandbox mit anonymisierten Daten; Datenklasse, Vertrag und Region werden pro Fall schriftlich geklärt.

F Prompt-Injection — „Kann man den Agenten austricksen?“

- A Nicht vertrauenswürdige Inhalte werden isoliert verarbeitet und als Daten behandelt, nie als Befehle — mehrschichtig im Live-Pfad. Versteckte Anweisungen werden zitiert und analysiert, aber nicht ausgeführt.

Fünf kluge Fragen, die Sie jedem Anbieter stellen sollten

- 01 Wo würden unsere Daten und das KI-Modell verarbeitet — in der EU, oder im eigenen Haus?
- 02 Welche Aktionen sollen automatisch laufen, welche brauchen zwingend eine menschliche Freigabe?
- 03 Wie wird protokolliert, wer wann welche Aktion ausgelöst hat — und wie lange aufbewahrt?
- 04 Wie ist der Agent gegen manipulierte Eingaben geschützt?
- 05 Mit welchem einzelnen, klar umrissenen Fall starten wir — hoher Nutzen, geringes Risiko?

Wer auf diese fünf Fragen keine konkreten, belegbaren Antworten bekommt, sollte kein Projekt starten — bei keinem Anbieter, auch nicht bei uns.

Worte, die Sie von uns nicht hören werden

Ehrlichkeit ist bei uns keine Notlösung, sondern die Kernbotschaft. Deshalb verzichten wir bewusst auf Formulierungen, die technisch nicht haltbar sind:

- „Unsere KI halluziniert nicht.“ — Kein Sprachmodell kann das garantieren. Haltbar ist: Unbelegtes wird ausgewiesen oder blockiert statt erfunden ausgegeben.
- „Das System ist deterministisch.“ — Die Code-Teile sind es (gleiche Eingabe, gleiche Ausgabe); der KI-Schritt bleibt probabilistisch und wird kontrolliert.
- „Automatisch DSGVO-konform / zertifiziert sicher.“ — Konformität entsteht pro Einsatz aus Architektur, Vertrag und Prüfung, nicht aus einem Label.
- „Unhackbar“, „military grade“, Erfolgszahlen ohne Messung. — Wenn dafür kein konkreter, aktueller Nachweis vorliegt, sagen wir es nicht.

AUF DEN PUNKT Wer selbst sagt, was das System NICHT kann, dem können Sie glauben, was es kann.

7 · Glossar

Jeder Begriff in einem Satz — zum schnellen Nachschlagen.

AGENTIC AI	KI, die nicht nur antwortet, sondern handelt: Ein Team von Agenten zerlegt, arbeitet parallel, führt geprüft zusammen.
KI-AGENT	Programm, das eine Aufgabe eigenständig in Schritten erledigt und Software bedient — unter Aufsicht.
ORCHESTRATOR	Der koordinierende Agent: zerlegt, verteilt, führt geprüft zusammen.
LLM / SPRACHMODELL	Das „Gehirn“, das Sprache versteht und erzeugt (z. B. Claude, Mistral).
TOKEN / KONTEXTFENSTER	Sprach-Häppchen / Kurzzeitgedächtnis des Modells (wie viel es auf einmal sieht).
PROMPT	Die Anweisung an das Modell.
FUNCTION-CALLING / WERKZEUG	Das Modell darf abgegrenzte Funktionen aufrufen (Termin anlegen, suchen) — ausgeführt vom Programm.
API	Eine kontrollierte digitale Serviceschalter-Schnittstelle: Software bittet andere Software um genau definierte Daten oder Aktionen.
OAuth	Ein zeitlich und fachlich begrenzter Besucherausweis für eine externe Anwendung — nicht das Passwort des Mitarbeiters weitergeben.
WEBHOOK	Eine digitale Türklingel: Ein Fremdsystem meldet sofort, dass etwas passiert ist, statt dass wir ständig nachfragen.
SFTP	Ein verschlossener digitaler Briefkasten für Dateiaustausch zwischen Systemen.
OBJEKT-SPEICHER / STORAGE	Das digitale Archiv für Originaldateien. Die Datenbank hält dazu nur den Index und die Rechte.
SIGNED URL	Ein kurz gültiger Besucherpasse zu genau einer privaten Datei — keine öffentliche Daueradresse.
RAG	Antwort erst in freigegebenen Quellen nachschlagen und mit Fundstelle belegen.
HALLUZINATION	Das Modell erfindet etwas plausibel klingendes — Gegenmittel: Zitat-Zwang und Blockade.

ERZEUGER ≠ PRÜFER	Wer ein Ergebnis erzeugt, gibt es nicht selbst frei; eine unabhängige Instanz falsifiziert.
PII	Personenbezogene Daten (Name, E-Mail, IBAN, Gesundheit).
MASKIERUNG / PSEUDONYM / ANONYM	Unkenntlich / durch Kürzel ersetzbar / nicht mehr rückführbar.
SSRF	Angriff: den Server austricksen, interne/verbotene Adressen abzurufen. Schutz: geprüfte Ausgangs-Abrufe.
PROMPT-INJECTION	Versteckte Befehle in Texten, die den Agenten kapern sollen. Schutz: Inhalte als Daten behandeln, isoliert verarbeiten.
RBAC	Rollenbasierte Rechte: jedes Werkzeug an eine Rolle gebunden.
RLS	Zeilenbasierte Rechte in der Datenbank: Selbst eine korrekte Abfrage sieht nur Daten des erlaubten Arbeitsbereichs.
SECRET / ENV	Ein geheimer Schlüssel in der Server-Umgebung. Er darf nie im Browser, Screenshot oder Quellcode landen.
HUMAN-IN-THE-LOOP	Der Mensch bestätigt/korrigiert — die Maschine handelt nicht allein.
MULTI-TENANT / ISOLATION	Ein sauberes Mandantenmodell verhindert, dass Mandant A Daten von Mandant B sieht.
AUDIT-TRAIL	Nachvollziehbares Protokoll: wer, wann, was. Revisionssicherheit braucht zusätzlich konkrete Unveränderbarkeits- und Aufbewahrungsnachweise.
DSGVO / EU AI ACT	EU-Datenschutz / EU-KI-Gesetz (Aufsicht, Protokolle, Doku für wichtige Systeme).
ON-PREMISE / SELF-HOSTED	Betrieb im eigenen Haus — die Daten verlassen es nicht.
OFFENE GEWICHTE (OPEN WEIGHTS)	Modell, dessen Innenleben einsehbar ist und das man selbst betreiben kann (Llama, Mistral, Qwen).
SOUVERÄNITÄT	Kontrolle darüber, wo Daten liegen und wer rechtlich Zugriff hat.
AVV / SCC / CLOUD ACT	Datenschutz-Verträge / EU-Standardklauseln / US-Zugriffsgesetz.
GUARDRAILS	Eingebaute Schranken, die verhindern, dass der Agent Unerlaubtes tut.
TEMPERATUR	Regler für Zufall/Kreativität des Modells — für Prüf-Aufgaben bewusst niedrig.

REIFEGRAD

Unser fünfstufiges Vokabular von Konzept bis Betrieb — damit ein Nachweis immer sagt, was er belegt.

DER MERKSATZ ZUM SCHLUSS

Nicht die klügste KI gewinnt, sondern die, der die Verantwortlichen vertrauen können. Der Schwarm liefert Perspektiven — der Mensch entscheidet — alles bleibt nachweisbar.

Stand 16.07.2026. Technische Aussagen sind aus dem aktuellen internen Review abgeleitet. Datenschutz-, Vertrags- und Compliance-Aussagen gelten nie pauschal: konkrete Architektur, Anbieter-Produkt, Vertrag und Einsatz werden je Kundenfall geprüft. Dieses Dokument ist ein Lehrdokument — keine Rechtsberatung und kein Angebot. Kontakt: Neon Arc · info@neonarc.com