

Gleicher Prompt, neues Modell: Warum ein Modellwechsel nicht harmlos ist

Was der Wechsel von Opus 4.8 auf Opus 5.x mit meinen Beitragsbildern angestellt hat und was das für KI im Unternehmen bedeutet

Neues Modell, in jeder Hinsicht besser, also umstellen. So läuft das in vielen Unternehmen, und so lief das auch bei mir. Ein paar Tage später stand ich auf einem Beitragsbild mit einer Lilie in der Hand vor einem Grabstein, auf dem in großen Lettern **ADFS** steht. Bestellt hatte das niemand.

Dieser Artikel zeigt an echten Beispielen, warum auch der Wechsel auf das **direkte Nachfolgemodell** nicht automatisch unkritisch ist. Bei Bildern sieht man die Auswirkungen sofort, und genau deshalb eignen sie sich so gut als Anschauungsmaterial. Bei Klassifizierungen, Zusammenfassungen oder Texten passiert dasselbe, nur leiser.

FAKTEN | Das Wichtigste in Kürze

- Gleicher Prompt, gleicher Bildgenerator. Gewechselt hat nur das Modell, das den Bildprompt schreibt: von Claude Opus 4.8 auf Opus 5.x.
- Alle harten Vorgaben wurden weiter eingehalten. Versoben hat sich die Auslegung weicher Begriffe wie „Humor“.
- Das Ergebnis: Friedhöfe, Abrissbirnen, Buchstabensalat. Kreativ, aber nicht mehr business-tauglich.
- Die Lehre: Ein Modellwechsel ist ein Release und braucht einen Gegencheck, auch beim direkten Nachfolger.

Woher die Bilder kommen

Jeden Tag erscheint auf boddenberg.de ein neues **Consulting Briefing**: ein aktuelles Thema aus Beraterperspektive, mit der Frage, was es für dich bedeutet und was jetzt zu tun ist. Statt anonymen Stockfotos bekommt jedes Briefing ein eigenes, thematisch passendes Bild mit Niko und mir.

Die Pipeline dahinter ist schlicht. Ein Claude-Modell liest das Briefing und schreibt daraus einen Bildprompt, ein Bildgenerator setzt ihn um. Im Prompt stecken feste Regeln: Niko steht zentral im Bild und ist hellwach, ich bin erkennbar, die Szene spielt in einer realistischen deutschen IT-Umgebung, weitere Menschen wirken glaubwürdig. Dazu kommt die Tonalität: **Humor ja, aber dezent**. Das lief über Monate stabil.

TIPP | Warum eigene Bilder ein guter Frühwarnmelder sind

Stockfotos sind austauschbar, eigene Bilder machen eine Marke wiedererkennbar. Der Nebeneffekt ist mindestens genauso wertvoll: Ein Bild, das jeden Tag neu entsteht, ist ein täglicher Dauertest für die ganze Pipeline. Mit Stockfotos wäre die Drift nie aufgefallen.

Vorher: So sah ein normales Bild aus



Abbildung 1: Vorher, Prompt-Modell Opus 4.8. Büroszene, Kolleginnen und Kollegen bei der Arbeit, Niko mittendrin. Unspektakulär und genau richtig.

So sahen die Briefing-Bilder über Monate aus. Das Thema ist erkennbar, die Szene glaubwürdig, niemand guckt entsetzt. Ein Bild, das man ohne Bauchschmerzen in einen LinkedIn-Feed voller Entscheider stellen kann.

Nachher: Die Galerie der Nebenwirkungen

Nach der Umstellung auf Opus 5.x kippte das innerhalb weniger Tage. Beim Durchsehen fallen vier Muster auf, und jedes ist für sich lehrreich.

Muster 1: Der Tod als Standardmetapher



Abbildung 2: Thema ADFS-Abschaltung. Das Modell veranstaltet gleich eine Trauerfeier im Rechenzentrum, inklusive Kerzen und Kranz.



Abbildung 3: Thema Kontenbereinigung an Hochschulen. Friedhof, Geister, Doktorhut auf dem Grabstein und ein Sarg voller Disketten.

Begriffe wie „abgekündigt“, „Altlast“ oder „Konten löschen“ übersetzt das neue Modell konsequent in Tod und Beerdigung. Das ist bildsprachlich clever und für sich genommen sogar witzig. Nur will kein Kunde sein Projekt beerdigt sehen, und eine Hochschule möchte ihre ehemaligen Studenten nicht als Geister über den Campus schweben lassen.

Muster 2: Metaphern werden wörtlich genommen



Abbildung 4: Thema Altsysteme ablösen. Eine Ausgrabungsstätte mit Spinnweben, Pinsel, Kelle und Spurensicherungs-Nummern.



Abbildung 5: Thema Migration. Die Abrissbirne kracht durch die Wand, während im Wohnzimmer gearbeitet wird.

„Modernisierung“ wird zur Archäologie, „Migration“ zur Abrissbirne. Das alte Modell hat solche Bilder höchstens angedeutet, das neue baut sie mit voller Ausstattung auf. Gemeint war ein Augenzwinkern, geliefert wurde ein Katastrophenfilm.

Muster 3: Der Berater lächelt, während alles einstürzt



Abbildung 6: Entsetzte Gesichter, eingerissene Wand, freiliegende Kabel. Und der Berater in der Mitte strahlt mit ausgebreiteten Armen.

Das ist das gefährlichste Muster, weil es nicht nur geschmacklich danebenliegt, sondern eine falsche Botschaft sendet. Wer das Bild ohne Kontext sieht, liest: **Der war's**. Für eine Beratungsseite ist das schlicht geschäftsschädigend.

WARNUNG | Die unfreiwillige Botschaft

Stell dir vor, ein Interessent scrollt morgens durch seinen Feed und sieht genau dieses Bild. Den Artikeltext liest er nicht, das Bild bleibt hängen. Die Erinnerung lautet dann nicht „kompetenter Berater für Migrationen“, sondern „das ist doch der, bei dem die Wand eingestürzt ist“. Kein Regelcheck der Welt schlägt hier an, denn formal ist alles erfüllt.

Muster 4: Buchstabensalat statt Botschaft



Abbildung 7: Schlüsselverwaltung mit jeder Menge Schildern, darunter das schöne Wort „Rechesznbum“.



Abbildung 8: Übergabe-Ordner mit Schleife, Haftnotizen und dem rätselhaften Hinweis „nicht thalten“.

Das neue Modell packt deutlich mehr Text in die Szenen: Schilder, Etiketten, Aufkleber. Je mehr Text der Bildgenerator setzen muss, desto höher die Wahrscheinlichkeit für peinliche Tippfehler. Und eine Fachseite, auf deren Bildern „Rechesznbum“ steht, verspielt Glaubwürdigkeit, bevor der erste Satz gelesen ist.

Der Grenzfall



Abbildung 9: Puzzle mit Schlüssel, Niko trägt das fehlende Teil im Maul. Verspielt, aber noch vertretbar.

Nicht jedes Bild ist misslungen. Dieses hier zeigt gut, wo die Linie verläuft: Die Metapher bleibt positiv, niemand kommt zu Schaden, nichts geht kaputt. Genau diese Art von Humor war mit „dezent“ gemeint.

Warum das passiert

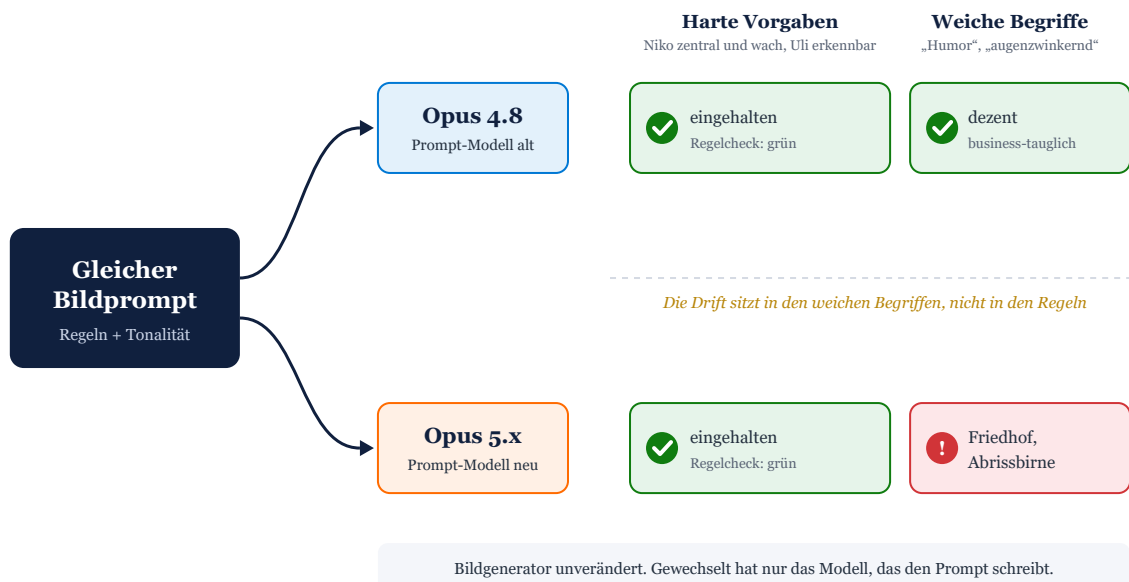


Abbildung 10: Harte Vorgaben überleben den Modellwechsel, weiche Begriffe nicht.

Prompts sind auf ein Modell **eingeschwungen**. Über Monate wird ein Prompt so lange nachgeschärft, bis das Ergebnis passt. Dabei kalibriert man ihn unbemerkt auf die Art, wie genau dieses eine Modell Begriffe auslegt. „Humor, aber dezent“ hieß für Opus 4.8 ein leichtes

Augenzwinkern. Für Opus 5.x ist dieselbe Formulierung eine Einladung zu deutlich mutigeren Bildideen.

Das ist ausdrücklich **kein Fehler des neuen Modells**. Es ist fähiger, nimmt Anweisungen ernster, interpretiert mehr und wählt kühnere Bilder. In fast jeder Hinsicht ist das ein Fortschritt. Nur war der alte Prompt nicht für diesen Fortschritt geschrieben.

WICHTIG | Besser heißt nicht gleich

Ein neues Modell ist in Benchmarks besser als sein Vorgänger. Dein Prompt ist aber kein Benchmark. Er ist ein über Monate abgestimmtes Zusammenspiel mit einem ganz bestimmten Modell. Wer das Modell wechselt, wechselt im Grunde die Software, auch wenn sich nur eine Versionsnummer ändert.

Art der Vorgabe	Beispiel	Verhalten beim Modellwechsel
Harte Regel	„Niko zentral und wach“	Bleibt stabil, leicht zu prüfen
Format und Struktur	„16:9, Büroumgebung“	Weitgehend stabil
Weicher Begriff	„humorvoll, aber dezent“	Driftet, schwer zu prüfen
Implizites Wissen	Was „business-tauglich“ bedeutet	Driftet am stärksten, steht nirgends geschrieben

Bei Bildern sieht man es, bei Labels nicht

Die Grabsteine sind eigentlich ein Glücksfall. Ein misslungenes Bild sieht jeder sofort, und man kann es einfach neu erzeugen. Derselbe Mechanismus wirkt aber überall, wo ein Modell weiche Begriffe auslegen muss: bei der Klassifizierung von Dokumenten, bei Zusammenfassungen, bei Metadaten, bei Texten in Hausstil.

Wenn nach einem Modellwechsel drei Prozent der Dokumente plötzlich als „intern“ statt als „vertraulich“ eingestuft werden, merkt das niemand. Es gibt keinen Grabstein, der ins Auge springt. Die Labels kippen leise, und das Problem fällt erst beim nächsten Audit auf, oder beim nächsten Datenabfluss.

Anwendung	Wie sich Drift zeigt	Wer es merkt
Beitragsbilder	Friedhof statt Büro	Jeder, sofort
Texte und Mails	Tonalität verschiebt sich	Aufmerksame Leser, nach Tagen
Zusammenfassungen	Mehr Interpretation, andere Gewichtung	Selten jemand
Klassifizierung und Labels	Einzelne Kategorien kippen	Niemand, bis zum Audit

WARNUNG | Wenn der Anbieter das Modell für dich wechselt

Noch heikler wird es, wenn du das Modell gar nicht selbst wählst. Bei Diensten wie Microsoft 365 Copilot entscheidet der Anbieter, welches Modell gerade antwortet, und stellt um, ohne zu fragen. Für interaktive Arbeit, bei der ein Mensch jedes Ergebnis sieht, ist das vertretbar. Für Hintergrundprozesse wie Klassifizierung ist es ein Risiko, das du kennen solltest.

Was daraus folgt: Modellwechsel als Release



Abbildung 11: Fünf Schritte, damit ein Modellwechsel keine Überraschung wird.

Ein Referenzset anlegen

Für jede KI-Funktion lohnt sich ein kleines Set typischer Fälle mit bekanntem Sollergebnis: dreißig Dokumente mit dem richtigen Label, zehn Bildprompts mit einem akzeptierten Bild, fünf Mails mit der gewünschten Zusammenfassung. Dazu ein paar bewusst schwierige Grenzfälle. Das ist ein Nachmittag Arbeit und spart Wochen der Ursachenforschung.

Alt gegen Neu laufen lassen

Vor jedem Umstieg läuft das Referenzset durch beide Modelle. Bei Klassifizierungen lässt sich das automatisch über die Trefferquote auswerten. Bei Bildern und Texten hilft nur eins: hinschauen. Fünf Minuten Sichtung hätten mir den Friedhof erspart.

Version pinnen statt Alias

Im Produktivbetrieb gehört eine **fest benannte Modellversion** in die Konfiguration, kein Alias, der automatisch auf das jeweils neueste Modell zeigt. Der Wert gehört in die Einstellungen, nicht in den Code, und die verwendete Modellversion gehört in jedes Protokoll. Wer in drei Monaten fragt, warum ein Dokument damals anders eingestuft wurde, bekommt dann eine Antwort.

Intern schnell, beim Kunden entscheidet der Kunde

Für meine internen Werkzeuge stelle ich weiterhin sofort auf das neueste Modell um, allerdings erst nach dem Gegencheck. Dort bin ich Testpilot und Nutznießer zugleich, und wenn etwas schiefgeht, bade ich es selbst aus. Bei Produkten, die beim Kunden laufen, ist das Modell Teil der freigegebenen Konfiguration, genau wie eine Softwareversion. Ich prüfe neue Modelle und spreche eine Empfehlung aus, ob und wann umgestellt wird, entscheidet der Kunde.

Weiche Begriffe nachschärfen

Nach einem Modellwechsel braucht der Prompt fast immer eine Überarbeitung. Weiche Begriffe wie „Humor“ werden durch konkrete Grenzen ergänzt: keine Friedhöfe und Grabsteine, keine Zerstörung, kein Chaos mit entsetzten Statisten, nur selten Text im Bild. Was vorher stillschweigend funktioniert hat, muss jetzt ausdrücklich dastehen.

FAKTEN | Checkliste vor dem Modellwechsel

- Gibt es ein Referenzset mit Sollergebnissen und Grenzfällen?
- Ist das Set durch altes und neues Modell gelaufen?
- Hat ein Mensch die Ergebnisse gesichtet, nicht nur die Statistik?
- Sind weiche Begriffe im Prompt durch konkrete Grenzen abgesichert?
- Ist die neue Modellversion fest eingetragen und wird sie protokolliert?
- Ist bei Kundensystemen geklärt, wer den Wechsel freigibt?

TIPP | Neue Modelle brauchen oft weniger Prompt

Ein fähigeres Modell versteht vieles, was man dem alten noch umständlich erklären musste. Statt den Prompt nach dem Wechsel nur mit weiteren Verboten aufzufüllen, lohnt sich ein Blick, was sich streichen lässt. Ein kürzerer, klarerer Prompt ist beim nächsten Wechsel auch wieder leichter zu prüfen.

Häufig gestellte Fragen

Ist das neue Modell also schlechter?

Nein. Es ist fähiger und interpretiert mutiger, genau das löst die Drift aus. Das Problem liegt nicht im Modell, sondern in einem Prompt, der auf das alte Modell abgestimmt war.

Betrifft das nur Claude?

Nein, das betrifft jeden Anbieter und jede Modellfamilie. Bei Diensten, in denen der Anbieter das Modell selbst auswählt, hast du den Wechsel nicht einmal in der Hand.

Gilt das auch für kleine Versionssprünge?

Ja, gerade dort. Bei einem großen Generationswechsel rechnet jeder mit Änderungen und prüft. Beim direkten Nachfolger schaltet man einfach um, und genau dann kommt die Überraschung.

Wie groß muss ein Referenzset sein?

Für die meisten Anwendungen reichen zwanzig bis fünfzig typische Fälle plus eine Handvoll Grenzfälle. Wichtiger als die Menge ist, dass die Fälle den echten Alltag abbilden und das Sollergebnis eindeutig ist.

Sollte man dann besser gar nicht mehr umstellen?

Doch, unbedingt. Neue Modelle bringen echte Verbesserungen, und wer dauerhaft auf alten Versionen sitzen bleibt, verschenkt Qualität und irgendwann auch den Support. Umstellen ja, aber mit Gegencheck und nicht blind per Klick.

WEITER | Mehr zu Claude im Unternehmen

Modelle, Bezugswege, Governance und Datenschutz im Überblick: [Claude im Unternehmen](#)

Du willst KI-Funktionen sauber einführen und Modellwechsel beherrschbar machen?
[Beratung zu Claude und KI-Governance](#)