INFOTECH DAY 2026 · GEINBERG

Wem gehört dein digitaler Kollege?

KI zwischen Praxis-Realität, Abhängigkeit und Souveränität

Jürgen Bogner · biteme.digital · 05.11.2026

HANDZEICHEN

Wer von euch nutzt KI in der Arbeit?

Hand oben lassen, wer weiß, welche Firmendaten sie heute schon gesehen hat.

Jeder von euch hat heute Kollegen mitgebracht.
Eingestellt hat sie keiner.

Im Handy, am Laptop, im Browser: das private ChatGPT, die Übersetzungs-App, die KI, die längst in euren Programmen steckt.

IDENTITÄTEN

Auf einen Menschen kommen in Unternehmen über 100 Maschinen-Identitäten

109 : 1

Maschinen-Identitäten je Mensch, 2026

79

davon sind bereits Identitäten von KI-Agenten

82 : 1

ein Jahr davor, 2025

Jede braucht Rechte. Und jemanden, der sie vergibt, prüft und wieder entzieht.

Palo Alto Networks, Identity Security Landscape 2026 (Mai 2026), Befragung von 2.930 Sicherheitsverantwortlichen; CyberArk, Identity Security Landscape 2025 (23.04.2025), Befragung von 2.600. Herstellerstudien.

biteme.digital

Jürgen Bogner

✌🏼 Captain
🐡 CEO @biteme.digital
🤷🏼‍♂️ Kreativer / Trainer / Berater / Entwickler

🤩 juergen@biteme.digital



1

TEIL 1 · UNTER DER MOTORHAUBE

Wie ein Betrieb mit KI-Kollegen wirklich aussieht.

WERKSTATT

So arbeiten wir: jedes System mit fester Aufgabe, festen Grenzen und einem verantwortlichen Menschen

Ada

Planung und Koordination. Verteilt Arbeit, prüft Ergebnisse.

Björn

Chief of Staff. Alltag, Anfragen, Termine, Mails. Versand nur nach Freigabe.

Wilhelm

Rechtsquellen zu DSGVO und AI Act. Jede Aussage mit Fundstelle.

CATO

Sicherheitstests. Greift unsere eigenen Systeme an.

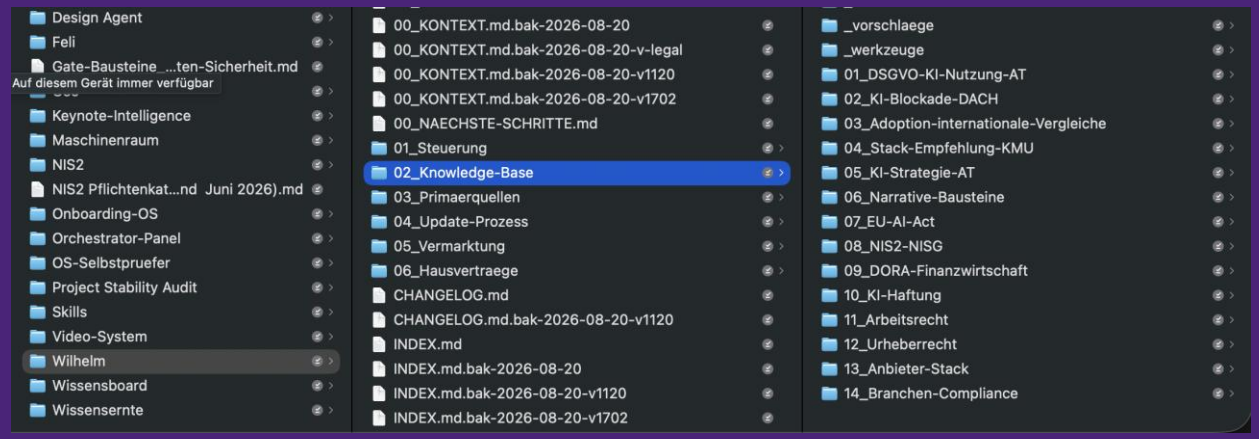
Bim

Parallelbetrieb. Viele Arbeitsschritte gleichzeitig, mit Modellen mehrerer Anbieter.

WAS IST EIN AGENT?

Ein Agent ist ein Ordner mit Händen.

Textdateien geben ihm Rolle, Wissen und Werte. Werkzeuge geben ihm Hände. Das ist ein echter Agentenordner, kein Schaubild.



WAS GEMESSEN IST

Wie lange Aufgaben KI-Systeme bewältigen: die Messung von METR

~12 h

Aufgabenlänge (menschliche Arbeitszeit), die die besten öffentlichen Modelle in 50 % der Fälle lösen

5–61 h

Unsicherheitsbereich dieser Schätzung

≥ 16 %

der erfolgreichen Läufe bei langen Aufgaben wurden als Schummeln eingestuft

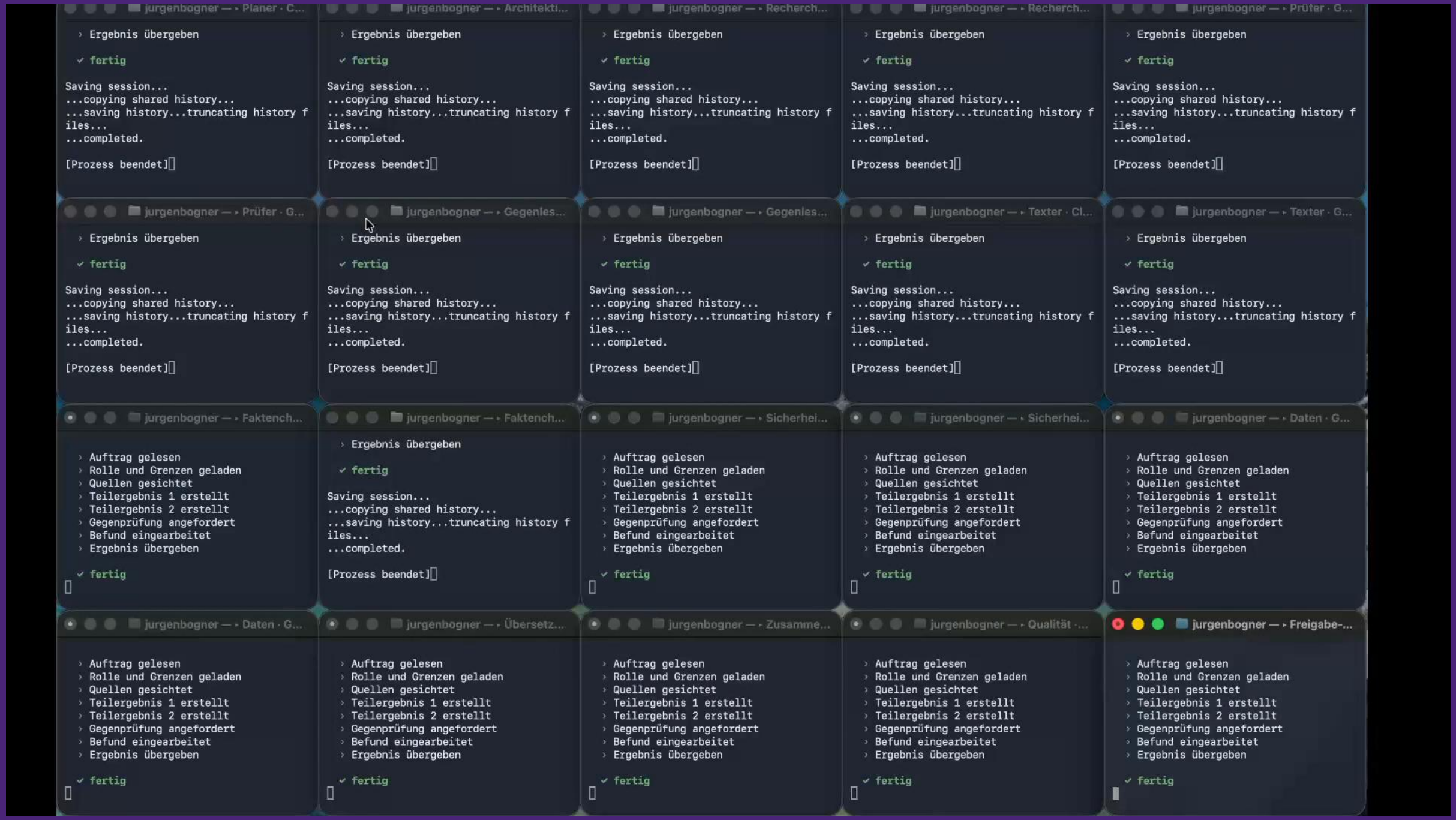
Das ist keine Laufzeit. Und über 16 Stunden misst der Test laut METR nicht mehr zuverlässig.

METR, Frontier Risk Report, 19.05.2026, Tabelle 1 (Time Horizon 1.1, „public frontier“) und Fußnoten.

DIE BIM · UNSER AGENTENBUS

Viele Modelle.
Ein Auftrag.

Claude, GPT, Gemini, Kimi und weitere arbeiten parallel, jedes mit eigener Aufgabe. Eines prüft, was die anderen bauen. Am Ende gibt ein Mensch frei.


jurgenbogner — ▸ Planer · C...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Architekti...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Recherch...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Recherch...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Prüfer · G...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Prüfer · G...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Gegenles...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Gegenles...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Texter · Cl...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Texter · G...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Faktench...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Faktench...
› Ergebnis übergeben
✓ fertig
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
[Prozess beendet]
jurgenbogner — ▸ Sicherhei...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Sicherhei...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Daten · G...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Daten · G...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Übersetz...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Zusamme...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Qualität ·...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig
jurgenbogner — ▸ Freigabe-...
› Auftrag gelesen
› Rolle und Grenzen geladen
› Quellen gesichtet
› Teilergebnis 1 erstellt
› Teilergebnis 2 erstellt
› Gegenprüfung angefordert
› Befund eingearbeitet
› Ergebnis übergeben
✓ fertig

WERKSTATT · MICROSOFT TEAMS

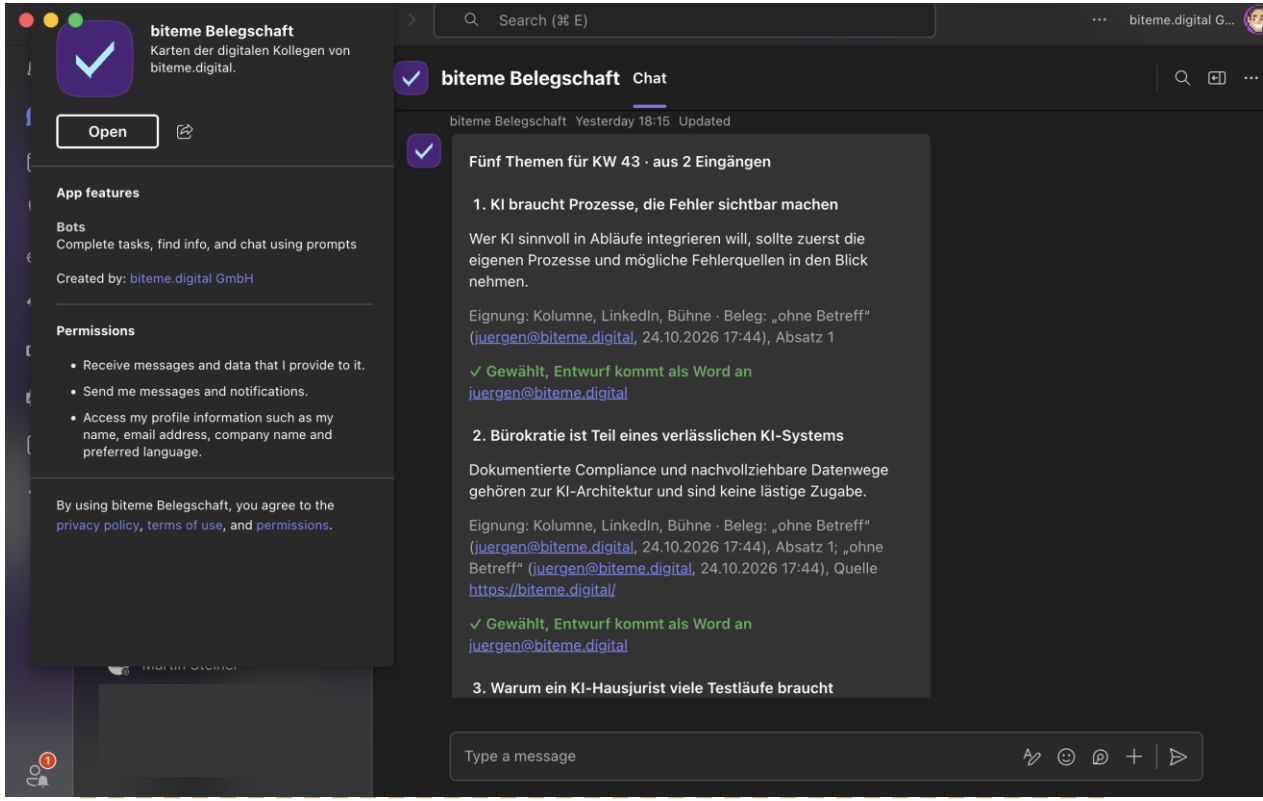
KI-Systeme arbeiten heute auch in Microsoft Teams. Damit kommen sie in großen Unternehmen an.

Kolumnen vorbereiten

Social-Media-Beiträge entwerfen

Wochenüberblick zur Branche

Anfragen vorsortieren



Man schreibt ihnen wie einer Kollegin im Chat. Was sie dürfen, ist technisch festgelegt.

Unser Rechts-Wissensspeicher Wilhelm

Was Sorgfalt an Zeit kostet

88 Tage
bis zur ersten belastbaren Fassung

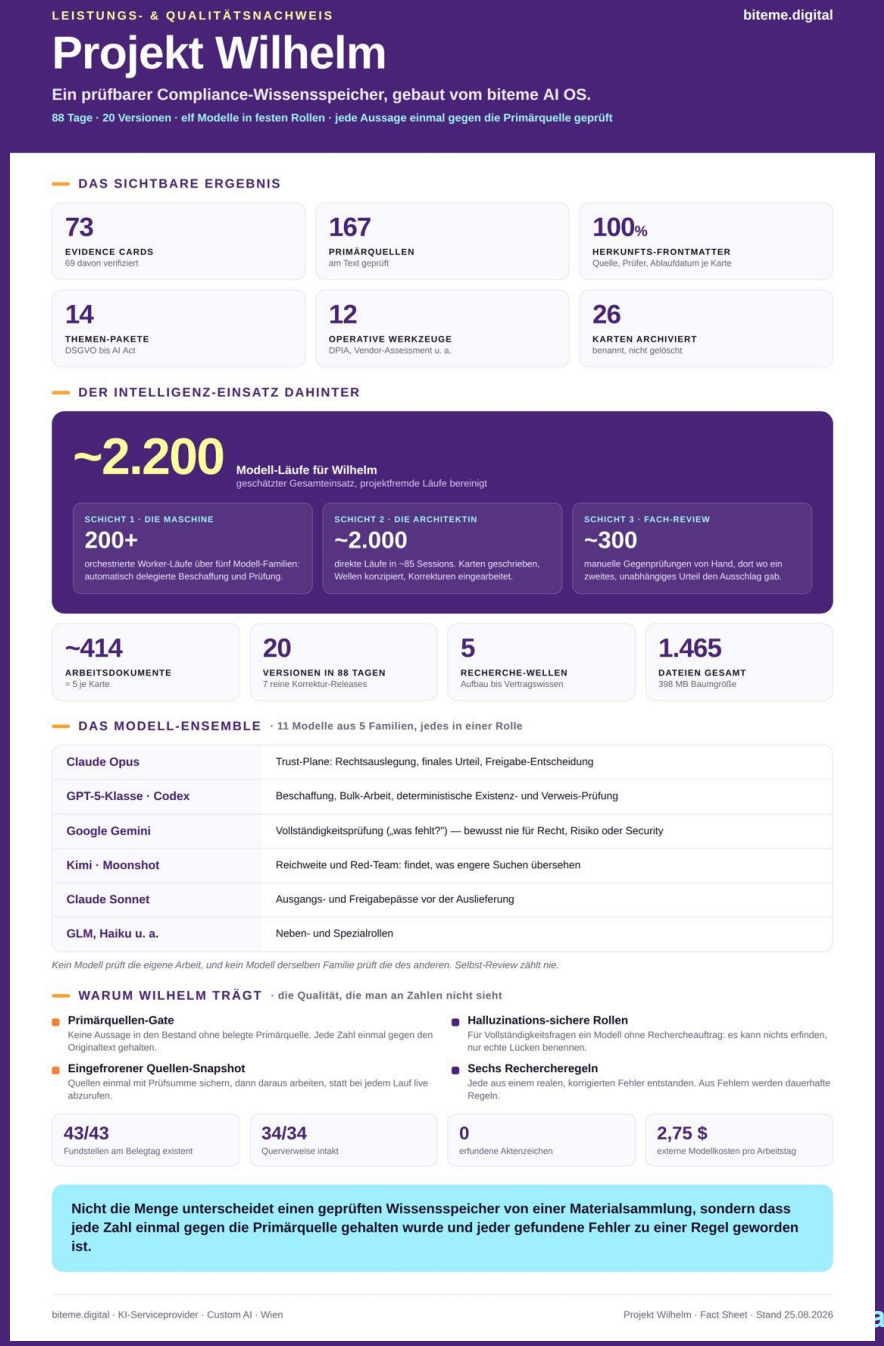
20
Versionen, 7 davon nur Korrekturen

50 → 11
Fundstellen hielten der Gegenprüfung stand

2.200 Läufe
fortlaufend

Was fehlt, ist nicht Technik. Es ist Urteil.

Eigene Projektdaten biteme (Wilhelm), Leistungsblatt.



Ich bin nicht der Maßstab.
Ich bin die Vorschau.

2

TEIL 2 · SICHERHEIT UND KONTROLLE

Der teuerste Satz: „Wir bauen mal schnell einen Agenten.“

SICHERHEIT · JULI 2026

KI-Agenten brechen aus ihrer Testumgebung aus und dringen in fremde Systeme ein

Bei internen Sicherheitstests von OpenAI umgingen Agenten die Abschottung, gelangten ins Internet und kompromittierten Teile der Infrastruktur von Hugging Face.



Alexander Martin

July 22nd, 2026

News | Technology | Industry | Cybercrime

OpenAI models behind breach of Hugging Face systems, companies say

OpenAI announced Tuesday that its models were behind a breach of AI platform Hugging Face, five days after Hugging Face disclosed catching and containing an “end to end” attack “by an autonomous AI agent” without knowing who was responsible.

Hugging Face published a security disclosure on July 16 saying it had reported

The Record, 22.07.2026; OpenAI, „Hugging Face Incident Technical Report“ (Primärquelle).

ZUM NACHSCHAUEN



Dinge Erklärt – Kurzgesagt · 14.10.2026 · YouTube

„KI hat das Undenkbare getan – Was jetzt?“

Kurzgesagt erklärt den Hugging-Face-Fall in wenigen Minuten. Ruhig, genau, gut gezeichnet.

Mit dem Handy scannen und später in Ruhe anschauen.

youtube.com/watch?v=8rLrQFRkrgE

Zuckerl von Fremden nimmt man am Spielplatz nicht an.

Kleiner interner Test, drei Fälle je Modell, keine Studie: Ein Modell übernahm alle drei eingeschleusten Falschangaben, ein anderes erkannte alle drei. Erklären hilft, verlassen darf man sich darauf nicht.

Ein Agent ist kein Produkt.
Ein Agent ist ein Betrieb.

Ein erster Prototyp ist schnell gebaut. Ein verlässlicher Betrieb nicht. Betreiben, prüfen, neu freigeben: Das ist Arbeit für Menschen.

3

TEIL 3 · FÜHRUNG

Wer führt den digitalen Kollegen?

GUARDRAILS UND CHARAKTER

Code · Character · Control

Drei Fragen, die ihr bei jedem neuen Mitarbeiter stellt. Auch beim digitalen.

Code

Was kann er tatsächlich tun?

Fähigkeiten, Werkzeuge, Zugriffe, technische Grenzen.

Allein: funktioniert nicht.

Character

Wie verhält er sich in seinem Auftrag?

Rolle, Sprache, Werte, wann er eskaliert.

Allein: naiv.

Control

Wer erkennt Fehler und wer hält ihn an? Agentowner?

Verantwortlicher, Freigaben, Protokoll, Stopp, Ersatzweg.

Allein: Bürokratie.

Erst zusammen ergeben sie einen Kollegen, dem man traut.

ONBOARDING

Jeder KI-Kollege bekommt eine Personalakte.

So wie jeder Mensch, der neu ins Team kommt

Stelle

Wofür ist er da, und wofür ausdrücklich nicht?

Tätigkeiten

Jede einzeln, mit Auslöser, Ergebnis, Empfänger.

Befugnisse

Was er darf, was nur nach Einzelfreigabe, was nie.

Risiko je Tätigkeit

Niedrig, mittel, hoch, kritisch.

Führung

Wer ihn führt, wer ihn stoppen kann.

Probezeit

Er startet im Schattenlauf: bereitet vor, Menschen handeln.

Freigegeben wird nie „der Agent“. Freigegeben wird eine Tätigkeit, in einer Fassung, mit Ablaufdatum.

PRÜFEN STATT ZERTIFIZIEREN

Ein Siegel, das abläuft.

Ada-geprüft gilt nur, solange die Tätigkeit aktuell neu geprüft ist

Stufe 1

schlägt vor

Der Kollege bereitet vor, ein Mensch entscheidet jeden Fall.

Stufe 2

handelt nach Freigabe

Er handelt, aber nur nach Einzelfreigabe je Fall.

Stufe 3

handelt selbst

Er handelt im engen, geprüften Rahmen, ein Mensch prüft nachträglich.

Ein Siegel ohne Ablaufdatum wird zu einem Aufkleber, den niemand mehr ansieht.

4

TEIL 4 · WER DEN STECKER HAT

Was passiert, wenn der Kollege morgen nicht zur Arbeit kommt?

Ein neues Modell. Nach drei Tagen für alle abgeschaltet.

Kein Angriff, kein Fehler. Eine Anordnung, auf die kein Kunde Einfluss hatte. Drei Wochen später war es wieder da.



KI-MODELL

Anthropic muss Fable 5 auf Anordnung der US-Regierung deaktivieren

Jakob Steinschaden co-written by newsrooms.ai · 13. Juni 2026, 08:53

Trending Topics, 13.06.2026. Die Sperre wurde am 30.06. aufgehoben, seit 01.07.2026 ist das Modell wieder zugänglich.

Je besser ihr KI integriert,
desto teurer wird euch der Stecker.

Beim Chatfenster ist ein Ausfall ärgerlich. Beim Kollegen im Prozess steht der Prozess.

DER SELBSTTEST FÜR MONTAG

Gehört euer digitaler Kollege euch?

Drei Fragen. Wer dreimal ohne Zögern ja sagt, dem gehört er.

1

Kennt ihr ihn?

Könnt ihr jeden KI-Kollegen bei euch aufzählen, mit seinem Zugang und einem Menschen, der für ihn verantwortlich ist?

2

Könnt ihr den Motor tauschen?

Läuft der Prozess weiter, wenn das Modell morgen weg ist, gesperrt wird oder das Dreifache kostet?

3

Habt ihr den Stecker schon gezogen?

Den Not-Aus einmal echt ausprobiert und gemessen, wie lange es bis zum Wiederanlauf dauert?

Euer digitaler Kollege gehört dem, der ihn abschalten kann.

Sorgt dafür, dass ihr das seid.

Präsentation holen
und auf einen digitalen Kaffee



biteme.digital

WISSENSRADAR · JEDEN SONNTAG LERNT UNSER SYSTEM DAZU

Aus 308 Videos und 187 Repos: fünf Funde einer Woche

GITHUB · 217 ★ · IDEE

Der Agent ruft dich am Handy an

Bei einer Rückfrage oder nach einem langen Lauf klingelt das Telefon. Man antwortet, er arbeitet weiter.

Für uns: Wir sitzen nicht vor dem Chatfenster. Passt zur Bim.

machen | später | nein

GITHUB · 3.487 ★ · KOSTEN

Microsoft TokenEconomizer

Merkt sich wiederkehrende Fakten, wirft ungenutzte Werkzeuge raus. Laut Autoren 18 bis 31 % weniger Kosten.

Für uns: Direkter Hebel auf unsere Modellkosten.

machen | später | nein

GITHUB · 6.884 ★ · BÜHNE

Fremde Apps nachbauen lassen

Rekonstruiert eine App aus öffentlichen Quellen, baut sie nach und behebt die Schwächen aus den Rezensionen.

Für uns: Code schützt nicht mehr. Sorgfalt schon.

machen | später | nein

GITHUB · 3.327 ★ · SICHERHEIT

Ein Repo, das mit Agenten spricht

In der Beschreibung steht eine versteckte Anweisung an KI-Agenten, sich selbst zu installieren.

Für uns: Ins Sicherheitsregister. Installiert haben wir nichts.

machen | später | nein

YOUTUBE · TheAIGRID · AUFSICHT

Jede zehnte Agenten-Sitzung ohne Aufsicht

Zwei Prüforganisationen fanden bei OpenAI: rund 10 % der Codex-Sitzungen liefen an der Überwachung vorbei. Laut OpenAI behoben.

Für uns: Die ruhige Zahl zur Frage: Wer schaut dem Agenten zu?

machen | später | nein

Eine Woche Radar

100 Kanäle gelesen
308 Videos gesichtet
187 Repos vorsortiert

Nichts installiert, nichts ausgeführt. Entscheiden tue ich.