



DEIN BEGLEITMATERIAL

Die KI spart Zeit. Wer bezahlt die Rechnung?

Modellrechnung · Prompt Caching · Notion

Fünf Handgriffe und Dein Wochenversuch

Studioporträt und Szenen mit KI erzeugt. Bildfassung vom 07.10.2026.

DIE GANZE RECHNUNG IN 60 SEKUNDEN

DER ERSTE ENTWURF IST SCHNELL.

FERTIG IST ER ERST NACH DER PRÜFUNG.

01 Was ist wirklich fertig?

Miss den gesamten Weg: Vorbereitung, Eingabe, Kontrolle, Nacharbeit und misslungene Versuche. Lege vorher fest, was ein brauchbares Ergebnis ist.

02 Was kostet der ganze Weg?

Verbrauch, Software, Einrichtung und eigene Zeit zählen zusammen. Ein niedriger Tokenpreis allein sagt wenig über die Kosten einer erledigten Aufgabe.

03 Was machen wir mit der Zeit?

Zeitgewinn, bewertete Kapazität und Geld auf dem Konto sind verschiedene Größen. Entlastung kann wertvoll sein, ohne sofort zusätzliche Einnahmen zu erzeugen.

DER DEUTSCHE AUFHÄNGER

Die Telekom plant laut WDR vom 5. Oktober KI-bedingte Entlastungen von 1,1 Milliarden Euro für 2027. Die zusätzlichen KI-Ausgaben beziffert der Bericht nicht. Das ist ein angekündigtes Ziel, keine nachgewiesene Nettoersparnis. Genau diese Unterscheidung nehmen wir mit an den Bürotisch.

60 VERMERKE. WAS BLEIBT ÜBRIG?

Erfundenes Beispiel für einen Monat. Keine Produktpreise, Marktmittelwerte oder Prognose für Deinen Betrieb.

ZEIT JE VORGANG

OHNE KI: 20 MINUTEN

3 MIN.

5 MIN.

2

MIT KI: 10 MINUTEN

3 Minuten Eingabe + 5 Minuten Kontrolle + 2 Minuten Nacharbeit

60 × 10 Minuten Zeitgewinn	10 Stunden
10 Stunden × 40 Euro bewertete Kapazität	400 Euro
Software und Verbrauch im Monat	–150 Euro
Einrichtung: 300 Euro / 3 Monate	–100 Euro

RECHNERISCHER SALDO

150 EURO

Kein automatisch entstehender Geldgewinn. Erst die Verwendung der freien Zeit entscheidet über zusätzliche Einnahmen, vermiedene Kosten oder bewusst gewählte Entlastung.

DIE NUTZENSCHWELLE

$250 / 40 = 6,25$ Stunden im Monat. Bei 60 Vorgängen sind das **6 Minuten 15 Sekunden je Fall**. Darunter deckt der bewertete Zeitgewinn die Zusatzkosten nicht. Qualität und Fehlerrisiken müssen zusätzlich passen.

Die Verteilung der Einrichtung ist eine Rechenannahme, keine steuerliche Abschreibung. Ab Monat vier entfällt sie nur ohne weitere Einrichtungskosten.

FÜNF HANDGRIFFE. KEIN TOKEN-YOGA.

01 Ergebnis festlegen

Zweck, Zielgruppe, Länge und Qualität bestimmen. Fehlende Angaben ausdrücklich markieren lassen.

02 Passenden Kontext geben

Relevante Auszüge statt des ganzen Archivs. Wichtige Einschränkungen und Ausnahmen erhalten.

03 Ausgabe begrenzen

Das benötigte Format bestellen. Weniger Leseaufwand hilft auch beim Festpreisabo.

04 Leistung passend wählen

Modelle an denselben typischen Fällen prüfen. Qualität, gesamte Kosten und Korrekturzeit vergleichen.

05 Schleifen begrenzen

Überarbeitungen und Werkzeugaufrufe deckeln. Ein Wunsch im Prompt ersetzt keine technische Geldgrenze.

KOPIERBARER AUFTRAG

Erstelle aus dem freigegebenen Text einen Gesprächsvermerk mit höchstens fünf Punkten. Nenne Entscheidungen, zuständige Personen, Termine und offene Fragen. Nutze nur Informationen aus dem Text. Markiere fehlende Angaben als „nicht genannt“. Benenne Widersprüche. Ergänze keine Vermutungen. Liefere nur den Vermerk. Text: [hier einfügen]

Danach prüfen: Stimmen Inhalt, Zuständigkeiten und Termine mit der Quelle überein?

DIE WERKBANK BLEIBT. DER AUFTRAG WECHSELT.

Wiederkehrende, identische Eingabeteile können bei geeigneten Modelldiensten zeitweise wiederverwendet werden. Die Antwort entsteht trotzdem neu.

EINE ANFRAGE, ZWEI TEILE

FESTE REGELN + VORLAGE

NEUER GESPRÄCHSTEXT

Identischer Anfang, wenn der Dienst es unterstützt

01

Gleich heißt wirklich gleich

Der relevante Anfang muss übereinstimmen. Nur sinngleiche Formulierungen reichen nicht. Wechselnde Angaben gehören hinter den festen Teil.

02

Treffer und erste Schreibkosten messen

Normale Eingaben, Cache-Lesen, gegebenenfalls Cache-Schreiben und Ausgaben getrennt erfassen. Den ersten Lauf und die Wiederholungen gemeinsam vergleichen.

NIKOS PRÜFFRAGE

Lohnt sich der Aufwand bei meiner tatsächlichen Nutzung? Mindestlänge, Lebensdauer, Modellunterstützung und Tarife hängen vom Dienst ab. Ein Treffer ist nicht garantiert. Ein Festpreisabo wird dadurch nicht automatisch günstiger. Rechne auch Deine Einrichtungszeit mit.

ORDNUNG, DIE ARBEIT SPART.

Notion verbindet Wissensseiten mit Datenbanken, deren Einträge sich filtern und sortieren lassen. Für den Versuch genügt eine kleine Struktur. Eine vorhandene Tabelle kann ebenso reichen.

01

VORLAGE

Freigegebener Auftrag, Quelle und Stand.

02

FÄLLE

Zeit, Kosten, Fehler und Prüfergebnis.

03

VERANTWORTUNG

Wer pflegt? Wer prüft? Wann erneut?

NOTION IST NICHT PROMPT CACHING.

Notion organisiert Wissen. Caching betrifft die Verarbeitung beim Modelldienst. Bei einer KI-Anbindung kontrollieren: Welche Seiten oder Abschnitte werden tatsächlich übertragen? Danach den Eingabeverbrauch prüfen.

DREI KLARE TIPPS VON NIKO

1

Durch brauchbare Ergebnisse teilen

20 Anläufe, 10 gute Vermerke: gesamte Verbrauchskosten durch 10 teilen. Eigene Arbeitszeit zusätzlich notieren.

2

Den teuersten Fall ansehen

Neben dem üblichen Verbrauch den Ausreißer untersuchen. Ein günstiger Lauf belegt keinen stabil günstigen Ablauf.

3

Zeit sofort eintragen

Suchen, Prüfung und Nacharbeit direkt festhalten. Nicht am Freitag aus dem Gedächtnis schätzen.

EINE AUFGABE. EINE EHRliche WOCHE.

MO

Maßstab setzen

Aufgabe, erlaubte Daten und Qualität festlegen. Bisherige Dauer messen.

DI

Klein einrichten

Vorlage bauen. Aufwand erfassen. Normalen und schwierigen Fall testen.

MI / DO

Fälle sammeln

Alle Versuche zählen, auch Abbrüche. Eingabe, Prüfung und Nacharbeit messen.

FR

Gemeinsam entscheiden

Beibehalten, ändern, länger testen oder stoppen. Seltene Fehler bleiben offen.

UNSERE AUFGABE:

Fall	Vorher	Mit KI*	Kosten	Geprüft?	Fehler / Nacharbeit
1					
2					
3					
4					
5					

* Minuten gesamt bis zum geprüften Ergebnis. Direkte Kosten einschließlich Fehlversuchen.

DIE GEWONNENE ZEIT VERWENDEN WIR FÜR:

ENTSCHEIDUNG + NÄCHSTER PRÜFTERMIN:

BELEGE STATT SPARFOLIEN.

01 / WDR, 5. Oktober 2026

Telekom-Planwerte für das Geschäft außerhalb der USA, Vergleich mit 2023: 1,1 Mrd. Euro für 2027; 2,5 Mrd. bei indirekten Kosten für 2030. Zusätzliche KI-Ausgaben im Bericht nicht beziffert.

02 / Wall Street Journal, 4. Oktober 2026

Spending on AI Is Becoming Almost Impossible for Businesses to Budget. Journalistischer Kostenanlass; möglicherweise zugangsbeschränkt.

03 / Chen et al., Preprint, Fassung 28. Mai 2026

The Price Reversal Phenomenon. Experimentelle Modellvergleiche; keine allgemeine Erfolgsquote für Unternehmen.

04 / Prompt Caching: Claude und Amazon Bedrock

Anbieterbedingungen und Nutzungsdaten. Mindestlänge, Lebensdauer und Tarife dienstabhängig. Technische Quellen geprüft am 6.10.2026.

05 / Notion: Grundlagen und Datenbanken

Funktionsbeschreibung des gemeinsamen Arbeitsbereichs. Eine bestehende Tabelle kann für den Versuch ebenso genügen.

06 / The Economist, 1. Oktober 2026

A hurricane is coming to tax systems. Titel und Einstieg als Diskussionsanlass. Keine Aussage über eine eingeführte deutsche KI-Steuer.

Weitere Technikquelle: Claude Platform, 8.9.2026

Amazon Bedrock: Prompt Caching

Notion: Einführung in Datenbanken

Buchbezug: Niko Martzy, Der KI Coach, bereitgestellte Innenfassung v2. Kapitel 9, S. 152 bis 155; Kapitel 10, S. 174; Kapitel 15, S. 274 bis 275; Kapitel 19, S. 349 bis 350. Gedruckte Seitenzählung. [Buchinformationen](#)

Transparenz: Whitesides Publishers. Niko Martzy ist Autor und Gründer der ecomex Business Academy; wirtschaftliches Interesse an KI-Weiterbildung. Rechnung: redaktionelles Beispiel. Drei hörbar gekennzeichnete Eigenwerbungen. Stimmen und Studioszene: KI-generiert. Redaktion und Freigabe: Menschen. Diese Fassung dient der Medienabnahme.