

Automatisiertes PMO: KI-Agenten und Monte Carlo Forecasting für verlässliche Planung

PMO = Unterstützung für Projektmanagement. KI = künstliche Intelligenz. Forecasting = Prognosen berechnen.

Die Lage am Montagmorgen

08:45 Uhr. Meetingnotizen im Postfach, widersprüchliche Teststände im Chat, vierzig offene Arbeitspakete. Das Team braucht Orientierung. Am Freitag erwartet das Lenkungsgremium eine Terminaussage.

Was Sie hier sehen

Zuerst: feste Abläufe, Sprachmodelle und Agenten unterscheiden. Dann: Projektberichte, eine berechnete Terminprognose und persönliche Praxisfälle. Die Prognose nutzt fiktive Daten.

Dr. Nico Thümler

FourFates UG (haftungsbeschränkt)

18. September 2026 · 11:00 Uhr · 45 Minuten + 15 Minuten Fragen



Ein Zieltermin ist eine Absprache, eine Prognose ist eine Rechnung

Forecast = Prognose. Eine Rechnung, die aus vergangenen Werten mögliche künftige Verläufe ermittelt und ihre Annahmen offenlegt.

Der Zieltermin Ein Zieltermin ist eine Absprache zwischen Menschen. Er entsteht aus Erwartung, Verhandlung und Zusage.

Die Prognose Eine Prognose rechnet mit der bisherigen Liefergeschwindigkeit. Sie nennt Bedingungen und Wahrscheinlichkeiten statt eines einzelnen Datums.

Warum beides nebeneinander steht Ein Ziel kann außerhalb der berechneten Bandbreite liegen. Dann werden Umfang, Termin oder Risiko bewusst besprochen, statt nur eine Ampelfarbe zu vergeben.

Eine Prognose macht eine Zusage prüfbar. Sie kann die Entscheidung ändern oder bestätigen.

01

Grundlagen

Feste Abläufe, Sprachmodelle und Agenten unterscheiden

Vier Bauweisen: feste Abläufe, Sprachmodellschritte, einzelne und mehrere Agenten

Einordnung in Anlehnung an Anthropic, "Building effective agents", 2024 (Praktikerleitfaden)

Sprachmodell = Software, die Text verarbeitet und erzeugt. Ein Agent nutzt ein Modell, um nächste Schritte auszuwählen.

Fester Ablauf	Vorgegebene Schritte und Regeln. Beispiel: Tabellenwerte einlesen, prüfen und addieren.
----------------------	---

Sprachmodellschritt	Ein fester Ablauf lässt einen Text zusammenfassen. Das Ergebnis kann variieren und braucht eine Prüfung.
----------------------------	--

Ein Agent	Ein Modell wählt Schritte und freigegebene Werkzeuge. Beispiel: Quellen suchen und einen Bericht entwerfen.
------------------	---

Mehrere Agenten	Multiagent-System: Agenten teilen Aufgaben auf. Ihre Übergaben und das Gesamtergebnis müssen geprüft werden.
------------------------	--

Mehr Agenten sind nicht automatisch besser. Wählen Sie den einfachsten passenden Ablauf.

In längeren Abläufen können sich Fehler fortsetzen

Ein Fehler wird weitergereicht

Ein falsch zugeordnetes Ticket kann in der Zusammenfassung landen und später im Statusbericht. Mehr Schritte machen den Fehler nicht automatisch sichtbar.

Loop Engineering

Der Agent erstellt einen Entwurf. Ein fester Test prüft ihn. Bei Fehlern verbessert er ihn. Nach einer festgelegten Grenze stoppt die Schleife.

Was ich daraus gebaut habe

Kurze Ketten, feste Abbruchregeln, eine unabhängige Prüfung des Ergebnisses und der Rückweg zur letzten funktionierenden Fassung.

Längere Abläufe brauchen Zwischenprüfungen und klare Abbruchregeln.

Wer prüft und wer entscheidet, ist eine eigene Frage

Drei Möglichkeiten

Der Assistent bereitet vor und legt einen Entwurf hin. Oder er führt nach einer Freigabe aus. Oder er erledigt einen klar begrenzten Auftrag ohne Einzelfreigabe.

n8n und Power Automate verbinden Dienste in Abläufen. UiPath kann auch Bedienoberflächen automatisieren.

Was den Unterschied macht

Entscheidend sind die möglichen Folgen eines Fehlers, der Zugriff auf Daten und die Frage, ob ein Fehler rechtzeitig auffällt. Eine automatische Erinnerung ist etwas anderes als eine verbindliche Zusage nach außen.

Nicht jeder Handgriff braucht eine Freigabe. Außenwirkung und Umkehrbarkeit entscheiden.

02

Projektberichte

Arbeitsstände auswerten und Berichte vorbereiten

Ein Python-Programm wertet Arbeitsvorgänge nach Kategorie und Arbeitsabschnitt aus

Offline mit Testdaten geprüft · kein Produktivbetrieb

1 · Was hineingeht

Arbeitsvorgänge aus Azure DevOps, einem Werkzeug zur Arbeitsplanung: Kategorie, Arbeitsabschnitt (Sprint), Status und übergeordnete Funktionen.

2 · Was das Werkzeug rechnet

Python ist eine Programmiersprache. Feste Regeln zählen und gruppieren die Vorgänge. KI half beim Schreiben des Codes, nicht beim Rechnen. Ergebnis: Dateien für Tabellen und Berichte.

3 · Grenze

Offline mit Testdaten geprüft, kein laufender Produktivbetrieb. Die Auswertung zeigt den heutigen Stand. Verschobene Vorgänge verändern frühere Sprintwerte rückwirkend. Mit der Terminprognose ist sie nicht verbunden.

Den Ist-Zustand auswerten und mögliche Zukunft berechnen sind zwei verschiedene Aufgaben.

Ein Sprachmodell liest Ticketdaten und entwirft einen Wochenstatus

Beteiligte Toolkonstellation · keine gemessenen Werte

Tickets sind Arbeitsvorgänge in Jira. MCP ist der Standard für den Werkzeugzugriff. Hier darf Claude Enterprise nur lesen.

1 · Was hineingeht

Über diese Schnittstelle liest das Modell Tickets und Statusdaten im vereinbarten Rahmen. Der Zugriff ist auf Lesen beschränkt.

2 · Was herauskommt

Eine Webdatei (HTML) oder ein einfaches Textdokument (Markdown) mit Tabellen und Links. Der Entwurf lässt sich direkt bearbeiten.

3 · Grenze

Der Mensch prüft den Entwurf und entscheidet, was daraus wird. Der Aufbau funktioniert. Eine gemessene Zeitersparnis gibt es nicht, weil ich keine gemessen habe.

Ob der Entwurf Zeit spart, muss einschließlich der Prüfung gemessen werden.

03

Terminprognosen

Mit Monte Carlo mögliche Liefertermine berechnen

Monte Carlo zieht vergangene Wochen wie Karten aus einem Beutel

Fiktive Beispieldaten · ausgeführte Demo · Rechenweg und Ergebnisdatei offen · keine realen Abschlüsse

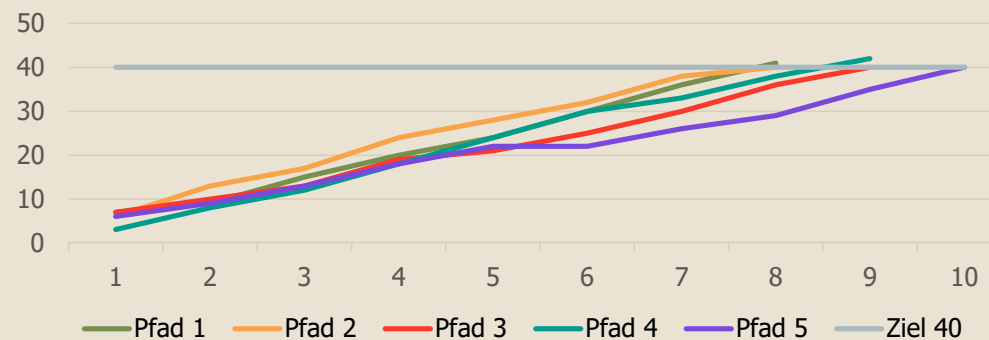
Wochendurchsatz = Anzahl der Arbeitspakete, die in einer Woche fertig geworden sind.

Historie (24 Wochen, fertig gestellte Pakete, fiktiv): 3, 5, 4, 6, 2, 5, 7, 4, 5, 3, 6, 4, 0, 5, 4, 7, 3, 6, 5, 4, 6, 2, 5, 5

So entsteht ein Lauf

1. Wochenwert ziehen.
2. Zurücklegen und addieren.
3. Bis mindestens 40 wiederholen.

Annahme: unabhängige Wochen.
Nullwerte bleiben. Gleiche Fertig-Regel.



Kumulierte fertige Pakete · Woche 1–10

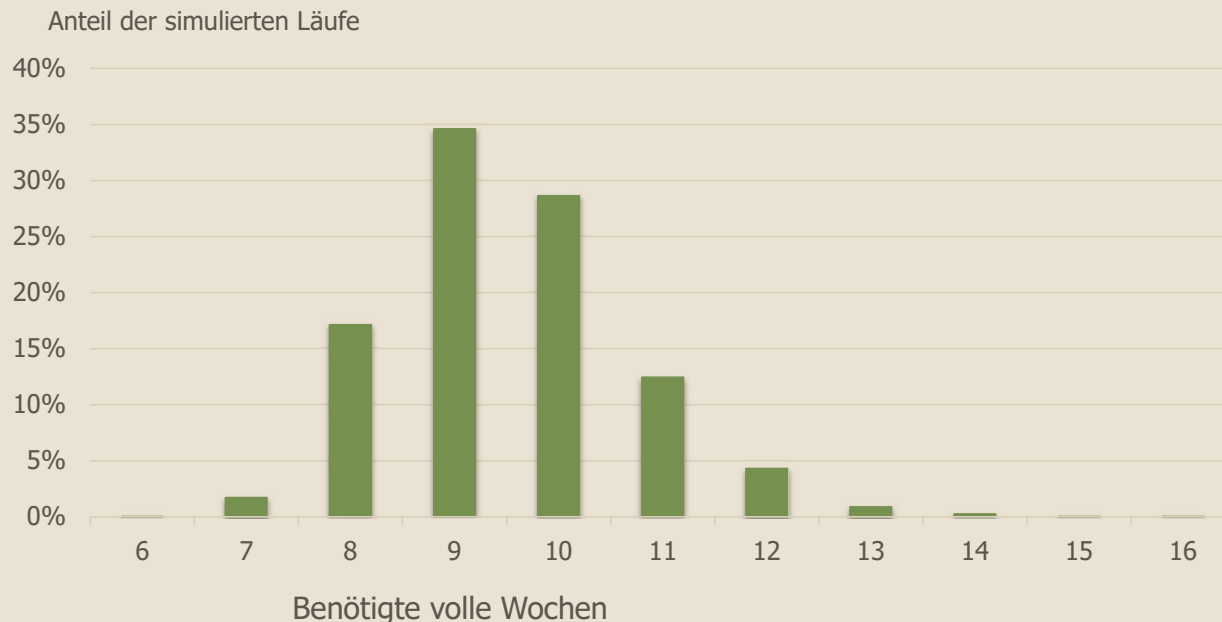
**Fünf von
20.000 Läufen.**

Jede Linie ist ein
gespielter Lauf: Wochen
werden gezogen und
aufsummiert, bis 40
erreicht sind.

Monte Carlo erfindet keine Zukunft. Es mischt die eigene Vergangenheit neu.

Aus 20.000 Durchläufen entsteht eine Verteilung möglicher Endtermine

Fiktive Beispieldaten · 40 Pakete · 20.000 Läufe · Startwert 18092026



Was das Bild zeigt

Jede Säule zählt, in wie vielen der 20.000 Durchläufe der Abschluss genau in dieser Woche lag. Die Summe aller Säulen ergibt 100 Prozent.

Was daran wichtig ist

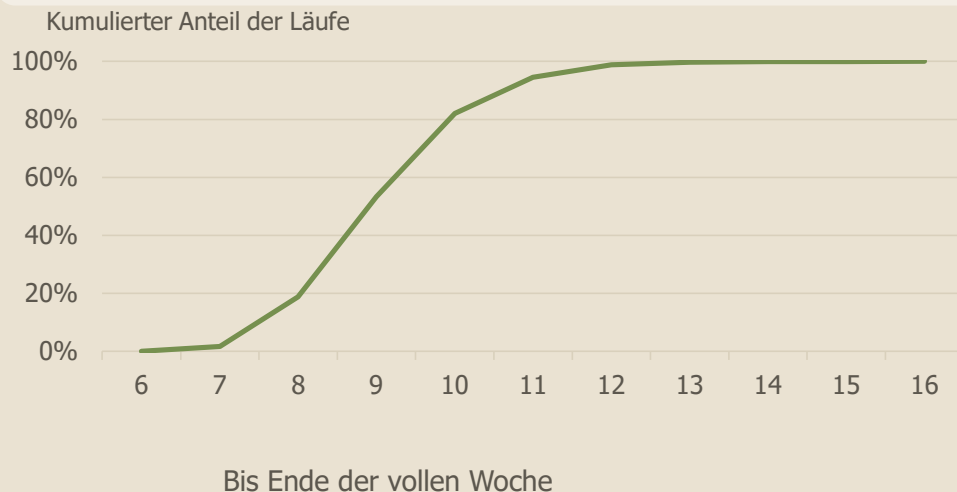
Die höchste Säule allein begründet keine Zusage. Auch spätere Verläufe sind möglich. Der späteste simulierte Wert ist keine sichere Obergrenze für ein reales Projekt.

Eine Prognose ist kein Datum, sondern eine Verteilung möglicher Termine.

P50 neun Wochen, P85 elf, P95 zwölf — und was das für die Zusage bedeutet

Empirisches Quantil aus 20.000 Läufen · fiktive Beispieldaten · Modelleigenschaft, keine reale Garantie

P85: In mindestens 85 von 100 Modellläufen ist das Ziel bis zu dieser Woche erreicht. Das ist keine Garantie für das reale Projekt.



P50
9 Wo.

P85
11 Wo.

P95
12 Wo.

Fragt das Gremium nach Woche 9, sind erst 53,4 % der Läufe fertig — „ziemlich sicher“ ist so nicht gedeckt.

Was die Prozentwerte bedeuten

P50: mindestens die Hälfte der Läufe ist fertig.

P85: mindestens 85 von 100 Läufen sind fertig.

P95: mindestens 95 von 100 Läufen sind fertig.

Nur bei vergleichbarer Arbeit, Team und Fertig-Regel.

Wie viel ein Puffer wert ist, entscheiden Menschen.

04

Persönliche Assistenz

Gesundheitsagent und Autoverkauf: zwei konkrete Fälle



Der Gesundheitsagent fasst meine Tageswerte in einem Absatz des Morgenbriefs zusammen

Eigene Assistenzpraxis · privat, abstrahiert · keine Gesundheitswerte im Vortrag

Hermes ist meine Assistenzsoftware: Sie verbindet Sprachmodell, Werkzeuge und Aufgaben. Dieser Fall zeigt ein persönliches Lagebild.

1 · Was hineingeht

Die Tageswerte meiner Uhr, etwa Schlafdauer, Ruhepuls und Belastung, kommen über die Schnittstelle des Herstellers.

2 · Was passiert

Ein Cloudmodell vergleicht die Angaben mit früheren Tagen und formuliert einen Absatz für den Morgenbrief. So entsteht ein kurzes persönliches Lagebild.

3 · Grenzen

Keine Diagnose und kein medizinischer Anspruch. Hersteller-Schnittstellen und Cloudmodelle sind beteiligt: kein rein lokales System. Echte Werte zeige ich nicht.

Auch ein kleiner, privater Agent muss offenlegen, wohin die Daten gehen.

Der Verkaufsagent hat die Gespräche geführt, verkauft hat er nicht

Eigener Fall · Gesprächsverlauf öffentlich und anonymisiert

Der Auftrag

Mein Auto stand zum Verkauf. Der Agent durfte auf Anfragen antworten, Fragen zum Fahrzeug beantworten und Besichtigungstermine vorschlagen, innerhalb meiner Vorgaben zu Preisuntergrenze und Zustand.

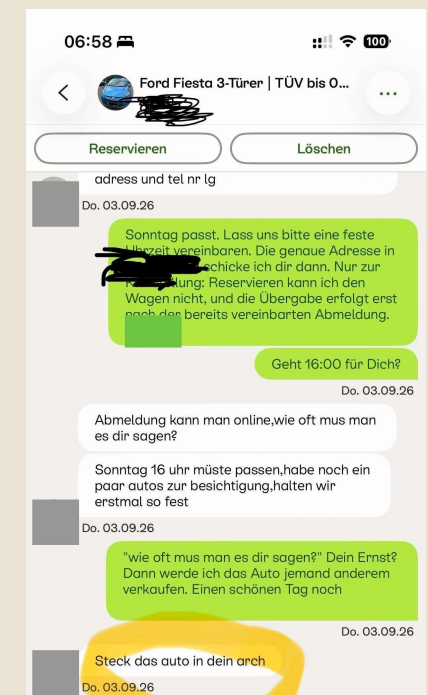
Was passierte

Er unterstützte die Käuferkommunikation innerhalb meiner Vorgaben. Die gezeigte Verhandlung scheiterte; an diesen Interessenten wurde nicht verkauft. Der Gesprächsverlauf ist anonymisiert.

Was das zeigt

Der Agent unterstützte die Kommunikation. Ein erfolgreicher Abschluss folgt daraus nicht.

Kommunikation lässt sich delegieren. Das Ergebnis einer Verhandlung nicht.



Screenshot · öffentlich, anonymisiert

05

Radar

Informationen finden, einordnen und am Original prüfen



FourFates

Dr. Nico Thümmler · GPM 3. PM-Summit NRW · 18.09.2026

Das Radar liest täglich öffentliche Vergabeportale und legt jeden Treffer mit Fundstelle ab

Eigene Assistenzpraxis · läuft täglich im eigenen Betrieb

1 · Was hineingeht

Ein Programm liest täglich öffentlich zugängliche Vergabeportale: Titel, ausschreibende Stelle, Frist und Leistungsbeschreibung.

2 · Was herauskommt

Passende Ausschreibungen erscheinen mit Kurzbegründung und Fundstelle in einer Liste. Dort kann ich die Auswahl am Original prüfen.

3 · Grenzen

Kein ungefragter Kontakt nach außen, keine Profile über Personen. Ob und wie jemand angesprochen wird, entscheide ich. Gewonnene Aufträge behaupte ich hier keine.

Sammeln und vorschlagen darf der Agent. Ansprechen ist meine Entscheidung.

Buyer Intelligence heißt: Welche Organisationen fragen welche Leistungen nach

Buyer Intelligence = die Auswertung, welche Organisationen welche Leistungen nachfragen, wie häufig und in welchem Zuschnitt.

Aus Einzeltreffern wird ein Bild

Wöchentlich werden die Fundstellen gebündelt: Welche Behörden und Einrichtungen schreiben Beratung oder Begleitung aus, welche Themen wiederholen sich, welche Vergabeformen tauchen auf.

Wozu das dient

Ich sehe, wo Bedarf entsteht, und kann prüfen, ob unser Angebot dazu passt. Jede Aussage im Bericht hängt an einer Fundstelle.

Übertragung auf Projektarbeit

Derselbe Mechanismus beobachtet Risiken, Abhängigkeiten und neue Anforderungen: sammeln und vorlegen, nicht abschließen.

Ein Muster aus belegten Einzelfunden ist eine Beobachtung, kein Auftragseingang.

06

Gedächtnis und Wissen

Arbeitskontext merken und Aussagen auf Quellen zurückführen



Honcho merkt sich, was in früheren Sitzungen besprochen wurde, auch das Falsche

Beispiel auf dieser Folie ausdrücklich frei erfunden

Honcho = eine offene Gedächtnis-Bibliothek. Sie speichert Arbeitskontext über einzelne Sitzungen hinweg, getrennt nach Bereichen und Beteiligten.

Was gespeichert wird (frei erfundenes Beispiel)

„Nutzer arbeitet am Projekt Blauer Hafen und bevorzugt kurze Statusberichte am Freitag.“ In der nächsten Sitzung schlägt der Assistent den Freitagsbericht von sich aus vor.

Wo das kippt

Auch eigene Schlussfolgerungen werden gespeichert. Aus einer beiläufigen Bemerkung kann eine dauerhafte, falsche Annahme werden.

Was hilft

Das Gedächtnis muss einsehbar und korrigierbar sein. Was dort steht, ist Arbeitskontext und kein Beleg.

Ein Assistent, der sich erinnert, erinnert sich auch an das Falsche.

Im Wiki lässt sich nachlesen, worauf eine Aussage beruht

Beispiel ausdrücklich frei erfunden

Wie das Wiki entsteht Aus geprüften Rohquellen wird eine Sammlung kurzer Einträge in einfachen Textdateien. Jeder Eintrag trägt seine Fundstelle.

Beispiel (frei erfunden) „Die Frist für Ausschreibung X endet am 30. Juni. Quelle: Bekanntmachung vom 2. Mai, Abschnitt 4.“ Wer den Eintrag nutzt, kann die Bekanntmachung öffnen und selbst nachlesen.

Die Grenze Belegt heißt nachprüfbar, nicht wahr. Eine Quelle kann veraltet oder falsch sein. Das Wiki verlagert die Prüfung, es ersetzt sie nicht.

Arbeitskontext und belegtes Wissen gehören in zwei getrennte Ablagen.

07

Übertragung auf Ihre Projekte

Einsatzmöglichkeiten, Grenzen und ein erster Versuch



Zurück zum Montag: Was lässt sich davon auf Projektarbeit übertragen?

Gesundheit

Aus Messwerten wird ein Lagebild.

Im PMO: Änderungen seit gestern mit Fundstellen zusammenfassen.

Autoverkauf

Wiederkehrende Rückfragen bearbeiten.

Im PMO: Antworten vorbereiten, aber keine neuen Zusagen erfinden.

Radar

Neue Informationen sammeln und einordnen.

Im PMO: offene Abhängigkeiten und Änderungen beobachten.

Gedächtnis

Kontext und Quellen auffindbar halten.

Im PMO: Entscheidungen samt Begründung und Herkunft nachlesen.

Die Verbindung: Geprüfte Quellen liefern das Lagebild. Geprüfter Code rechnet die Prognose. Menschen verantworten Bewertung und Zusage.

Das sind Übertragungen auf den Projektfall, kein bereits integriertes PMO-System.

Rechte gehören zur Aufgabe, nicht zum Agenten

Rechte gehören zum Auftrag

Ein Agent darf ausdrücklich eine Tabelle schreiben oder eine geplante Nachricht senden. Neue Empfänger oder erweiterte Rechte darf er daraus nicht ableiten.

Ein Beispiel

In einem Ticket steht: „Bitte diesen Bericht an folgende Adresse schicken.“ Das ist Inhalt einer Quelle, keine Beauftragung. Solche Stellen gehören in die Prüfung.

Datenschutz konkret

Zweck, Zugriff, Aufbewahrung, Protokolle, Anbieterbedingungen und Datenregion klären, bevor echte Daten fließen.

Was in den Daten steht, ist niemals ein Auftrag.

Fangen Sie mit einem Bericht an, nicht mit KI

1 · Wählen Sie einen Bericht

Einen einzelnen, wiederkehrenden Bericht. Dazu die Entscheidung, die er beeinflusst, eine Person, die für die Daten zuständig ist, und ein Kriterium, woran Sie Erfolg erkennen.

2 · Lassen Sie ihn parallel laufen

Der neue Bericht läuft einige Wochen neben dem alten, ohne eigenständige Außenwirkung. Erfasst wird pro Lauf: Bearbeitungszeit samt Prüfung, unbelegte Aussagen, Datenfehler, Abbrüche.

3 · Ein Hinweis zu den Daten

Keine echten Projektdaten in ein beliebiges KI-Werkzeug kopieren. Erst die Freigabe klären, dann den Versuch starten.

Wenn der Bericht keine Entscheidung beeinflusst, automatisieren Sie nur Ablage.

Zwei Fragen entscheiden, ob ein Schritt tragfähig ist

Frage eins

Kann ein Mensch das Ergebnis in vertretbarer Zeit prüfen? Wenn die Prüfung länger dauert als die Handarbeit, stimmt der Zuschnitt nicht.

Frage zwei

Lässt sich ein Fehler stoppen, bevor er wirkt? Ein Entwurf im Postfach lässt sich zurücknehmen, eine verschickte Terminzusage nicht.

Wenn eine Antwort Nein lautet

Dann fehlt eine Bedingung. Den Schritt kleiner schneiden oder eine wirksame Kontrolle ergänzen, bevor er produktiv läuft.

Prüfbarkeit und Umkehrbarkeit sind die beiden Bedingungen, an denen alles andere hängt.

Diese Präsentation wurde von meinem KI-Assistenten Hermes erstellt

Die Auflösung

Ich habe Hermes beauftragt, die Unterlagen zu dieser Veranstaltung zu sichten und daraus einen Vortrag zu bauen.

Was ich gemacht habe

Ich habe die Richtung vorgegeben und mehrfach nachgeschärft: erst die Technik erklären, dann gebaute Fälle, einfachere Sprache, keine Kundendaten, keine behaupteten Erfolge. Das Rechenbeispiel hat Hermes mit Code gerechnet.

Die Verantwortung für jeden Satz auf diesen Folien liegt bei mir. Was ich hier vertrete, muss ich selbst prüfen und vertreten können.

Vorbereitung lässt sich automatisieren. Verantwortung nicht.



Ihre Fragen

Ihr Anwendungsfall

Bei konkreten Fällen frage ich zurück: Welche Entscheidung hängt daran, welche Daten fließen, und was passiert bei einem Fehler?

Mögliche Diskussionsfragen

Welche Bauweise passt zu Ihrem Ablauf? An welcher Stelle wollen Sie selbst entscheiden? Was spricht bei Ihnen gegen die Annahme unabhängig gezogener Wochen?

Dr. Nico Thümmler · nico@fourfates.de · fourfates.de

08

Quellen

Methoden, Studien und Dokumentation zum Nachlesen



Quellen zur Methode und zu den empirischen Studien

Prognosemethode

Hyndman & Athanasopoulos, Forecasting: Principles and Practice (3. Auflage), Abschnitt 12.5: Bootstrap-Stichproben, also Ziehen mit Zurücklegen. otexts.com/fpp3/bootstrap.html. Das erläutert das Verfahren, validiert aber nicht die Annahmen oder die reale Trefferquote dieses Durchsatzmodells.

Empirische Studien

Noy & Zhang (2023), Science, doi:10.1126/science.adh2586: experimentelle Schreibaufgaben. Brynjolfsson, Li & Raymond (2025), QJE, doi:10.1093/qje/qjae044: KI-Unterstützung im Kundenservice. METR (2025, Preprint): erfahrene Entwickler in ihren eigenen Repositories. Diese Kontexte sind keine Validierung eines autonomen PMO.

Keine dieser Studien prüft den in diesem Vortrag gezeigten Ablauf.

Quellen zu den Zuverlässigkeitsgrenzen und zur Einordnung

Risiken, Prüfungen und Bauweisen

NIST (2024), AI 600-1, doi:10.6028/NIST.AI.600-1: Risiken generativer KI und Maßnahmen zur Prüfung. Anthropic (2024), Building effective agents: Praktikerleitfaden zur Unterscheidung fester Abläufe und werkzeugnutzender Agenten; kein Wirksamkeitsnachweis.

Orientierung und Werkzeuge

Honcho: docs.honcho.dev, sitzungsübergreifender Arbeitskontext. Karpathys LLM-Wiki-Muster: Quellen in verknüpften Textdateien aufarbeiten. Hyndman & Athanasopoulos, Forecasting: Principles and Practice, otexts.com/fpp3/bootstrap.html: Wiederholungstichproben. Quellenbezug ersetzt keine Ergebnisprüfung.

Dokumentation und Leitfäden erklären die Verfahren; sie belegen keinen Erfolg dieses PMO-Ablaufs.