AEGIS

Agent Engine - Governed Intent System

GOVERNANCE FÜR KI-AGENTEN

Autonome KI-Agenten. Ohne Kontrollverlust.

AEGIS trennt das Denken, das Prüfen und das Ausführen. Kein Agent handelt, ohne dass eine unabhängige Instanz zugestimmt hat - lokal betrieben, lückenlos nachvollziehbar, in Ihrer Hand.

Pilotprogramm 2027

Begrenzte Pilotplätze für ausgewählte Unternehmen

aegis-stack.com

DAS PROBLEM

KI-Agenten handeln längst selbstständig

Sie lesen Dateien, schreiben E-Mails, rufen Systeme auf. Die meisten Lösungen am Markt vertrauen dabei darauf, dass das Modell schon das Richtige tut. Doch ein Sprachmodell unterscheidet nicht zuverlässig zwischen Ihrer Anweisung und einem Text, den es gerade verarbeitet.

Mehr Rechte als nötig

01

Agenten erhalten oft pauschale Rechte auf Dateiablagen, Postfächer und Fachsysteme - mehr, als die Aufgabe erfordert. Jede überzählige Berechtigung ist eine Angriffsfläche, die ein manipulierter Agent nutzen kann.

Handeln ohne Freigabe

02

Was ein Agent plant, setzt er häufig im selben Schritt um. Eine eingeschleuste Anweisung in einer E-Mail oder einem PDF genügt, und aus einer Zusammenfassung wird ein Versand nach außen.

Keine Nachvollziehbarkeit

03

Wer hat was wann veranlasst, auf welcher Grundlage? Ohne belastbares Protokoll fehlt die Grundlage für Fehleranalyse, Haftungsfragen und Nachweise gegenüber Aufsicht und Kunden.

Die Frage ist nicht, ob ein Modell Fehler macht.
Die Frage ist, was es dann tun darf.

DAS PRINZIP

Kein Agent prüft sich selbst

AEGIS arbeitet nach dem Drei-Instanzen-Prinzip: drei technisch getrennte Instanzen, wie die drei Schichten unserer Bildmarke. Jede hat genau eine Aufgabe - und keine kann die andere übergehen.

1 PLANUNGSINSTANZ



Denken

Ein Sprachmodell versteht die Aufgabe und zerlegt sie in Schritte. Es liefert für jeden Schritt einen strukturierten Vorschlag, einen Intent, hat aber weder Zugangsdaten noch Zugriff auf Werkzeuge.

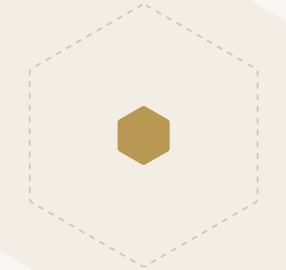
2 PRÜFINSTANZ



Prüfen

Deterministischer Code ohne Sprachmodell prüft jeden Intent gegen Ihre Regeln und signiert kryptografisch nur, was alle Prüfungen besteht. Diese Instanz plant nichts und führt nichts aus.

3 AUSFÜHRUNGSINSTANZ



Ausführen

Prüft die Signatur erneut, führt den freigegebenen Schritt aus und schreibt Ergebnis oder Ablehnung in ein Protokoll, das sich nachträglich nicht unbemerkt verändern lässt.

Ohne gültige Signatur passiert nichts. Selbst ein vollständig manipuliertes Modell kann nur Vorschläge erzeugen.

KONTROLLE IM DETAIL

Was die Prüfinstanz kontrolliert

Die Prüfinstanz bewertet keine Absichten, sondern nachweisbare Eigenschaften jedes Intents: Format, Werkzeug, Stufe, Ziel und Gültigkeit. Dieselbe Eingabe führt immer zum selben Ergebnis.

Schema

Der Intent muss exakt dem erwarteten Format entsprechen. Unbekannte Felder oder fehlende Angaben führen zur Ablehnung.

Werkzeug

Die Aktion muss auf der Positivliste stehen. Es gilt Default-Deny: Was nicht ausdrücklich erlaubt ist, ist verboten.

Aktionsstufe

Sie legen fest, welche Stufe ein Agent höchstens erreichen darf. Liegt der Intent darüber, wird er abgelehnt.

Ziele

Inhalte aus dem Internet nur von freigegebenen Domains. Das Internet gilt grundsätzlich als nicht vertrauenswürdig.

Gültigkeit

Abgelaufene Zeitstempel und bereits verwendete Freigaben werden verworfen. Eine erteilte Freigabe gilt nur einmal.

AKTIONSSTUFEN

Je mehr Wirkung, desto höher die Stufe

Jede Aktion gehört zu einer Stufe. Die Stufe ist Teil des signierten Intents und lässt sich nicht nachträglich herabsetzen.

0

Stufe 0: Nur lesen

Dokument öffnen, Wissensbasis durchsuchen, Ordner auflisten

1

Stufe 1: Lokale Zustandsänderung

Notiz anlegen, Entwurf speichern, Datei in einen Arbeitsordner schreiben

2

Stufe 2: Wirkung nach außen

E-Mail versenden, Daten an ein Fremdsystem übertragen, Termin für Dritte buchen

NACHVOLLZIEHBARKEIT

Jede Entscheidung belegbar

Jeder Schritt landet im Audit-Protokoll, auch jede Ablehnung - mit Begründung: welches Werkzeug, welche Stufe, welche Regel.

audit.log — Beispiel

```
09:14:02  planen      intent a41c ordner_lesen /akten/2026 stufe 0   vorgeschlagen
09:14:02  prüfen      a41c schema ok werkzeug erlaubt signiert        freigegeben
09:14:03  ausführen   a41c signatur gültig hash 9e2f..c1             protokolliert
09:14:07  planen      intent b7d0 shell rm -rf /tmp/export           vorgeschlagen
09:14:07  prüfen      b7d0 werkzeug nicht in positivliste             abgelehnt
09:14:09  ausführen   a41c nonce bereits verwendet                   abgelehnt
```



Das Protokoll ist nur anhängend. Jeder Eintrag enthält den SHA-256-Hash seines Vorgängers, so entsteht eine Hash-Kette. Wird ein Eintrag nachträglich verändert oder entfernt, passt die Kette nicht mehr, und die Ausführungsinstanz verweigert beim Start den Betrieb.

- **Wer hat die Aufgabe ausgelöst?**
- **Was hat der Agent geplant?**
- **Welche Regel hat entschieden?**
- **Was wurde tatsächlich ausgeführt?**

IM VERGLEICH

Drei Wege, KI-Agenten zu betreiben

	EINFACHER AGENT	MIT MENSCHLICHER FREIGABE	AEGIS
Wer entscheidet über die Ausführung?	Das Modell selbst	Ein Mensch, meist ohne Kontext	Feste Regeln in einer getrennten Instanz
Wirkung von Prompt Injection	Direkt bis zur Aktion	Hängt von der Aufmerksamkeit ab	Auf zulässige Werkzeuge und Stufen begrenzt
Skaliert bei vielen Aktionen?	Ja	Nein, Freigaben werden zur Routine	Ja, Regeln prüfen jede Aktion gleich
Nachweis im Nachhinein	Lückenhafte Logs	Klicks ohne Begründung	Signierte Intents in einer Hash-Kette
Zugangsdaten beim Modell?	Ja	Ja	Nein

DATENHOHEIT

Ihre Daten bleiben bei Ihnen

AEGIS läuft auf Ihrer eigenen Infrastruktur, vom Server im Haus bis zur Umgebung mit NAS-Speicher. Sie entscheiden, welche Modelle zum Einsatz kommen und ob das System Ziele außerhalb Ihres Netzes erreichen darf.

Offline-fähig

Der Kern arbeitet ohne Internetverbindung, Modelle laufen lokal.

Kein Training mit Ihren Daten

Dokumente, Protokolle und Wissensbasis verlassen Ihre Infrastruktur nicht.

Freie Wahl der Modelle

Modell- und anbieterunabhängig, ohne Vendor-Lock-in.

Speicherort nach Wahl

Protokolle und Wissensbasis liegen auf Ihrem Server oder NAS.

EINSATZ

Für wen AEGIS gedacht ist

Für Organisationen, die KI-Agenten produktiv nutzen wollen, deren Daten aber nicht in fremde Hände gehören und deren Entscheidungen nachweisbar sein müssen.

Kanzleien und Steuerberatung

Mandantendaten unterliegen der beruflichen Verschwiegenheit. Agenten dürfen recherchieren und Entwürfe vorbereiten, aber nichts ohne Prüfung nach außen geben.

Gesundheitswesen

Gesundheitsdaten gehören zu den besonders geschützten Kategorien der DSGVO. Lokaler Betrieb und strikte Aktionsstufen halten sensible Daten im Haus.

KMU mit Nachweispflichten

Wer Kunden, Prüfern oder Aufsicht Nachweise schuldet, braucht ein Protokoll, das zeigt, welcher Agent was mit welcher Freigabe getan hat.

Teams, die unabhängig bleiben wollen

AEGIS ist an keinen Modellanbieter gebunden. Sie wählen die Modelle und tauschen sie, ohne die Kontrolle neu aufzubauen.

AI Act, DSGVO und NIS2

AEGIS begrenzt Rechte technisch, lässt jede Aktion von einer unabhängigen Instanz prüfen und protokolliert sie manipulationssicher. Das unterstützt Ihre Pflichten aus AI Act, DSGVO und NIS2, ersetzt aber nicht deren rechtliche Bewertung. AEGIS stellt keine Konformität her, es liefert die technischen Bausteine, auf die sich Ihre Konformitätsarbeit stützen kann.

PILOTPROGRAMM 2027

Pilotplatz anfragen

Das Pilotprogramm ist keine Testversion, sondern eine begleitete Einführung: Wir setzen AEGIS gemeinsam mit Ihnen in einem klar umrissenen Prozess auf Ihrer Hardware ein, mit festen Ansprechpersonen aus unserem Entwicklungsteam. Pilotunternehmen nutzen AEGIS vor dem allgemeinen Marktstart 2027 und bestimmen mit, welche Werkzeuge und Anbindungen zuerst entstehen.

01 Anfrage
Sie schildern Ihr Vorhaben kurz über das Formular. Wir melden uns persönlich.

02 Erstgespräch
Prozesse, Datenlage, Infrastruktur und Ziele. Danach wissen beide Seiten, ob ein Pilot sinnvoll ist.

03 Konzept
Welcher Agent welche Werkzeuge nutzen darf, welche Aktionsstufen gelten und woran wir den Erfolg messen.

04 Pilotbetrieb
AEGIS läuft in Ihrer Umgebung. Wir werten die Protokolle gemeinsam aus und schärfen die Regeln nach.

Unverbindlich anfragen

Erzählen Sie uns kurz, wofür Sie KI-Agenten einsetzen möchten. Die Zahl der Pilotplätze ist begrenzt.

aegis-stack.com/pilotprogramm

AEGIS
Agent Engine - Governed Intent System

AEGIS wird entwickelt von oQueo, Agentur für KI, Software und Digital.
oQueo e.U., Rehgasse 5, 2514 Möllersdorf, Österreich
oqueo@oqueo.com · aegis-stack.com · © 2026 oQueo e.U. · Stand Oktober 2026