WHITEPAPER, NOVEMBER 2026

KONZEPT



iogentic

Agentic Workflows

Regulierte Branchen, am Beispiel von Bauausschreibungen

Wie KI-Agenten Dokumentenarbeit im deutschen Bau- und Versorgungssektor übernehmen können und wie ihre Ergebnisse geprüft werden, bevor eine Person sie freigibt. Das Fallbeispiel ist das Leistungsverzeichnis einer öffentlichen Ausschreibung.

Dieses Konzeptpapier stammt vom Gründungsteam von iogentic. Die Plattform ist nicht veröffentlicht. Das Fallbeispiel und unsere Evaluierung beruhen auf synthetischen Projekten, und die Fotos sind KI-generierte Bilder.

Kurzfassung

Die wichtigsten Punkte, jeweils mit Quelle.

Deutsche Unternehmen führen KI ein, und ein wachsender Teil von ihnen befasst sich mit Agenten: Software, die Dokumente liest, Tools aufruft und Entwürfe erstellt. Ausschreibungen, Genehmigungen und Sicherheitsnachweise folgen schriftlichen Regeln und werden von einer namentlich benannten Person freigegeben. Deshalb muss ein Entwurf geprüft werden, bevor sich jemand darauf verlässt. Dieses Papier betrachtet einen Fall, das Leistungsverzeichnis einer öffentlichen Bauausschreibung, und beschreibt, wie die geplante Agentenplattform von iogentic ihn bearbeiten soll.

01 **Verbreitung**

37 Prozent der deutschen Unternehmen mit zehn oder mehr Beschäftigten nutzten 2026 KI [1]. In einer Bitkom-Umfrage setzen 11 Prozent der Unternehmen, die KI nutzen oder den Einsatz planen, bereits KI-Agenten ein [2].

02 **Ausschreibungen**

Knapp die Hälfte der 199.334 öffentlichen Vergaben, die für 2024 gemeldet wurden, waren Bauaufträge [3]. Für jeden davon muss die Leistung so beschrieben sein, dass alle bietenden Unternehmen sie gleich verstehen [4].

03 **Modellergebnisse**

In unserem Lauf hat keines von fünf aktuellen Modellen alle Aufgaben bestanden, die wir bewerten konnten [5]. In veröffentlichten Benchmarks scheitern Agenten oft an der Wiederholung einer Aufgabe, die sie zuvor gelöst haben [6, 7].

04 **Workflow-Design**

Das Design legt die Reihenfolge der Arbeit fest und formuliert die Prüfungen vor dem Entwerfen. Jeder Entwurfsagent bekommt eine getrennte Prüfinstanz, und an jedem Freigabepunkt steht eine Person [8, 9].

05 **Protokolle**

Die KI-Verordnung (AI Act) verlangt von Hochrisiko-KI-Systemen eine Protokollierung, technische Dokumentation und menschliche Aufsicht [10, 11, 12]. Die Plattform soll für jedes Ergebnis sechs Protokolle aufbewahren, damit eine prüfende Person nachvollziehen kann, was geschehen ist.

Übersicht

01 **Regulierte Arbeit** 04

Deutsche Unternehmen führen KI ein, während ihre Dokumentenarbeit schriftlichen Regeln folgt und mit einer Freigabe endet.

02 **Öffentliche Ausschreibungen** 06

Der öffentliche Bau wächst voraussichtlich schneller als die Branche insgesamt, und seine Ausschreibungen beschreiben die Leistung in einem Leistungsverzeichnis.

03 **Grenzen der Modelle** 09

Unser Evaluierungslauf und veröffentlichte Studien.

04 **Agentic Workflows** 11

Workflows und Agenten in einem Lauf.

05 **Die Plattform** 12

Eine geplante Runtime für Agenten, mit gemeinsamen Diensten in einem OS, einem abgeschlossenen Bereich je Organisation und einer App je Workflow.

06 **Der Harness** 15

Designstudie der Oberfläche.

07 **Fallbeispiel** 18

Vergabeunterlagen für eine 20-kV-Kabeltrasse, vom ersten Workshop bis zur erteilten Freigabe.

08 **Souveränität** 22

Wo Daten, Modell und Log laufen.

09 **Einstieg** 23

Ein Pilot in vier Schritten.



01

KAPITEL 01

Regulierte Arbeit

Deutsche Unternehmen führen KI ein, während ihre Dokumenten-arbeit schriftlichen Regeln folgt und mit einer Freigabe endet.

Verbreitung und Regeln

KI-Nutzung, Fachkräfte und die im November 2026 geltenden Regeln.

37 %

der deutschen Unternehmen mit zehn oder mehr Beschäftigten nutzten 2026 KI [1]

11 %

der Unternehmen, die KI nutzen oder den Einsatz planen, setzen laut Bitkom-Umfrage bereits KI-Agenten ein [2]

25,4 %

der Beschäftigten in Engpassberufen sind 55 Jahre oder älter [13]

Verbreitung

Die KI-Nutzung in deutschen Unternehmen ist gewachsen. Destatis zählt 37 Prozent der Unternehmen mit zehn oder mehr Beschäftigten, die 2026 KI nutzten, und 73 Prozent bei Unternehmen mit 250 oder mehr Beschäftigten [1]. Bitkom meldet 57 Prozent für Unternehmen mit 20 oder mehr Beschäftigten [2], und eine Nachrichtenmeldung zu einer ifo-Umfrage nennt 54,5 Prozent im Juni 2026 [14]. Die Erhebungen beruhen auf unterschiedlichen Grundgesamtheiten, daher unterscheiden sich die Werte.

Die Nutzung von Agenten ist noch gering. Von den Bitkom-Befragten, die KI nutzen oder den Einsatz planen, setzen 11 Prozent KI-Agenten ein, und 29 Prozent planen das [2]. Im Bau begann die KI-Nutzung später als in anderen Branchen. Laut derselben Nachrichtenmeldung stieg die KI-Nutzung im Bauhauptgewerbe innerhalb von drei Jahren von 7,1 auf 39,8 Prozent [14]. Die Nutzung generativer KI lag im Bau 2025 dagegen noch unter 15 Prozent [15].

Fachkräfte

Fachkräfte für Ausschreibungen und Pläne sind knapp. In den relevanten Bauberufen fehlten 2023 im Durchschnitt 70.323 Fachkräfte [16]. In zwei Kernberufen der Bauplanung sind 27,9 und 25,3 Prozent der Beschäftigten 55 Jahre oder älter [13]. Die Arbeitsproduktivität im deutschen Bau ist seit den 1990er-Jahren kaum gewachsen [16]. Beschäftigte in Architektur- und Ingenieurbüros verbringen etwa ein Viertel ihrer Zeit mit Arbeit außerhalb ihrer Projekte [17].

Regeln

Im selben Zeitraum haben sich mehrere Regeln für KI geändert. Die KI-Verordnung (AI Act) gilt grundsätzlich ab dem 2. August 2026, für einige Hochrisiko-KI-Systeme gelten spätere Termine [18]. Der Data Act gilt seit dem 12. September 2025 [19], und das deutsche NIS2-Umsetzungsgesetz wurde im Dezember 2025 verkündet [20]. Im Unternehmen bestimmt der Betriebsrat mit, wenn technische Einrichtungen Beschäftigte überwachen können, und er ist über Arbeitsverfahren zu unterrichten, die KI einsetzen [21, 22].



02

KAPITEL 02

Öffentliche Ausschreibungen

Der öffentliche Bau wächst voraussichtlich schneller als die Branche insgesamt, und seine Ausschreibungen beschreiben die Leistung in einem Leistungsverzeichnis.

Markt und Ausschreibungen

Bauvolumen, öffentliche Vergaben und die Regeln für Vergabeunterlagen.

544,48 Mrd.
Euro Bauvolumen in Deutschland 2025, Schätzung des DIW [23]

knapp 48 %
der 199.334 öffentlichen Vergaben 2024 waren Bauaufträge [3]

10 + 4 %
des Planerhonorars für Gebäude: Vorbereitung und Mitwirkung bei der Vergabe [24]

Markt

Das DIW schätzt das deutsche Bauvolumen für 2025 auf 544,48 Milliarden Euro, davon 97,62 Milliarden Euro im öffentlichen Bau [23]. Für 2026 und 2027 erwartet es, dass der öffentliche Bau real um 6,7 Prozent pro Jahr wächst, schneller als die Branche insgesamt [23]. Im Jahr 2024 zählte das Baugewerbe 20.857 Betriebe mit 20 oder mehr Beschäftigten [25].

Ausschreibungen

Öffentliche Auftraggeber in Deutschland meldeten für das Jahr 2024 insgesamt 199.334 Vergaben im Wert von 135,2 Milliarden Euro. Auf Bauaufträge entfielen knapp 48 Prozent der Anzahl und 47,7 Milliarden Euro des Werts [3]. Oberhalb des EU-Schwellenwerts, der für Bauleistungen seit Januar 2026 bei 5.404.000 Euro liegt, gelten die europäischen Vergaberegeln [26, 27].

Die VOB/A verlangt, die Leistung eindeutig und so erschöpfend zu beschreiben, dass alle Unternehmen die Beschreibung im gleichen Sinne verstehen [4]. Das Gesetz gegen Wettbewerbsbeschränkungen (GWB) fordert eine Beschreibung, die so eindeutig wie möglich ist [28]. Nach der HOAI entfallen auf die Vorbereitung der Vergabe 10 und auf die Mitwirkung bei der Vergabe 4 Prozent des Planerhonorars für Gebäude [24].

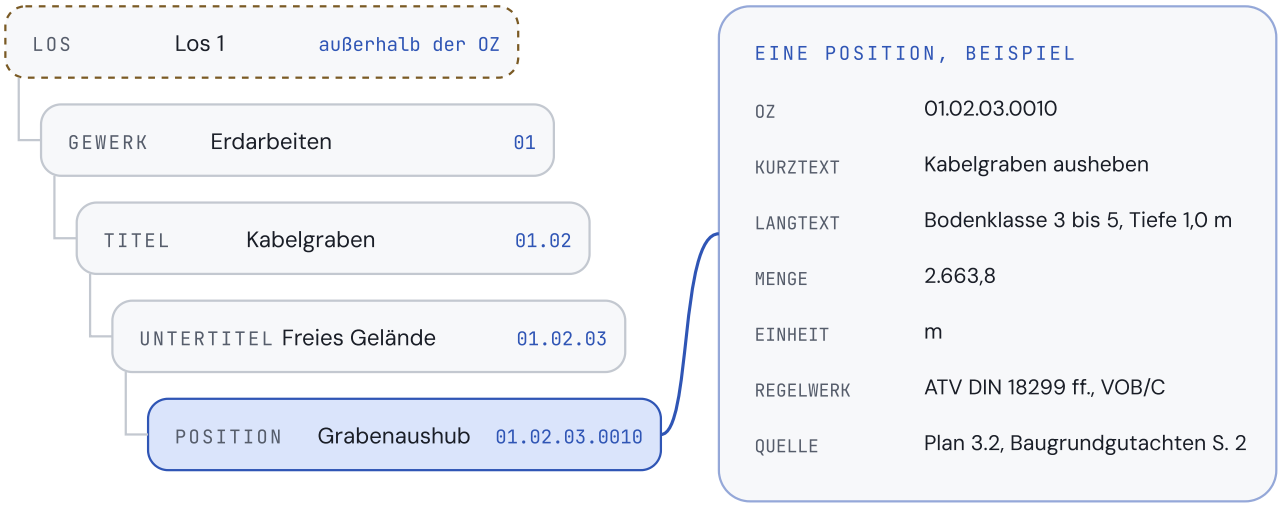
Kapazität

Öffentliche Arbeitgeber schrieben 2025 fast 12.400 Stellen in Bauberufen auf allen Qualifikationsstufen aus, 73 Prozent davon in Kommunen [29]. Eine Auswertung von fünf abgeschlossenen Bauprojekten in Bremen aus dem Jahr 2017 nannte Fehler im Leistungsverzeichnis als eine von fünf Ursachengruppen für Nachträge [30]. Sie schlug eine konsequente Prüfung des Leistungsverzeichnisses nach dem Vier-Augen-Prinzip vor, die Zeit und Fachkenntnis braucht [30].

BEISPIELPOSITION

Leistungsverzeichnis

Ebenen, Positionen, Standardtexte und die GAEB-Austauschphasen.



Ein Leistungsverzeichnis (LV) ist hierarchisch aufgebaut. Gewerke, Titel und Untertitel gliedern die Leistung. Die Positionen auf der untersten Ebene beschreiben sie, jeweils mit Menge, Einheit und einem Verweis auf die technischen Regeln der VOB/C [31]. Die Ordnungszahl nummeriert jedes Element, und GAEB DA XML erlaubt bis zu fünf Ebenen und 14 Stellen [32]. Eine Ausschreibung kann außerdem in Lose aufgeteilt werden, und die Richtlinie eines Betreibers führt die Lose außerhalb der Ordnungszahl [33].

Standards legen die Texte fest. Das STLB-Bau liefert Standardtexte für Positionen und wird von den Arbeitskreisen des GAEB erstellt [34]. Die ATV DIN 18299 enthält die allgemeinen Regelungen für Bauarbeiten jeder Art [35]. Die DIN 276 gliedert die Kosten [36], und die DIN 18290 verknüpft Bauwerksmodelle mit GAEB-Leistungsverzeichnissen [37].

GAEB DA XML ist das Austauschformat, aktuell in Version 3.3 und mit einer Betaversion 3.4 [38]. Seine Phasen beschreiben, wer was sendet: X83 ist die Angebotsaufforderung vom Auftraggeber an die Bieter, X84 das Angebot, das zurückkommt, und X85 ein Nebenangebot [39]. Eine Landesbauverwaltung etwa lässt Bieter die Struktur der Datei nicht ändern [40] und verlangte 2024 noch Version 3.2 [40].



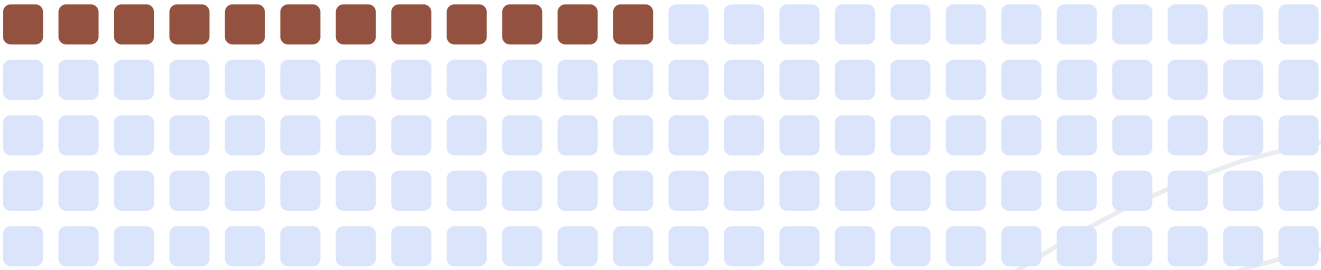
GEMESSEN, SYNTHETISCHE DATEN

Grenzen der Modelle

Unser Evaluierungslauf vom 8. November 2026 mit synthetischen Projekten.

Wir haben im November 2026 fünf aktuelle Open-Weight-Modelle an zwei synthetischen Ausschreibungsprojekten getestet. Jedes Modell arbeitete in einer Agentenschleife mit Tools an vier Aufgaben, dreimal je Projekt, insgesamt also 120 Versuche [5]. Es sind die Aufgaben einer ausschreibenden Stelle: Projektfakten erfassen, Verweise prüfen, Positionen des Leistungsverzeichnisses entwerfen und die Vorbemerkungen schreiben.

Kein Modell hat alle Aufgaben bestanden, die wir bewerten konnten. Ein Modell bestand die Projektfakten, ein anderes die Positionen. Die Verweise waren die leichteste Aufgabe, und vier Modelle haben sie bestanden. In 12 der 120 Versuche enthielt die Antwort eine Zahl, die kein Tool geliefert hatte [5].



Ein Quadrat je Versuch. Rot: 12 Versuche mit einer Zahl, die kein Tool geliefert hat, zusammengefasst; ihre Lage hat keine Bedeutung.

MODELL	T1 PROJEKTFAKTEN	T2 VERWEISE	T3 LV-POSITIONEN	T4 VORBEMERKUNGEN
Modell A	3,5 nicht bestanden	3,2 nicht bestanden	4,5 nicht bestanden	5,2 kein Urteil
Modell B	3,9 nicht bestanden	6,0 bestanden	5,4 bestanden	5,5 kein Urteil
Modell C	2,9 nicht bestanden	5,5 bestanden	3,8 nicht bestanden	5,3 kein Urteil
Modell D	4,2 bestanden	6,0 bestanden	4,4 nicht bestanden	6,0 kein Urteil
Modell E	3,9 nicht bestanden	6,0 bestanden	4,9 nicht bestanden	6,0 kein Urteil

Mittlere Bewertung aus sechs Versuchen auf einer Skala von 1 bis 6, wobei 6 die beste Bewertung ist. Bestanden: Mittelwert mindestens 4 und 80 % der Versuche mit mindestens 4. T4 hat kein Urteil, weil die Prüfmodelle nicht kalibriert sind. Modellnamen werden nicht genannt. [5]

Veröffentlichte Studien

Studien zur Zuverlässigkeit von Agenten, zu erfundenen Antworten und zu Prüfmodellen.

Auch veröffentlichte Studien berichten von Problemen mit der Zuverlässigkeit. Auf tau-bench löste 2024 ein führender Agent weniger als die Hälfte seiner Aufgaben und wurde weniger konsistent, wenn dieselbe Aufgabe wiederholt wurde [6]. Eine Studie mit 15 Modellen aus dem Jahr 2026 stellte fest, dass die jüngsten Zuwächse an Leistungsfähigkeit die Zuverlässigkeit nur wenig erhöht haben [7]. Wenn Modelle Fragen zu Dokumenten beantworten, die ihnen vorliegen, steigt der Anteil erfundener Antworten mit der Länge des Kontexts [41].

Ein zweites Modell als Prüfinstanz hat Grenzen. Ein starkes Prüfmodell kann etwa so oft mit Menschen übereinstimmen, wie Menschen untereinander übereinstimmen [42]. Einfache Übereinstimmungsraten überschätzen das allerdings, weil sie zufällige Übereinstimmung nicht berücksichtigen [43]. Unsere beiden Prüfmodelle erreichten ein Kappa von 0,5 und blieben damit unter der von uns gesetzten Schwelle von 0,6. Deshalb zählt noch keine Bewertung durch ein Prüfmodell [5].

Unser Vorgehen beruht auf den drei unten genannten Maßnahmen. Ihre Wirkung haben wir noch nicht in einem Lauf gemessen. Schon der passende Kontext macht Antworten faktentreuer [44], und besseres Retrieval hat in einem veröffentlichten Test die fehlgeschlagenen Abrufe um zwei Drittel verringert [45].

Fachwissen

Regeln, Begriffe und Fehlerfälle der ausschreibenden Stelle, im Workshop festgehalten, bevor ein Agent etwas entwirft.

Workflows mit Prüfungen

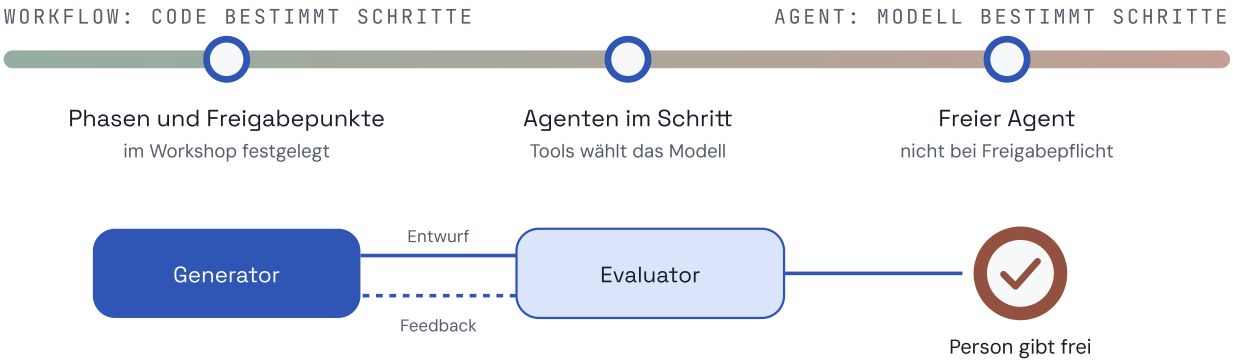
Feste Phasen, ein Prüfagent getrennt vom Entwurfsagenten, Code-Prüfungen für jedes Ergebnis und Menschen an den Freigabepunkten.

Modelle mit Kontext

Ihre Regeln, Normen und früheren Fälle gehen über Konnektoren als Kontext an das Modell.

Agentic Workflows

Workflows und Agenten in einem Lauf.



Anthropic unterscheidet zwei Arten von Systemen. In einem Workflow gibt Code den Weg vor, und das Modell füllt die Schritte aus. In einem Agenten entscheidet das Modell selbst über den nächsten Schritt und darüber, welches Tool es nutzt [8]. Regulierte Arbeit braucht beides. Die Reihenfolge der Arbeit ist durch Regeln festgelegt, aber innerhalb eines Schritts braucht es Urteilsvermögen, etwa um ein Baugrundgutachten zu lesen oder einen Plan einer Vorlagenposition zuzuordnen.

Das Design stützt sich auf drei veröffentlichte Muster. Im Evaluator-Optimizer-Muster erstellt ein Modellaufruf den Entwurf, ein anderer prüft ihn und gibt Feedback [8]. Checkpoints halten den Lauf dort an, wo eine Person entscheiden muss [8], und OpenAI beschreibt menschliches Eingreifen als Schutzmaßnahme bei Aktionen mit hohem Risiko [9]. Evals messen einen Ausgangswert, bevor jemand einen Prompt nachjustiert [9], und sie sollten neben der endgültigen Antwort auch den Weg der Tool-Aufrufe bewerten [46].

Agenten nutzen Tools über eine Standardschnittstelle. Anthropic hat das Model Context Protocol im November 2024 als offenen Standard vorgestellt, der Modelle mit Daten und Tools verbindet [47]. Seine Spezifikation verlangt die Zustimmung der nutzenden Person, bevor ein Tool verwendet wird [48].

ELEMENT	IM KONZEPT DER PLATTFORM
Feste Phasen und Freigabepunkte	Im Workshop festgelegt. Ein Lauf kann keinen Freigabepunkt überspringen.
Entwurfs- und Prüfagent	Der Prüfagent arbeitet getrennt vom Entwurfsagenten und gibt Befunde zurück.
Menschen an Freigabepunkten	Freigabepunkt 1 vor dem Entwurf, Freigabepunkt 2 vor der Herausgabe.
Evals vor dem Nachjustieren	Eingebaute Testfehler und Code-Prüfungen laufen vor Freigabepunkt 2.
Tools über MCP	Konnektoren zu Ihren Systemen, im Beispiel nur lesend.



05

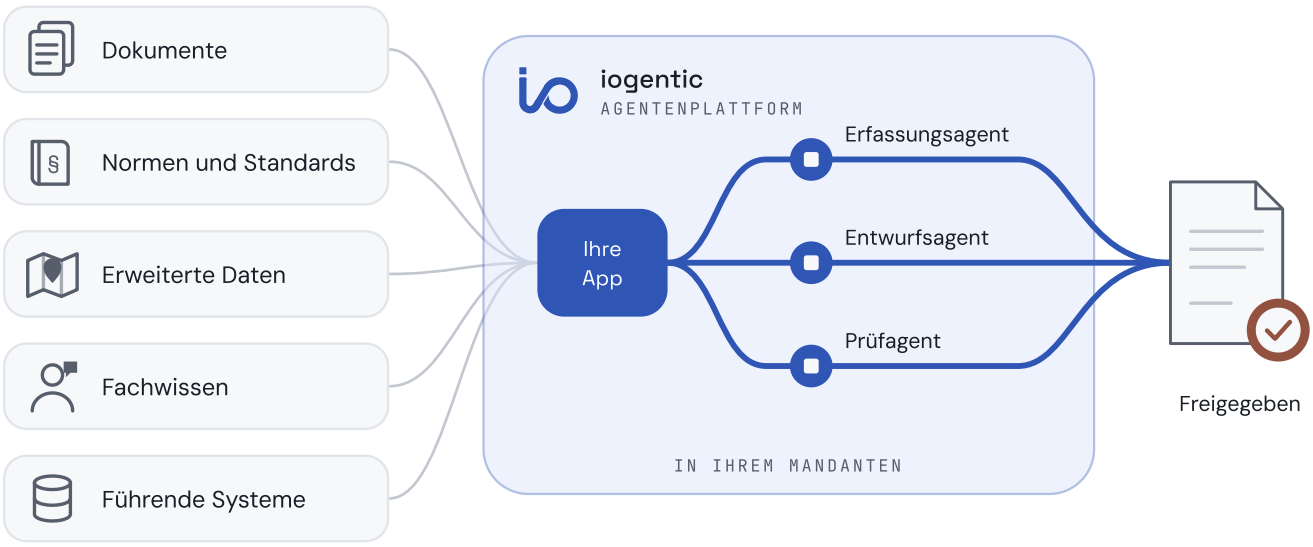
KAPITEL 05

Die Plattform

Eine geplante Runtime für Agenten, mit gemeinsamen Diensten in einem OS, einem abgeschlossenen Bereich je Organisation und einer App je Workflow.

Architektur

Von den Eingaben bis zum freigegebenen Ergebnis.



Nach unserem Plan lernen wir zuerst den Workflow in einem Workshop mit den Menschen kennen, die die Arbeit machen. Dann bauen wir ihn als App auf einem Agentic OS und testen jeden Schritt, bevor sich jemand darauf verlässt. Wo eine Regel eine Person verlangt, soll der Lauf auf ihre Freigabe warten.

Die Plattform besteht aus drei Produkten. Der iogentic Harness ist die App, in der Menschen Agenten bauen, ausführen und prüfen. Er entsteht als native App. Pipesmith (pipesmith.ai) soll Agenten bauen, prüfen und ablegen. CI und Registry von Pipesmith sind in Entwicklung, die QA ist ein Vorschlag. Die iogentic Runtime soll einen Agenten nur aus einem geprüften Build starten, ihn in einer Sandbox ausführen und jeden Schritt aufzeichnen. Sie ist ein Konzept ohne Code.

PRODUKT	STUFE	STAND AM 24. NOVEMBER 2026
iogentic Harness	Entwicklung	In Entwicklung, native App
Pipesmith	Build, Prüfung, Ablage	CI und Registry in Entwicklung, QA als Vorschlag
iogentic Runtime	Lauf	Konzept, noch kein Code

KONZEPT, NICHT GEBAUT

Agentic OS

Im Konzept stellt das OS jedem Workflow dieselben acht Dienste bereit.



OS: ACHT GEMEINSAME DIENSTE

Ein Mandant ist der abgeschlossene Bereich einer Organisation, mit eigenen Daten, Zugangsdaten, Modellen und Agenten und ohne Verbindung zu anderen Mandanten. Eine App ist ein Workflow mit eigener Oberfläche. Apps tauschen Daten nur über typisierte Ereignisse auf dem Bus aus, und Ihre Systeme bleiben das führende System.

Scheduler

Ein Lauf ist ein Prozess mit neun Zuständen, von „angelegt“ bis „abgeschlossen“.

Wissensspeicher

Wissen wird aus Ihren Systemen gelesen. Agenten schreiben Einträge, und eine Person gibt sie frei.

Agentenablage

Agenten und Fähigkeiten werden mit einem Manifest wie Pakete installiert.

Modell-Gateway

Zulässige Modelle und Regionen je Mandant. Modell und Version werden je Aufruf protokolliert.

Konnektoren

Treiber zu Ihren Systemen, mit zulässigen Tools je Konnektor.

Audit-Log

Einträge werden nur angefügt und getrennt von den Nachweisen geführt.

Event-Bus

Typisierte Ereignisse: Schritt, Freigabepunkt wartet, Freigabe erteilt, Nachweis ergänzt, Export.

Prüfstand

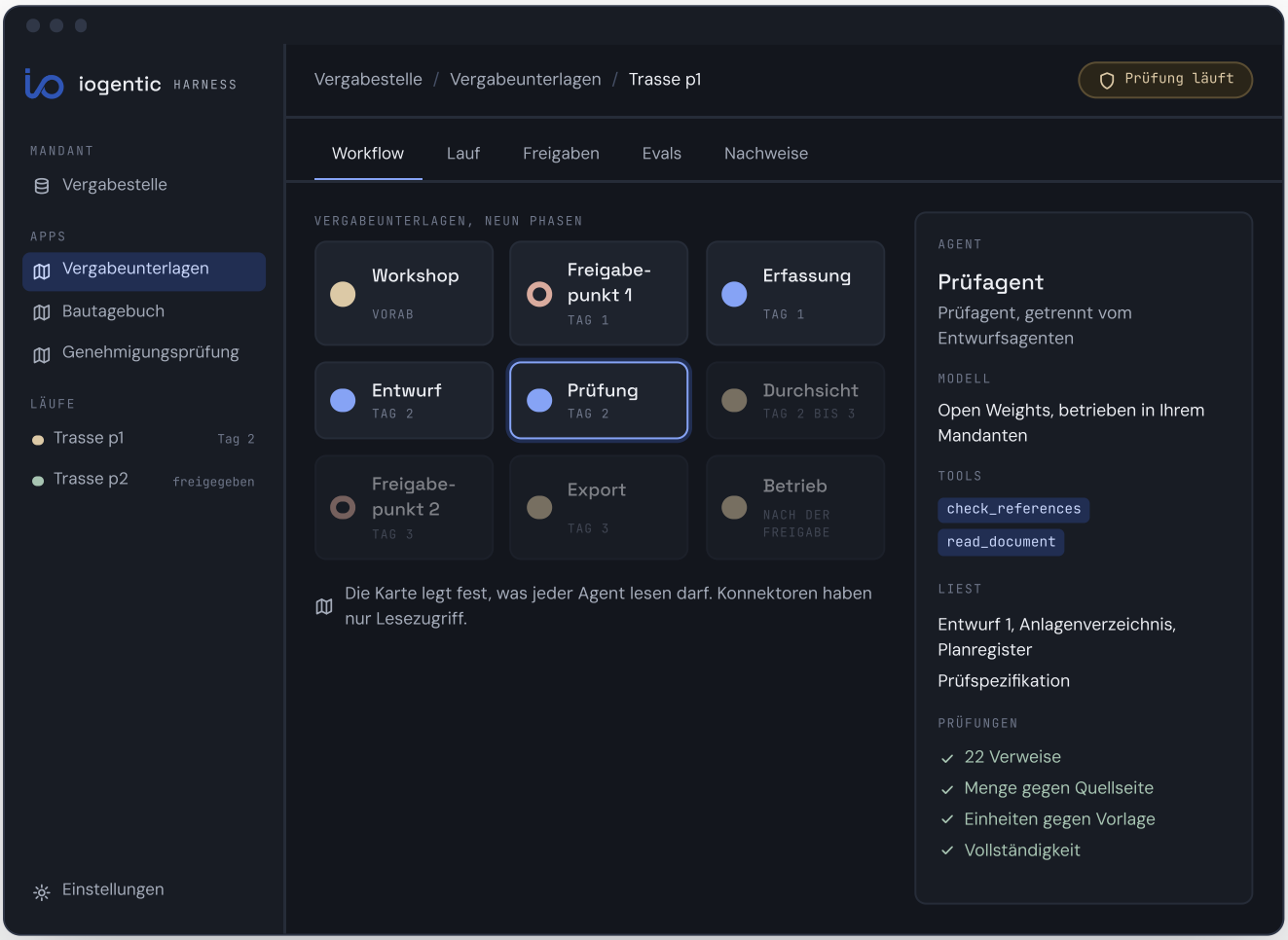
Code-Prüfungen und Prüfmodelle laufen gegen absichtlich eingebaute Fehler.

UI-STUDIE, BEISPIELDATEN

Der Harness

Designstudie des Harness als Oberfläche der Plattform.

Der Harness ist die Oberfläche, über die Menschen die Plattform nutzen. Er entsteht als native Desktop-App. Er zeigt das Gespräch mit einem Agenten und jeden Tool-Aufruf mit seiner Dauer, und er fragt nach einer Freigabe, bevor der Agent etwas ändert. Die Designstudie unten nutzt Beispieldaten und zeigt vier Ansichten: den Workflow einer App, einen Lauf, die Freigabe an Freigabepunkt 2 und die Evaluierungsmatrix. Die Studie ist nicht gebaut.

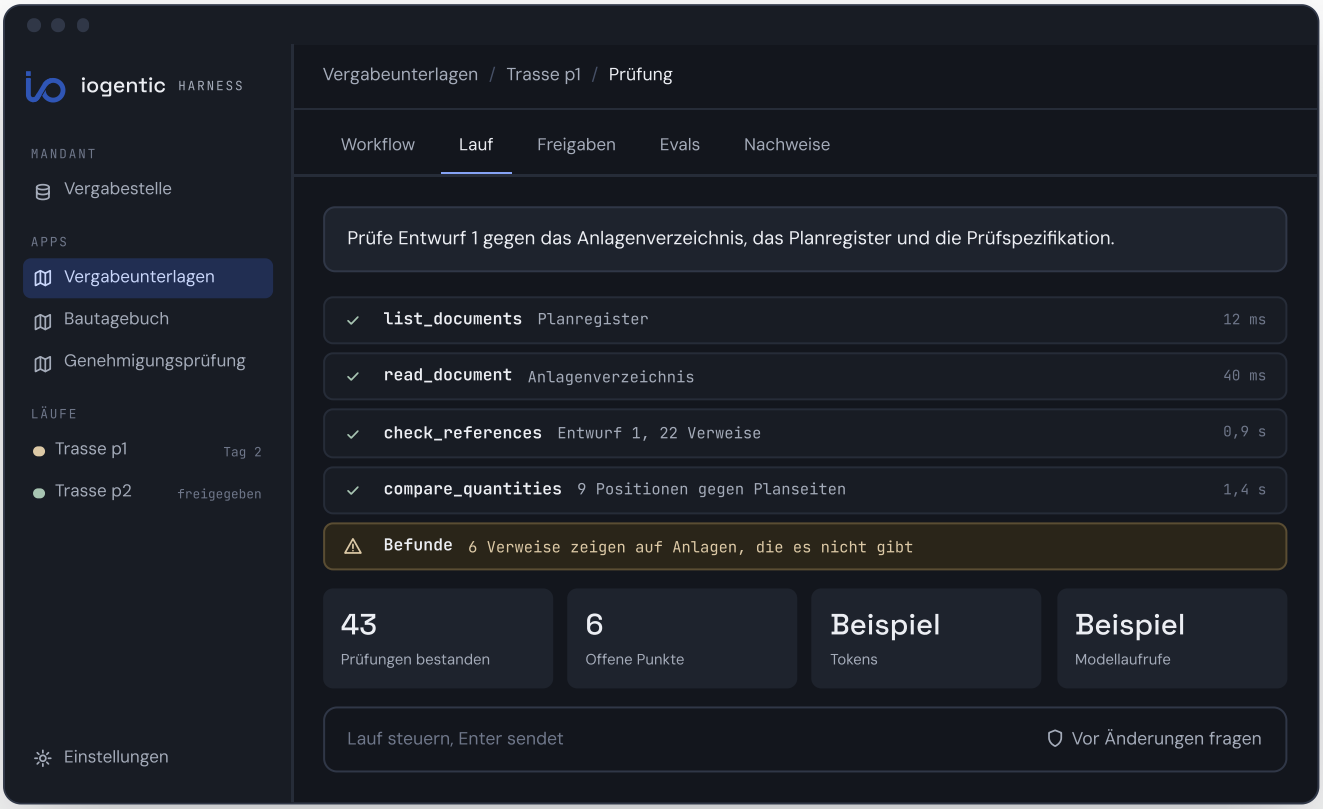


Workflow Der Inspektor listet Tools, Lesezugriffe und Prüfungen des Prüfagenten auf.

UI-STUDIE, BEISPIELDATEN

Lauf

Ein Lauf des Prüfagenten zum ersten Entwurf.

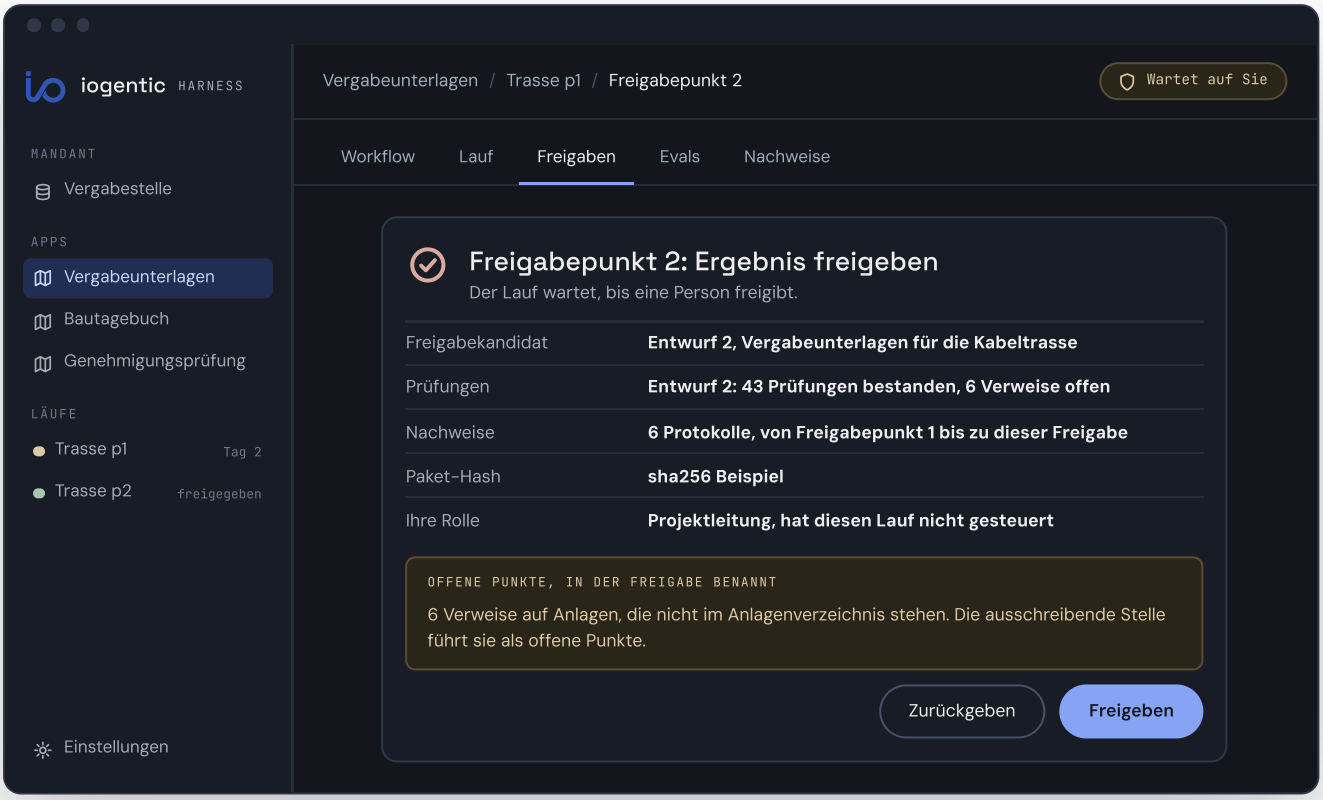


Lauf Jeder Tool-Aufruf mit seiner Dauer. Die Befunde des Prüfagenten gehen an die ausschreibende Stelle.

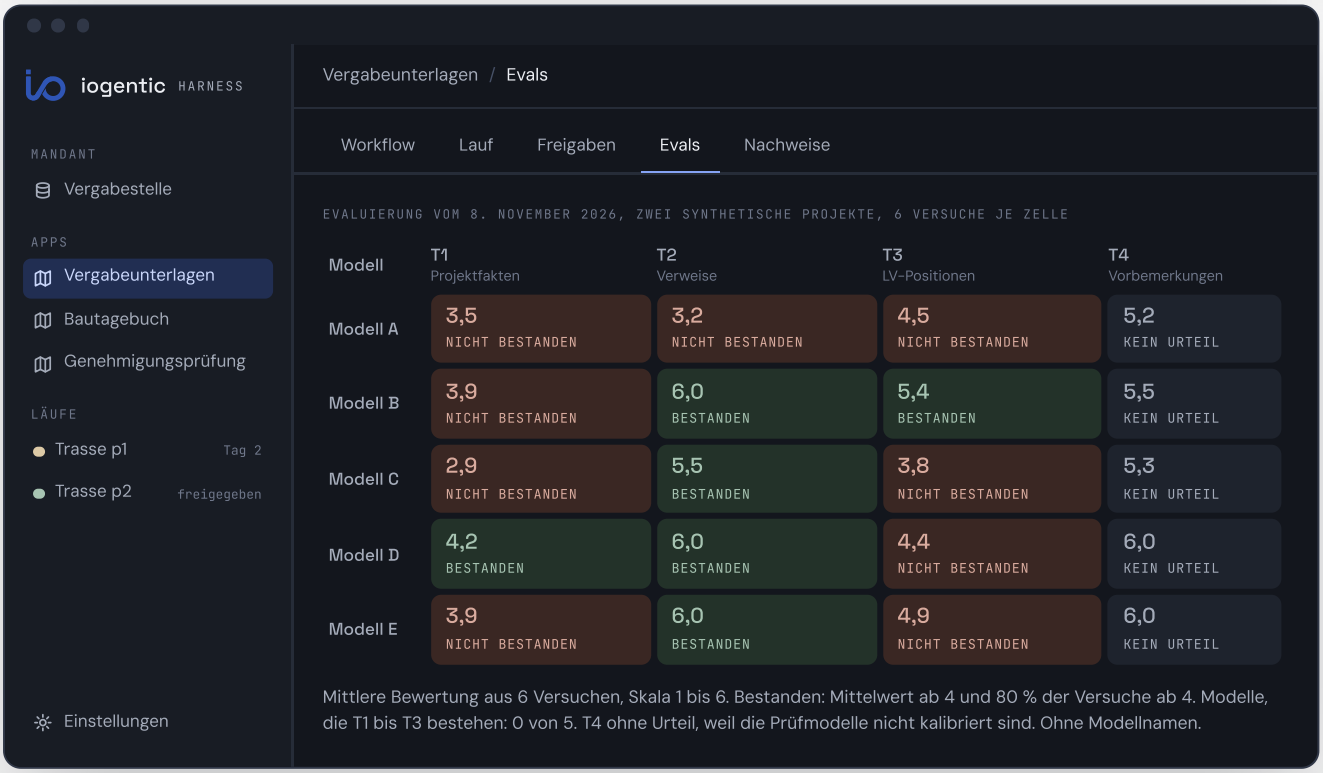
Im Konzept zeigt jeder Lauf die Aufgabe, jeden Tool-Aufruf mit seiner Dauer, die bestandenen Prüfungen und die Befunde. Wo eine Regel eine Person verlangt, wartet der Lauf und zeigt an, dass er auf eine Person wartet. Das Team kann einen Lauf steuern, während er arbeitet, und jede Änderung an Dateien braucht vorher eine Freigabe.

Freigabe und Evaluierung

Die Freigabe an Freigabepunkt 2 und die Evaluierungsmatrix.



Freigabepunkt 2 Freigabe mit benannten offenen Punkten.



Evals Die Matrix unseres Laufs vom 8. November 2026, gemessen an synthetischen Projekten, mit anonymisierten Modellen. [5]



07

KAPITEL 07

Fallbeispiel

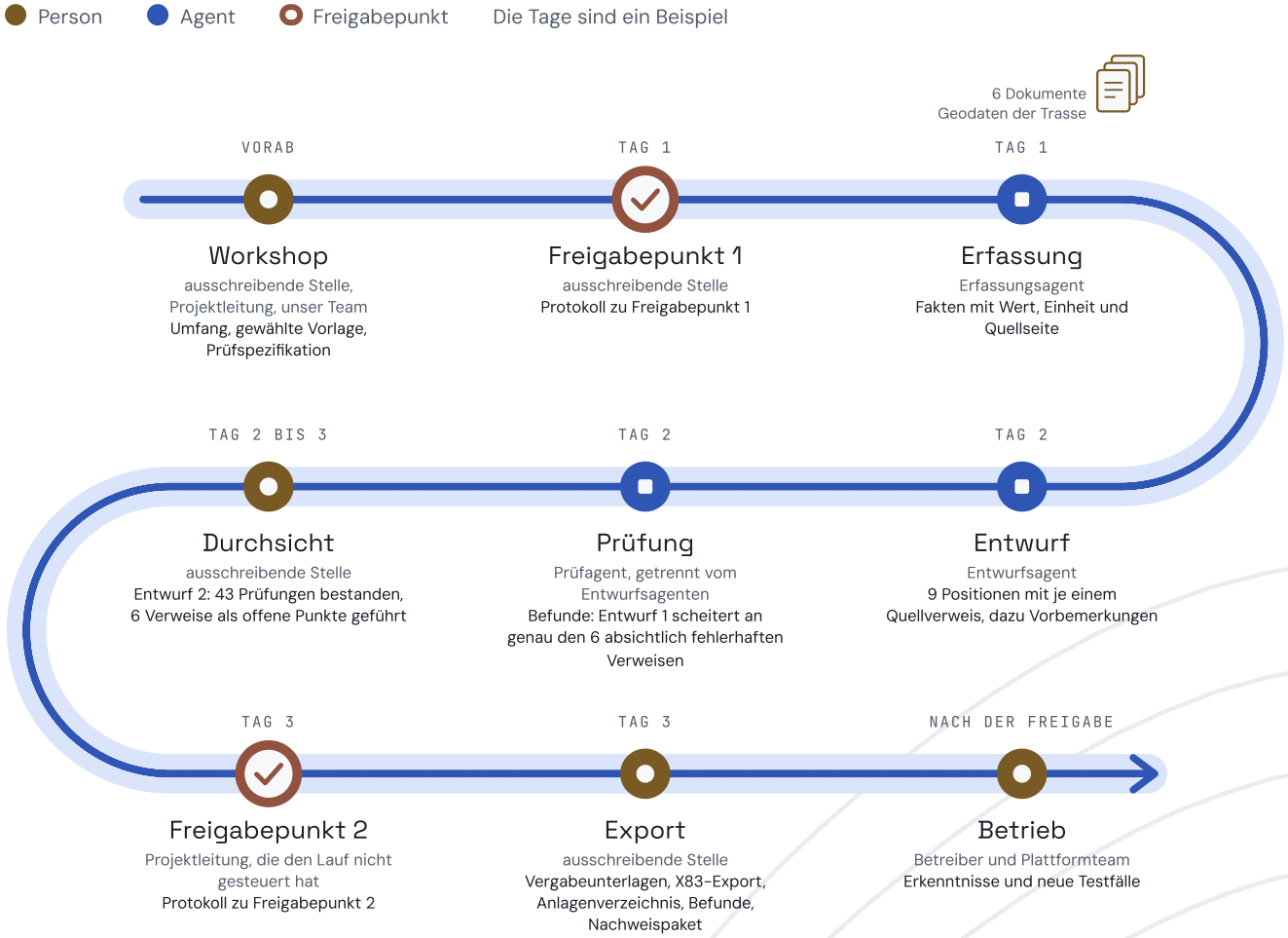
Vergabeunterlagen für eine 20-kV-Kabeltrasse, vom ersten Workshop bis zur erteilten Freigabe.

Das Projekt ist synthetisch, der Lauf ist vorab festgelegt. Kein Modell wurde aufgerufen, und es wurden keine Kundendaten verwendet.

SYNTHETISCH, VORAB FESTGELEGT

Neun Phasen

Phasen, Beteiligte und Freigabepunkte des vorab festgelegten Laufs.

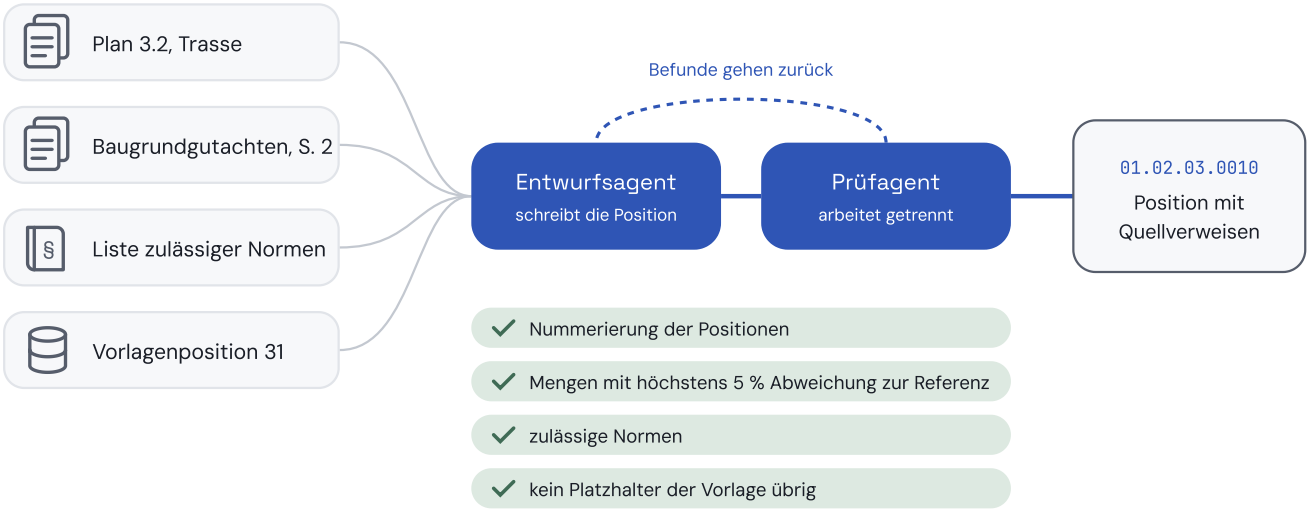


20-kV-Kabeltrasse, 2.663,8 m, 17 Kreuzungen: 11 Wege, 1 Hauptverkehrsstraße, 1 Bahnstrecke, 4 Gewässer. 25 Fakten, 22 Verweise, davon 6 absichtlich fehlerhaft, 9 Positionen. Die Referenzfälle sind noch nicht fachlich geprüft. [49]

BEISPIEL

Eine Position entwerfen

Wie eine Position entworfen und geprüft wird.



Im Beispiel geht jede Position von Quellen aus: dem Plan, dem Baugrundgutachten, der Liste zulässiger Normen und der Vorlagenposition. Der Entwurfsagent schreibt die Position mit einem Verweis auf jede Quellseite. Der Prüfagent arbeitet getrennt davon, vergleicht Mengen mit den Planseiten und Einheiten mit der Vorlage und schickt Befunde zurück [8]. Positionen, die nicht bestehen, gehen an den Entwurfsagenten zurück. Die ausschreibende Stelle sieht die bestandenen Positionen zusammen mit den Befunden.

Protokolle

Die sechs geplanten Protokolle und die KI-Verordnung.



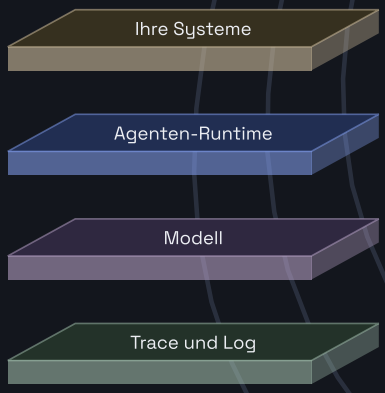
Im Konzept hält der Lauf für jedes Ergebnis sechs Protokolle fest, vom Protokoll zu Freigabepunkt 1 bis zur erteilten Freigabe. Sie zeigen, wer den Umfang festgelegt und wer das Ergebnis geprüft hat.

Die KI-Verordnung (AI Act) verlangt von Hochrisiko-KI-Systemen ein Risikomanagementsystem, eine technische Dokumentation, eine automatische Protokollierung und eine wirksame menschliche Aufsicht [50, 10, 11, 12]. Ob ein Agent, der Vergabeunterlagen entwirft, als hochriskant gilt, ist im Einzelfall zu prüfen. Für Hochrisiko-KI-Systeme nach Anhang III gelten die Pflichten ab dem 2. Dezember 2027, nachdem der Digital-Omnibus den Termin verschoben hat [18, 51]. Die Tabelle ordnet vier der sechs geplanten Protokolle diesen Artikeln zu. Das ist unsere eigene Lesart, ohne rechtliche Prüfung.

PROTOKOLL	INHALT	KI-VERORDNUNG
Protokoll zu Freigabepunkt 1	wer den Umfang festgelegt hat, Eingaben und Prüfspezifikation	Art. 14, menschliche Aufsicht [12]
Trace und Prompts	jeder Schritt, Tool-Aufruf und Modellaufruf	Art. 12, Aufzeichnungspflichten [11]
Modelle und Tools	welches Modell und welche Tools, mit Versionen	Art. 11, technische Dokumentation [10]
Prüfergebnisse	Code-Prüfungen und eingebaute Fehler gegen die Anforderung	Art. 9, Risikomanagement [50]

Souveränität

Daten, Modell und Log können in der EU bleiben, und Sie wählen, wo jede Ebene läuft.



Ihre Systeme	Bleiben, wo sie sind. Agenten lesen über Konnektoren.
Agenten-Runtime	Rechenzentrum in Deutschland, EU-Cloud-Region oder Ihre eigenen Server.
Modell	Selbst betriebene Open Weights, ein in der EU gehosteter Dienst oder per Opt-in eine API in einer EU-Region.
Trace und Log	Als Dateien exportiert. Die Aufbewahrung legen Sie fest.

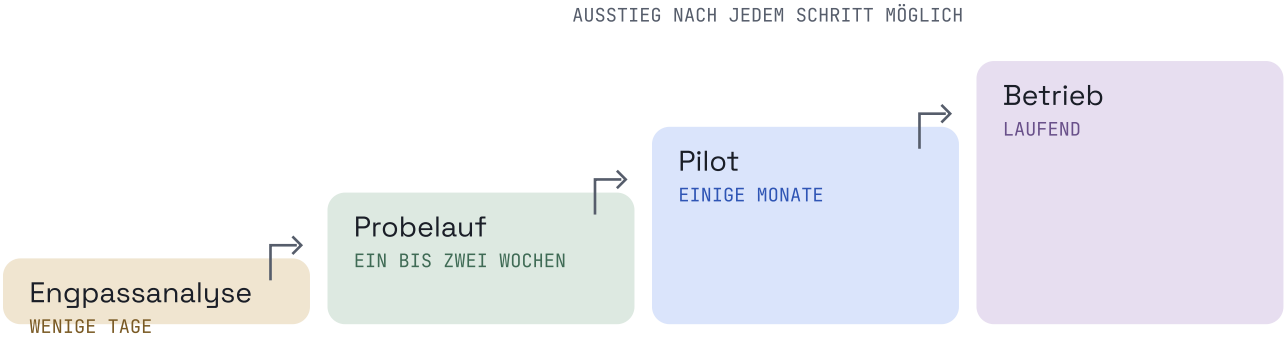
Die Datenschutz-Grundverordnung (DSGVO) gibt Menschen das Recht, nicht einer ausschließlich auf automatisierter Verarbeitung beruhenden Entscheidung unterworfen zu werden [52]. NIS2 teilt Einrichtungen in besonders wichtige und wichtige Einrichtungen ein, jeweils mit eigenen Pflichten [53]. ISO/IEC 42001 spezifiziert ein Managementsystem für KI [54], und das Bundesamt für Sicherheit in der Informationstechnik (BSI) beschreibt die Risiken generativer KI-Modelle für Industrie und Behörden [55]. Wird KI eingeführt, kann der Betriebsrat eine sachverständige Person hinzuziehen [56]. Bevor ein Pilot echte Dokumente verarbeitet, wird sein Aufbau mit Ihrer oder Ihrem Datenschutzbeauftragten anhand dieser Texte geprüft.

Wir sind keine Juristinnen und Juristen, und wir haben kein Zertifikat. Ob ein Aufbau den Regeln genügt, hängt vom Anwendungsfall ab. Die Plattform soll die Protokolle erzeugen, die eine prüfende Stelle nachprüfen kann.

Einstieg

Ein Pilot in vier Schritten, von der Engpassanalyse bis zum Betrieb.

Der erste Pilot würde die Vergabeunterlagen für ein Vorhaben abdecken, von vollständigen Plansätzen bis zum freigegebenen Entwurf, parallel zu Ihrem üblichen Prozess. Die Schritte drei und vier brauchen die Runtime, die heute ein Konzept ist. Preise, Dauer und Abbruchkriterien vereinbaren wir mit Ihnen vor der ersten Runde.



Eine Übersicht, wo sich Agenten in Ihrer Arbeit lohnen könnten, und ein Blick auf die Daten.

Ein erster Entwurf zu einem echten Fall, teils vor Ort, mit einer Schätzung und ihrer Herleitung.

Ergebnisse, die namentlich benannte Personen in Ihrer Organisation freigegeben haben.

In Ihrer Umgebung oder einer souveränen Cloud in Deutschland, betrieben von uns oder von Ihrem Team.

Was Sie behalten

- Eine Prozesslandkarte
- Ein Entwurf zu einem echten Fall
- Freigegebene Ergebnisse mit je sechs Protokollen
- Dokumentation, wie jeder Agent gebaut und getestet ist
- Ihre Plansätze, Vorlagen, Prüfspezifikationen und Ihr Fachwissen

Was der Pilot misst

- Bearbeitungstage je Ausschreibung, aus einer Zeiterfassung
- Fehler in Anlagenverweisen
- Annahmequote an Freigabepunkt 2 in der ersten Runde
- Kosten je Ausschreibung: Modell, Hosting und Bearbeitungsstunden

Glossar

Die Begriffe aus Bauausschreibungen und zur Plattform, die in diesem Whitepaper vorkommen.

Agent	Ein Modell, das seinen nächsten Schritt und das passende Tool selbst wählt.
Agentenablage	Die Registry des OS, aus der Agenten und Fähigkeiten mit einem Manifest wie Pakete installiert werden.
App	Ein Workflow auf der Plattform, mit eigener Oberfläche, eigenen Agenten, Regeln und Konnektoren.
ATV	Allgemeine Technische Vertragsbedingungen, die technischen Regeln der VOB/C, etwa DIN 18299.
Evaluierung	Ein wiederholbarer Test eines Agenten an bekannten Fällen, mit Bewertungen und einer Bestehensregel.
Freigabepunkt	Eine Stelle im Lauf, an der eine namentlich benannte Person freigibt. Der Lauf wartet, bis die Freigabe vorliegt.
GAEB	Gemeinsamer Ausschuss Elektronik im Bauwesen, Herausgeber des Austauschformats GAEB DA XML.
Harness	Die App, in der Menschen Agenten bauen, ausführen, prüfen und freigeben.
HOAI	Die Honorarordnung für Architekten und Ingenieure in Deutschland, mit neun Leistungsphasen.
Kappa	Ein Maß für die Übereinstimmung zwischen zwei Bewertenden, das zufällige Übereinstimmung herausrechnet.
Leistungsverzeichnis	Kurz LV: die gegliederte Beschreibung der Leistung in einer Ausschreibung.
Mandant	Der abgeschlossene Bereich einer Organisation auf der Plattform: ihre Daten, Zugangsdaten, Modelle und Agenten.
MCP	Model Context Protocol, ein offener Standard, der Modelle mit Tools und Daten verbindet.
Ordnungszahl	Die Nummer eines Elements im Leistungsverzeichnis, etwa 01.02.03.0010.
VOB	Vergabe- und Vertragsordnung für Bauleistungen, die Regeln für die Vergabe von Bauleistungen und die Verträge darüber.
X83	Die GAEB-Phase für die Angebotsaufforderung, vom Auftraggeber an die Bieter gesendet.

Quellen

56 Quellen, nummeriert nach ihrer ersten Nennung. Am 24. November 2026 hat ein Agent jede Quelle geöffnet und ein zweiter sie mit der Aussage abgeglichen. Der Status steht bei jedem Eintrag.

[1] Destatis: Mehr als jedes dritte Unternehmen nutzt 2026 künstliche Intelligenz. 2026-11-24. https://www.destatis.de/DE/Presse/Pressemitteilungen/2026/11/PD26_408_52911.html *teilweise geprüft*

[2] Bitkom: Erstmals nutzt Mehrheit Unternehmen KI. 2026-09-14. https://www.bitkom.org/Presse/Presseinformation/Erstmals-nutzt-die-Mehrheit-der-Unternehmen-KI *geprüft*

[3] BMWE, erstellt von Destatis: Vergabestatistik, Bericht für das Gesamtjahr 2024. 19.08.2026. https://www.bundeswirtschaftsministerium.de/Redaktion/DE/Publikationen/Wirtschaft/vergabestatistik-gesamtjahr-2024.pdf *geprüft*

[4] lxgesetze.de (nichtamtliche konsolidierte Fassung): VOB/A 2019, § 7. Stand laut Seite: „Zuletzt geändert am 25.9.2023“. https://lxgesetze.de/vob-a-2019/7 *geprüft, Sekundärquelle*

[5] Gründungsteam von iogentic: Evaluation report, run of 8 November 2026 (internal). 8. November 2026, abgerufen am 24.11.2026. interner Bericht, auf Anfrage erhältlich *eigene Quelle*

[6] Yao, Shinn, Razavi, Narasimhan: tau-bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains (arXiv:2406.12045). 2024-06-17. https://arxiv.org/abs/2406.12045 *geprüft*

[7] Rabanser, Kapoor, Kirgis, Liu, Utpala, Narayanan: Towards a Science of AI Agent Reliability (arXiv:2602.16666). 2026-02-18. https://arxiv.org/abs/2602.16666 *geprüft*

[8] Anthropic: Building effective agents. 2024-12-19. https://www.anthropic.com/engineering/building-effective-agents *teilweise geprüft*

[9] OpenAI: A practical guide to building agents. 2026-01-15. https://openai.com/business/guides-and-resources/a-practical-guide-to-building-ai-agents/ *geprüft*

[10] Europäische Kommission, AI Act Service Desk: AI Act Article 11. 2026-07-27. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-11 *geprüft*

[11] Europäische Kommission, AI Act Service Desk: AI Act Article 12. 2026-07-27. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-12 *geprüft*

[12] Europäische Kommission, AI Act Service Desk: AI Act Article 14. 2026-07-27. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-14 *geprüft*

[13] KOFA, IW Köln: KOFA Studie 04/2026, Wertvolle Erfahrung sichern: Ältere Beschäftigte im Unternehmen halten. 2026. https://www.kofa.de/media/Publikationen/Studien/KOFA-Studie_04-2026_AEltere_Beschaeftigte.pdf *geprüft*

[14] onvista (Agenturmeldung zum ifo): Ifo: Gut jedes zweite Unternehmen nutzt KI. 2026-06-05. https://www.onvista.de/news/2026/06-05-ifo-gut-jedes-zweite-unternehmen-nutzt-ki-in-der-breite-angekommen-0-20-26518961 *geprüft, Sekundärquelle*

[15] IAB (Institut für Arbeitsmarkt- und Berufsforschung), verbreitet über idw: IAB press release on generative AI use (IAB-Betriebspanel). 2026-05-05. https://idw-online.de/en/news870242 *geprüft*

[16] IW Köln, im Auftrag des Hauptverbands der Deutschen Bauindustrie: Gutachten: Bedeutung der Bauwirtschaft. 2024-06-04. https://iwkoeln.de/fileadmin/user_upload/Studien/Gutachten/PDF/2024/Gutachten_2024-Bedeutung-Bauwirtschaft.pdf *teilweise geprüft*

[17] Baulinks, Bericht über eine Umfrage von BIngK, VBI, BAK und AHO: Ingenieurbüros unter Druck: Längere Projektlaufzeiten und hoher Bürokratieaufwand. 2026. https://www.baulinks.de/webplugin/2026/1602.php4 *geprüft, Sekundärquelle*

[18] Europäische Kommission, AI Act Service Desk: AI Act Article 113 (consolidated as at 27 July 2026). 2026-07-27. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-113 *geprüft*

[19] Europäische Kommission: Data Act explained. abgerufen am 24.11.2026. https://digital-strategy.ec.europa.eu/en/factpages/data-act-explained *geprüft*

[20] Bundesanzeiger Verlag (recht.bund.de): Bundesgesetzblatt Teil I 2025 Nr. 301. 2025-12-05. https://www.recht.bund.de/bgbl/1/2025/301/VO.html *teilweise geprüft*

[21] BMJV (gesetze-im-internet.de): BetrVG §87. aktuelle Fassung. https://www.gesetze-im-internet.de/betrvg/__87.html *geprüft*

[22] BMJV (gesetze-im-internet.de): BetrVG §90. aktuelle Fassung. https://www.gesetze-im-internet.de/betrvg/__90.html *geprüft*

[23] DIW Berlin: Bauwirtschaft kehrt zurück auf Wachstumskurs, Wochenbericht 5/2026. 2026. https://www.diw.de/de/diw_01.c.996899.de/publikationen/wochenberichte/2026_05_1/bauwirtschaft_kehrt_zurueck_auf_wachstumskurs.html *teilweise geprüft*

[24] Bundesministerium der Justiz (gesetze-im-internet.de): § 34 HOAI. Fassung 2021, abgerufen am 24.11.2026. https://www.gesetze-im-internet.de/hoai_2013/__34.html *geprüft*

[25] Statistisches Bundesamt (Destatis): Betriebe, tätige Personen und Umsatz im Baugewerbe. Tabelle, Berichtsjahr 2024, abgerufen am 24.11.2026. https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Bauen/Tabellen/betriebe.html *geprüft*

[26] Rödl & Partner: Neue EU-Schwellenwerte ab 1.1.2026. 03.11.2025. https://www.roedl.com/insights/neue-eu-schwellenwerte-ab-ersten-ersten-zweitausendsechsundzwanzig/ *teilweise geprüft, Sekundärquelle*

Quellen, Fortsetzung

[27] Bundesministerium der Justiz (gesetze-im-internet.de): § 106 GWB Schwellenwerte. in Kraft, abgerufen am 24.11.2026. https://www.gesetze-im-internet.de/gwb/__106.html *geprüft*

[28] BMJV (gesetze-im-internet.de): GWB §121. aktuelle Fassung. https://www.gesetze-im-internet.de/gwb/__121.html *geprüft*

[29] Institut der deutschen Wirtschaft (IW): Kommune sucht Bauingenieur, IW-Kurzbericht Nr. 41. 17.06.2026. https://www.iwkoeln.de/studien/thomas-puls-kommune-sucht-bauingenieur.html *teilweise geprüft, Sekundärquelle*

[30] Senator für Wirtschaft, Arbeit und Häfen, Bremen: Auswertungsergebnisse der Kostenkontrolle, Vorlage 19/455-L, point 3. 2017-11-09. https://www.wirtschaft.bremen.de/sixcms/media.php/13/19_455_L-Vorlage%20Kostenkontrolle%20Bauprojekte_B.pdf *teilweise geprüft*

[31] baunetzwissen.de: Leistungsverzeichnis. undatiert, gelesen am 24.11.2026. https://www.baunetzwissen.de/drucken/leistungsverzeichnis-8494320 *teilweise geprüft, Sekundärquelle*

[32] tab.de (Herausgeber nicht geprüft): Flyer GAEB DA XML. undatiert. https://www.tab.de/download/554109/GAEB_DA_XML_Flyer.pdf *geprüft, Sekundärquelle*

[33] Flughafen München (Richtlinie des Betreibers): Richtlinie zur LV-Erstellung, section 8. 2010-03. https://Munich-Airport.De/_b/0000000000000000007428832bb5d8b282c/richtlinie-zur-lv-erstellung.pdf *geprüft, Sekundärquelle*

[34] bauprofessor.de: Neue Generation STLB-Bau. 2020-01-06. https://www.bauprofessor.de/news/neue-generation-stlb-bau/ *geprüft, Sekundärquelle*

[35] DIN Media: DIN 18299 VOB Teil C. 2023-09. https://www.dinmedia.de/de/norm/din-18299/369847346 *geprüft*

[36] DIN Media: DIN 276 Kosten im Bauwesen. Seite gelesen am 24.11.2026. https://www.dinmedia.de/de/norm/din-276/293154016 *geprüft*

[37] DIN Media: DIN 18290-2. 2023-11. https://www.dinmedia.de/de/norm/din-18290-2/369464914 *geprüft*

[38] GAEB: GAEB Datenaustausch. Seite gelesen am 24.11.2026. https://www.gaeb.de/de/produkte/gaeb-datenaustausch/ *geprüft*

[39] DIN (Veranstaltung „BIM nutzt GAEB"): BIM nutzt GAEB, Vortrag Datenaustausch. 2017-05-12. https://www.din.de/resource/blob/236706/d25969398ae4db7247ed8a5118ffb2ee/top04-bim-nutzt-gaeb-2017-05-12-vortrag-gaeb-datenaustausch-rheinberg-data.pdf *geprüft, Sekundärquelle*

[40] Niedersächsisches Landesamt für Bau und Liegenschaften: AVA-Pflichtenheft Version 2024-01, section 2.1. 2024-01. https://nlbl.niedersachsen.de/download/208208/AVA-Pflichtenheft_des_SBN_Version_2024-01.pdf *teilweise geprüft*

[41] JV Roig: How Much Do LLMs Hallucinate in Document Q&A Scenarios? (arXiv:2603.08274). 2026-03-09. https://arxiv.org/abs/2603.08274 *geprüft*

[42] Zheng u. a.: Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena (arXiv:2306.05685). 2023-06-09. https://arxiv.org/abs/2306.05685 *geprüft*

[43] Norman, Rivera, Hughes: Reliability without Validity (arXiv:2606.19544). 2026-06-17. https://arxiv.org/abs/2606.19544 *geprüft*

[44] Lewis u. a.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (arXiv:2005.11401). 2020-05-22. https://arxiv.org/abs/2005.11401 *geprüft*

[45] Anthropic: Introducing Contextual Retrieval. 2024-09-19. https://www.anthropic.com/news/contextual-retrieval *geprüft*

[46] Google Cloud: A methodical approach to agent evaluation. 2025-11-18. https://cloud.google.com/blog/topics/developers-practitioners/a-methodical-approach-to-agent-evaluation *geprüft*

[47] Anthropic: Introducing the Model Context Protocol. 2024-11-25. https://www.anthropic.com/news/model-context-protocol *geprüft*

[48] Projekt Model Context Protocol: Model Context Protocol Specification. 2026-07-28. https://modelcontextprotocol.io/specification/latest *geprüft*

[49] iogentic: Worked example of the tender package. 24. November 2026. https://iogentic.ai/os *eigene Quelle*

[50] Europäische Kommission, AI Act Service Desk: AI Act Article 9. 2026-07-27. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-9 *geprüft*

[51] EEAS (Spiegelung einer Mitteilung des Rates): Artificial Intelligence: Council gives final green light to simplify and streamline rules. 2026-06-30. https://eeas.europa.eu/delegations/chile/artificial-intelligence-council-gives-final-green-light-simplify-and-streamline-rules_en *geprüft*

[52] gdpr-info.eu (nichtamtlicher Text): GDPR Article 22. k. A. https://gdpr-info.eu/art-22-gdpr/ *geprüft, Sekundärquelle*

Quellen, Fortsetzung

[53] BMJV (gesetze-im-internet.de): BSIG §28. 2025. https://www.gesetze-im-internet.de/bsig_2025/__28.html *geprüft*

[54] IEC Webstore: ISO/IEC 42001:2023. 2023-12-18. https://webstore.iec.ch/publication/90574 *geprüft*

[55] BSI: Aktualisierte Fassung veröffentlicht: Chancen und Risiken generativer KI-Modelle. 2025-01-21. https://www.bsi.bund.de/DE/Service-Navi/Presse/Alle-Meldungen-News/Meldungen/Chancen-Risiken-generative-KI_250121.html *geprüft*

[56] BMJV (gesetze-im-internet.de): BetrVG §80. aktuelle Fassung. https://www.gesetze-im-internet.de/betrvg/__80.html *geprüft*



iogentic

Kontakt

Wir bieten ein Gespräch von dreißig Minuten ohne Folien an. Schildern Sie uns eine Aufgabe aus Ihrer Organisation, und wir sehen uns gemeinsam an, ob ein Agent dabei helfen könnte.

hello@funkeltronic.ai

Whitepaper, November 2026. Die Plattform ist nicht veröffentlicht. Alle Fotos sind KI-generierte Bilder. Das Fallbeispiel und der Evaluierungslauf nutzen synthetische Projekte. Die Quellen wurden am 24. November 2026 gelesen.