Schluss mit dem Limit

15 Wege, mehr aus Claude herauszuholen — ohne mehr zu bezahlen.

Oliver Raffaele Desogus · odeso.ai

01

Neues Thema, neues Gespräch

Claude hat kein Gedächtnis zwischen den Nachrichten. Es bekommt bei jeder Frage die komplette bisherige Unterhaltung erneut mitgeschickt — Nachricht eins bis vierzig, und dann deine neue Frage. Das ist der Grund, warum sich ein langes Gespräch am Ende zäh anfühlt und warum du schneller ans Limit stößt, als du gerechnet hast: Du bezahlst bei jeder Frage noch einmal für alles, was vorher da war.

Richtig teuer wird das, wenn in einem Chat Dinge liegen, die nichts miteinander zu tun haben. Wenn du morgens einen LinkedIn-Post geschrieben hast, mittags ein Angebot kalkuliert und abends fragst, wie lange Reis kocht, dann wird für die Reisfrage der Post und das Angebot mitgeschleppt. Claude braucht davon nichts.

Also: ein Thema, ein Gespräch. Sobald du merkst, dass du das Thema wechselst — neuer Chat. Das kostet dich zwei Sekunden und bringt dir nebenbei bessere Antworten, weil nichts Fremdes mehr im Kontext liegt, an dem Claude sich orientieren könnte.

Ein Anhaltspunkt: Wenn du deinem Chat einen Titel geben müsstest und dir keiner einfällt, der auf alles darin passt, sind es schon zwei Gespräche.

02

Zusammenfassen und neu starten

Auch wenn du sauber ein Thema pro Chat hältst, wächst ein Gespräch. Die zwanzigste Nachricht trägt neunzehn Nachrichten mit sich, die erste trug keine. Und das Meiste davon ist Weg, nicht Ergebnis: verworfene Entwürfe, Irrwege, Korrekturen, drei Anläufe für eine Überschrift. Claude braucht das nicht mehr, es wird trotzdem bei jeder weiteren Frage mitgeschickt.

Was du willst, ist der Stand — nicht die Geschichte, wie du dahin gekommen bist.

Deshalb: Bevor ein Gespräch zäh wird, bitte Claude um eine Zusammenfassung. Entscheidungen, Zwischenstände, offene Punkte, wichtige Festlegungen. Die kopierst du als erste Nachricht in ein neues Gespräch und arbeitest dort weiter. Du hast dasselbe Wissen, aber ohne die Umwege.

Wann der richtige Moment ist, merkst du daran, dass Antworten länger brauchen, dass Claude Dinge wieder aufgreift, die längst geklärt waren, oder dass du selbst nach oben scrollen musst, um den Stand zu finden. Wenn du scrollen musst, muss Claude es auch — nur eben jedes Mal.

Bei Projekten mit mehreren Sitzungen lohnt es sich, die Zusammenfassung zusätzlich in einer Datei abzulegen. Dann startest du morgen nicht bei null, sondern bei Punkt sieben.

```
Fasse zusammen, was wir bisher erarbeitet haben: getroffene Entscheidungen,
Zwischenstände, offene Punkte, alles, worauf wir uns festgelegt haben.
Kein Verlauf, kein Nacherzählen – nur der Stand.
Kurz genug, dass ich es als erste Nachricht in ein neues Gespräch kopieren kann.
```

03

Nur die relevanten Dateien anhängen

Ein Anhang verschwindet nicht, nachdem Claude ihn gelesen hat. Er liegt im Gespräch und geht bei jeder weiteren Nachricht wieder mit. Ein Ordner mit zwölf Dateien, von denen zwei zur Frage gehören, kostet dich also nicht einmal etwas, sondern in jeder Runde.

Dazu kommt ein zweiter Effekt, der mit dem Limit nichts zu tun hat und trotzdem zählt: Je mehr Nebensächliches im Kontext liegt, desto eher greift Claude auf das Falsche zurück. Du gibst fünf alte Angebote mit, weil eines davon als Vorlage dienen soll — und bekommst eine Mischung aus allen fünf.

Also vor jedem Anhang eine Frage: Muss Claude das gelesen haben, um diese Aufgabe zu lösen? Wenn du zögerst, häng es nicht an. Nachreichen kannst du immer.

Bei großen Dokumenten reicht fast immer der eine Abschnitt statt der ganzen Datei. Kopier die drei Seiten, um die es geht, statt das Handbuch mit zweihundert anzuhängen. Und für Aufgaben, die überhaupt keinen Dateibezug haben — eine Formulierung, eine Idee, eine Einschätzung — häng gar nichts an.

Wenn du im selben Chat später etwas anderes vorhast, mach lieber einen neuen auf. Sonst schleppst du die Anhänge des ersten Themas durch das zweite.

04

Dokumente als Text statt Bild

Bilder werden nicht als Text verarbeitet. Sie werden in Bildbausteine zerlegt, und davon gibt es umso mehr, je größer und detaillierter das Bild ist. Eine Seite als Screenshot belegt deshalb deutlich mehr Platz im Gespräch als dieselbe Seite als reiner Text. Du bezahlst für Ränder, Logo, Tabellenlinien und Schlagschatten mit — für alles, was auf dem Bild zu sehen ist, nicht nur für die Zahlen, die du wissen wolltest.

Bei PDFs gibt es zwei Sorten, und der Unterschied ist genau dieser: Solche, in denen echter Text steckt — du erkennst sie daran, dass du den Text markieren und kopieren kannst — und gescannte, die im Grunde eine Sammlung von Bildern sind. Die zweite Sorte ist die teure.

Also: markieren, kopieren, einfügen. Bei einem Scan können die Vorschau am Mac und die meisten PDF-Programme den Text erkennen und als Text exportieren; das dauert einmal eine Minute und zahlt sich bei jeder Rückfrage aus.

Bilder gehören dorthin, wo das Bild selbst die Information ist: ein Diagramm, ein Layout, ein Screenshot einer Fehlermeldung, ein handschriftlicher Zettel. Für Zahlen, Fließtext und Tabellen ist Text immer der günstigere Weg — und der genauere obendrein.

05

Werkzeuge abschalten, die du nicht brauchst

Damit Claude ein Werkzeug benutzen kann, muss es wissen, dass es existiert und wie es zu bedienen ist. Diese Beschreibung liegt im Gespräch — bei jeder Nachricht, ob das Werkzeug benutzt wird oder nicht. Wer Websuche, Google Drive, den Kalender und drei weitere Konnektoren aktiviert hat, schleppt bei jeder Frage die Bedienungsanleitung für alle sieben mit, auch bei der Frage nach einem Synonym.

Der zweite Effekt wiegt schwerer: Eingeschaltete Werkzeuge werden auch benutzt. Eine aktive Websuche führt dazu, dass Claude bei Gelegenheit nachschlägt, obwohl es die Antwort auch so gewusst hätte — und der Inhalt der gefundenen Seiten landet dann ebenfalls im Gespräch und wird von da an mitgeschickt. Aus einer Frage werden so drei Seiten Fremdtext.

Geh deshalb einmal durch die Einstellungen und schalte aus, was du im Alltag nicht brauchst. Für den größten Teil dessen, wofür du Claude benutzt — schreiben, umformulieren, ordnen, erklären, prüfen — brauchst du überhaupt kein Werkzeug.

Anschalten kannst du sie in dem Moment, in dem du sie wirklich brauchst. Das ist ein Klick. Dieser Weg ist deutlich billiger als der umgekehrte, bei dem alles immer an ist, für den Fall der Fälle.

WENIGER DURCHLÄUFE BRAUCHEN

06

Beim ersten Mal genau sagen

Jede Korrekturrunde ist eine vollständige neue Anfrage — mit allem, was vorher im Gespräch stand, samt der Fassung, die du gerade verwirfst. Wenn du viermal nachbesserst, hast du nicht einmal bezahlt, sondern fünfmal, und die letzte Runde war die teuerste von allen. Der Preis für eine ungenaue erste Nachricht wird also nicht vorne fällig, sondern hinten, und er wächst mit jeder Runde.

Ungenau heißt dabei fast nie, dass du das Ziel nicht genannt hast. Es heißt, dass du die Bedingungen nicht genannt hast. „Schreib mir einen Text über X" lässt offen, wie lang, für wen, in welchem Ton, in welchem Format — und Claude muss raten.

Vier Angaben entscheiden fast jede Aufgabe: was entstehen soll, für wen und wozu, wie lang, und was auf keinen Fall vorkommen darf.

Der letzte Punkt wird am häufigsten vergessen und spart am meisten. Wenn du weißt, dass du keine Emojis willst, keine Aufzählungslisten und kein „in der heutigen schnelllebigen Zeit", dann schreib es hin. Sonst streichst du es hinterher — und bezahlst für das Streichen eine weitere vollständige Runde.

```
Aufgabe: [was soll entstehen – Text, Liste, Tabelle, Entwurf]
Für wen: [Adressat und wozu er es liest]
Umfang: [Wortzahl, Zeilen, Anzahl]
Ton: [sachlich, direkt, freundlich – und was du gerade nicht willst]
Nicht vorkommen: [Emojis, Floskeln, Aufzählungen, Superlative]
Ausgabe: [nur der Text, ohne Vorrede und ohne Nachbemerkung]
```

07

Claude vorher Rückfragen stellen lassen

Wenn eine Aufgabe unterbestimmt ist, rät Claude. Es liefert etwas Vollständiges, das plausibel aussieht — und das genau an den Stellen danebenliegt, die du nicht genannt hast. Dieses vollständige Ergebnis kostet dich einen ganzen Durchlauf — und es bleibt danach im Gespräch stehen, auch wenn du es verwirfst. Deine Korrektur trägt den verworfenen Entwurf also mit.

Fragen kosten dagegen fast nichts. Fünf Rückfragen und deine Antworten darauf sind ein Bruchteil dessen, was ein falscher Entwurf plus Korrektur verbraucht. Und du merkst beim Beantworten selbst, was du eigentlich willst — das ist der unterschätzte Teil.

Häng deshalb an komplexere Aufgaben einen Satz an, bevor du losschickst: erst fragen, dann arbeiten. Antworte dann kurz, Stichworte reichen, und lass erst danach den Entwurf bauen.

Das lohnt sich überall dort, wo es ein „richtig" gibt, das nur du kennst: Angebote, Bewerbungen, Absagen, alles mit Zahlen, alles mit einem bestimmten Adressaten. Bei einer Rechtschreibprüfung oder beim Umformatieren einer Liste ist es dagegen reine Zeitverschwendung — da gibt es nichts zu fragen.

> Fang noch nicht an. Stell mir zuerst die Fragen, die du brauchst, um das richtig zu machen – höchstens fünf, nur die, deren Antwort das Ergebnis wirklich verändert. Ich beantworte sie, danach legst du los.

08

Mehrere Aufgaben in einer Nachricht

Jede einzelne Nachricht bringt den gesamten bisherigen Verlauf noch einmal mit. Wenn du erst „fasse den Text zusammen" schickst, dann „liste die Kernaussagen", dann „schlag drei Überschriften vor", dann geht der Ausgangstext dreimal über die Leitung und die jeweils vorherige Antwort zusätzlich. Fragst du alles auf einmal, geht der Text einmal rüber und du bekommst alles drei zurück.

Nummerier die Teilaufgaben dabei. Claude arbeitet eine nummerierte Liste zuverlässig ab, und du siehst am Ergebnis sofort, ob etwas fehlt — bei einem Fließtext-Auftrag mit drei Wünschen darin fällt der dritte gern unter den Tisch.

Es gibt eine Grenze, und die ist wichtig: Bündeln lohnt sich, solange die Teilaufgaben unabhängig voneinander sind. Wenn Aufgabe zwei davon abhängt, wie Aufgabe eins ausgefallen ist, und du das Zwischenergebnis erst sehen willst, dann trenn sie. Sonst baut Claude auf einer Grundlage weiter, die du noch gar nicht abgenommen hast, und am Ende korrigierst du alle drei Teile statt einem.

Faustregel: Alles, was du ohnehin gleich nachgefragt hättest, gehört in dieselbe Nachricht.

```
Hier ist der Text. Erledige bitte alle drei Aufgaben in einer Antwort:

1. Fasse ihn in fünf Sätzen zusammen.
2. Zieh die drei Kernaussagen als Stichpunkte heraus.
3. Schlag drei Überschriften vor, jede unter acht Wörtern.
```

09

Die alte Nachricht bearbeiten

Ein Fehlstart bleibt stehen. Wenn die erste Antwort danebengeht und du als nächste Nachricht „nein, ich meinte eigentlich …" hinterherschickst, bleibt beides im Gespräch — der Fehlversuch und deine Korrektur. Ab diesem Punkt trägt jede weitere Nachricht beides mit, bis zum Ende des Gesprächs. Drei solcher Fehlstarts hintereinander, und ein spürbarer Teil deines Chats besteht aus Dingen, die niemand mehr braucht.

Es gibt einen zweiten Weg, und der arbeitet anders herum: Du kannst eine bereits abgeschickte Nachricht nachträglich bearbeiten und neu senden. Dann wird der Gesprächsverlauf ab dieser Stelle neu aufgebaut. Die missglückte Antwort ist damit aus dem laufenden Zweig heraus und geht nicht mehr mit.

Wenn die erste Antwort also in die falsche Richtung geht, weil deine Frage nicht eindeutig war: nichts hinterherschieben. Geh auf deine eigene Nachricht, wähle Bearbeiten, formulier sie so, wie du sie gleich beim ersten Mal hättest schreiben sollen, und schick sie neu ab. Das geht in der App genauso wie im Browser.

Wann nicht: wenn die erste Antwort brauchbar war und du darauf aufbauen willst. Dann brauchst du sie im Verlauf. Bearbeiten ist für Fehlstarts da, nicht für Fortschritt.

10

Nur den falschen Teil neu machen

„Mach das nochmal, aber besser" führt dazu, dass der komplette Text erneut erzeugt wird. Alles daran kostet — der Teil, der schon gut war, genauso wie der, der nicht stimmte. Und du bekommst ein Ergebnis, in dem sich auch die gelungenen Stellen verändert haben, also liest du alles noch einmal von vorn und vergleichst. Der zweite Durchlauf ist damit teurer als der erste, obwohl nur ein Absatz falsch war.

Sag stattdessen, welcher Teil bleibt und welcher nicht — und, das ist der entscheidende Zusatz, dass du nur den geänderten Teil zurückwillst. Ohne diesen Satz liefert Claude meistens den vollständigen Text mit dem neuen Absatz darin. Dann steht derselbe Text zweimal im Gespräch und wird von da an doppelt mitgeschickt.

Bei längeren Sachen hilft es, gleich beim ersten Auftrag um nummerierte Abschnitte oder Zwischenüberschriften zu bitten. Dann kannst du beim Nachbessern präzise darauf zeigen, statt zu umschreiben, welche Stelle du meinst — und ein „Abschnitt 3" ist eindeutiger als „der Teil mit dem Beispiel", von dem es womöglich zwei gibt.

```
Abschnitt 1, 2 und 4 bleiben unverändert – die passen.
Schreib nur Abschnitt 3 neu: er ist zu allgemein, ich will ein konkretes
Beispiel darin und höchstens vier Sätze.
Gib mir nur diesen einen Abschnitt zurück, nicht den ganzen Text.
```

11

Kleine Aufgaben, kleines Modell

Claude ist nicht ein Modell, sondern mehrere, und sie sind für unterschiedlich schwere Arbeit gebaut. Haiku 4.5 ist das kleine und schnelle. Sonnet 5 ist das Alltagsmodell und bei den meisten voreingestellt. Opus 5 und Fable 5 sind die großen. Die großen denken gründlicher, und gründlicher denken heißt mehr Arbeit pro Antwort. Wer jede Kleinigkeit vom größten Modell erledigen lässt, steckt diesen Aufwand in Aufgaben, bei denen es nichts zu überlegen gibt — und hat ihn nicht mehr übrig, wenn die Arbeit kommt, die ihn braucht.

Die Faustregel: Alles, wo es genau eine richtige Antwort gibt und du sie im Zweifel selbst hinbekommen hättest, kann Haiku. Rechtschreibung und Grammatik, Umformatieren, Listen sortieren, Adressen aus einem Text ziehen, Standardtext übersetzen, eine kurze Zusammenfassung.

Sonnet 5 ist der Normalfall für alles dazwischen. Zu Opus 5 oder Fable 5 greifst du dort, wo wirklich geurteilt werden muss: die Struktur für etwas Neues, ein Text, der überzeugen soll, eine Analyse mit Widersprüchen darin, alles, was über mehrere Schritte aufeinander aufbaut.

Umschalten kannst du über dem Eingabefeld, auch mitten im Gespräch. Ein Unterschied noch: Wie viel in ein Gespräch passt, wird in Token gemessen — so heißen die Häppchen, in die Text zerlegt wird, ein längeres Wort besteht aus mehreren davon. Haiku 4.5 fasst 200.000 davon, die drei großen fassen eine Million. Bei kurzen Aufgaben merkst du das nicht, bei einem langen Gespräch schon.

Nebenbei antwortet Haiku spürbar schneller: bei Kleinkram kein Kompromiss, sondern das angenehmere Werkzeug.

Arbeit über den Tag verteilen

Das Nutzungslimit ist kein Tageskonto, das um Mitternacht auf null zurückspringt. Es zählt in einem rollierenden Fenster: Was du verbraucht hast, wird nach und nach wieder frei, sobald es hinten aus dem Fenster herausrutscht. Deshalb erlebst du das Limit als etwas Unberechenbares — du hast gestern Abend viel gearbeitet und stößt heute Vormittag an, obwohl doch „ein neuer Tag" ist.

Wie lang dieses Fenster ist, hängt vom Tarif ab, und die Werte ändern sich. Merk dir deshalb das Prinzip statt einer Zahl: Entscheidend ist nicht, wie viel du an einem Tag verbrauchst, sondern wie dicht du es zusammendrängst.

Praktisch heißt das: Leg die großen Sachen nicht in einen Block. Wenn du weißt, dass heute drei umfangreiche Aufgaben anstehen, verteil sie über den Tag — vormittags eine, nach dem Mittag die zweite, am späten Nachmittag die dritte. Dazwischen liegt ohnehin genug Arbeit, die ohne Claude läuft.

Und wenn du trotzdem anstößt: Das ist der Moment für die Dinge, die du vor dir herschiebst — Buchhaltung, Telefonate, Aufräumen. Das Fenster läuft weiter, ob du davorsitzt und wartest oder nicht.

13

Claude nicht für alles benutzen

Jeder Versuch zählt, auch der, der nichts bringt. Wer dreimal versucht, ein Bild erzeugen zu lassen, hat drei Versuche verbraucht und kein Bild. Das ist der stillste Verbrauch von allen, weil er sich nicht wie Arbeit anfühlt.

Vier häufige Fehlgriffe:

Bilder erzeugen. Claude kann Bilder ansehen und beschreiben, aber keine erstellen. Dafür gibt es andere Werkzeuge.

Aktuelle Daten. Kurse, Wetter, Spielstände, Öffnungszeiten. Entweder Claude sucht im Netz — dann landen ganze Seiten in deinem Gespräch — oder es antwortet aus einem Wissensstand, der ein paar Monate alt ist. Eine Suchmaschine ist schneller und genauer.

Exaktes Rechnen. Eine Tabellenkalkulation rechnet richtig, ein Sprachmodell rechnet plausibel. Bei allem, was auf den Cent stimmen muss, nimm ein Programm, das rechnet — oder lass Claude die Formel bauen und rechne sie dort.

Einzelne Fakten nachschlagen. Eine Hausnummer, eine Jahreszahl. Dafür ein Gespräch aufzumachen ist der umständlichste Weg zu einer Antwort.

Prüf also kurz, bevor du fragst: Geht es um Sprache oder um Denken? Dann ist Claude richtig. Geht es um Nachschlagen, exaktes Rechnen oder um etwas, das kein Text ist? Dann nimm das Werkzeug, das dafür gebaut wurde, und heb dir dein Fenster für die Arbeit auf, die sonst niemand macht.

NICHT ZWEIMAL DASSELBE ERKLÄREN

14

Wiederkehrende Anweisungen einmal festhalten

Wiederholung ist der teuerste Verbrauch überhaupt, weil er unsichtbar bleibt. Wenn du jedes Mal, bevor du ein Angebot schreiben lässt, zehn Zeilen tippst — wie deine Angebote aufgebaut sind, welcher Ton, welche Bausteine immer vorkommen, was nie drinsteht —, dann schickst du diese zehn Zeilen jedes Mal neu. Zwanzig Angebote im Quartal, zwanzigmal derselbe Text.

Der zweite Schaden ist größer als der erste: Weil du die Anweisung jedes Mal aus dem Kopf formulierst, ist sie jedes Mal ein bisschen anders. Mal erwähnst du die Zahlungsziele, mal nicht. Das Ergebnis schwankt, obwohl die Aufgabe dieselbe war — und du kannst dich auf nichts verlassen.

Schreib den Ablauf deshalb einmal auf, statt ihn jedes Mal zu erklären. In der Weboberfläche ist die Stelle dafür ein Projekt: Deine Anweisungen liegen als Projektwissen darin, und jedes Gespräch in diesem Projekt kennt sie, ohne dass du sie tippst.

Wer mit Claude Code arbeitet, hat dafür Skills — eine Textdatei, in der steht, was passieren soll, was Claude dafür wissen muss und wie das Ergebnis auszusehen hat. Kein Code, nur Text. Danach genügt ein Aufruf statt einer Erklärung.

Die Faustregel ist für beides dieselbe: Wenn du eine Anweisung zum dritten Mal tippst, gehört sie nicht in den Chat, sondern in eine Datei.

```
Ich habe dir das jetzt in mehreren Gesprächen ähnlich erklärt.
Schreib daraus eine dauerhafte Anweisung, die ich einmal hinterlegen kann:
was zu tun ist, was du dafür von mir brauchst, wie das Ergebnis aussehen soll.
Nur die Regeln, keine Beispiele aus diesem Gespräch.
```

15

Dauerhafte Einstellungen einmal hinterlegen

Das ist der kleine Bruder von Punkt 14, und er wird noch häufiger übersehen. Es gibt Vorgaben, die nicht nur für einen bestimmten Ablauf gelten, sondern für alles, was du tust: dass du auf Deutsch arbeitest, dass du keine Emojis willst, dass du ganze Sätze lieber liest als Stichpunktlisten, dass Claude beim Widerspruch nachfragen soll statt höflich zuzustimmen.

Wer das in jedem Gespräch neu hinschreibt, tippt dieselben zwei Sätze hundertmal im Jahr. Schlimmer ist der andere Fall: Du vergisst sie in genau dem Gespräch, in dem es darauf ankam, bekommst ein Ergebnis im falschen Ton und bezahlst eine Korrekturrunde für etwas, das dauerhaft hätte geregelt sein können.

In den Claude-Einstellungen gibt es einen Bereich für persönliche Vorgaben: was Claude über dich wissen soll und wie es antworten soll. Was dort steht, gilt in jedem neuen Gespräch, ohne dass du es erwähnst.

Halt es kurz und schreib nur hin, was wirklich immer gilt. Vorgaben, die nur für ein Thema gelten, gehören in ein Projekt und nicht in die allgemeinen Einstellungen — sonst schleppst du sie überall mit, auch dort, wo sie bloß im Weg stehen. Zwei Sätze, die stimmen, sind mehr wert als zwanzig, die meistens nicht passen.

Und wenn du dieselbe Anweisung zum zehnten Mal tippst?
Dann gehört sie in einen Skill. Im Odeso Skill Vault liegen 20 fertige — 29 €, einmalig, kein Abo.

Odeso · Oliver Raffaele Desogus · od@odeso.ai · Impressum