Hilft ein Glossar gegen den KI-Drift? Selbst ausprobieren.

Begleitmaterial zur Folge 2 des Kanals «Stephan Fischli · Agentic Orchestration» · Stand November 2026

Im Video habe ich vier Testreihen zu je fünf Läufen gemessen: Die KI modelliert denselben Prozess, einmal ohne Glossar und dreimal mit einem Glossar, das von Reihe zu Reihe mehr kann. Mit diesem Paket habt ihr dasselbe Material: die Unterlagen der fiktiven Firma 4Health, das Glossar so, wie es am Ende des Versuchs bestätigt war, und die Verbregel. Damit könnt ihr ausprobieren, was ein Glossar in eurem eigenen Werkzeug bewirkt, oder einfach nachsehen, wie ein kuratiertes Glossar aufgebaut ist.

Im Paket

- `01_Unterlagen_4Health/`: die 28 Dokumente des Projekts «Sales & Order Management»: Prozessbeschreibungen, AGB, Schnittstellenbeschreibung, Interviewnotizen, E-Mails, Excel-Listen, eine alte Schulung. Englisch, Deutsch, Französisch und Italienisch, und wie in jeder echten Firma widersprechen sie sich, vor allem in ihren Wörtern.
- `02_Glossar_4Health.csv`: das Glossar aus dem Versuch, 487 bestätigte Begriffe, mit Begriff je Sprache, Synonymen, abgelehnten Wörtern, Definition und Fundstellen. Öffnen mit Excel.
- `03_Verbliste.csv`: die Verbregel aus den Reihen B2 und B3: ein verbindliches Verb je Art von Arbeit, dazu die fachlichen Verben, die stehen bleiben.
- `04_Aufgabe.txt`: die Bitte an die KI für Reihe A (ohne Glossar) und Reihe B (mit Glossar und Verbliste).
- `05_Anleitung.pdf`: dieses Dokument, mit Messblatt und den Werten aus dem Video.
- `06_Messblatt.csv`: das Messblatt zum Ausfüllen in Excel.
- `07_Artikel_DE.pdf` und `08_Article_EN.pdf`: der Artikel zur Folge, mit allen Zahlen und Quellen, auf Deutsch und Englisch.
- `…_EN`: Aufgabe, Anleitung und Messblatt auf Englisch. Glossar und Verbliste tragen deutsche und englische Spaltentitel.

Alle Unterlagen sind erfunden; 4Health, Personen, Partner und Abläufe sind fiktiv. Vermerke wie «vertraulich» in einzelnen Dokumenten gehören zur Geschichte. Das Glossar wurde aus allen 122 Dokumenten des Falls gewonnen; die 94 Dokumente aus dem ersten 4Health-Projekt (Einführung von BPM) sind nicht im Paket, darum verweisen einzelne Fundstellen auf Dokumente, die hier fehlen. Das Glossar ist so, wie es im Versuch verwendet wurde, nicht nachträglich poliert: Ihr findet darin auch Fehler der KI, etwa Synonyme mit der falschen Sprache oder Begriffe aus Verträgen, die mit der Bestellung wenig zu tun haben.

So ist das Glossar entstanden

Von Hand schreibt niemand ein Glossar aus über hundert Dokumenten in vier Sprachen. Darum lief es in drei Schritten, so wie im Video gezeigt:

1 **KI-Vorschläge**
Die KI hat die Unterlagen gelesen und Begriffe mit Bedeutung und Fundstelle gesammelt: 2072 Vorschläge, einer pro Wort, nicht pro Begriff.

2 **Kuration**

Die KI ordnet jeden Vorschlag einem Begriff zu und prüft, ob er zum Fokus des Projekts gehört. Aus fünf Zeilen für die Gutschrift wird ein Begriff mit einem Namen je Sprache; die anderen Wörter werden Synonyme.

3 **Bestätigung**

Ein Begriff gilt erst, wenn ihn jemand vom Fach bestätigt. Abgelehnte Wörter werden ausdrücklich markiert. Am Schluss von Hand ergänzt: die drei Kernobjekte des Prozesses, die die KI nicht gefunden hatte.

Wie man eine Zeile liest

Begriff (EN)	DE	Synonyme	Abgelehnt
sales order	Kundenauftrag	B1 sales order (en), SAP sales order (en), Verkaufsauftrag (de)	ERP order (en): im Modell → sales order
credit note	Gutschrift	A/R-Gutschrift (de)	–
parcel	Paket	Sendung (de), package (en), shipment (en)	–
production order	Produktionsauftrag	Fertigungsauftrag (de)	–

Synonyme sind erlaubt und werden erkannt, im Modell aber durch den verbindlichen Begriff ersetzt. **Abgelehnte** Wörter dürfen gar nicht vorkommen. «ERP order» ist bei 4Health ein solches Wort: Im Modell heisst es immer «sales order».

Selbst ausprobieren

1 Reihe A: fünfmal ohne Glossar

Öffnet für jeden Lauf eine **neue** Unterhaltung (ohne Verlauf, ohne Projektwissen), hängt die Unterlagen an und schickt die Bitte für Reihe A aus `04_Aufgabe.txt`. Keine Rückfragen beantworten, nichts nachbessern. Speichert jede Antwort.

2 Reihe B: fünfmal mit Glossar und Verbliste

Dasselbe, aber mit der Bitte für Reihe B und zusätzlich `02_Glossar_4Health.csv` und `03_Verbliste.csv` im Anhang.

3 Zählen

Legt pro Reihe alle Schrittnamen der fünf Läufe untereinander und zählt:

- **verschiedene Schrittnamen**: wie viele verschiedene Namen kommen vor? «Check order» und «Validate order» zählen als zwei.
- **verschiedene Verben**: wie viele verschiedene erste Wörter?
- **gleiche Hauptschritte**: wie viele Hauptschritte heissen in allen fünf Läufen gleich?

Tragt die Zahlen ins Messblatt ein und vergleicht Reihe A mit Reihe B.

> **Zu viele Dateien?** Manche Werkzeuge nehmen nur wenige Anhänge. Dann nehmt in beiden Reihen dieselbe Auswahl, zum Beispiel SO-001, SO-008, SO-010, SO-011, SO-014, SO-015 und SO-017. Wichtig ist nur, dass A und B dieselben Unterlagen bekommen.

> **Worauf achten.** Kommt das Glossar überhaupt an? Ein Glossar, das die KI nicht liest, wirkt nicht; im Versuch kamen zuerst nur 3 von 473 Begriffen bei der KI an. Und wo die Namen gleich werden: Bleibt auch der Zuschnitt der Schritte gleich? Ein Glossar regelt Wörter, nicht, was ein eigener Schritt ist.

Werte aus dem Video

Prozess «B2C order-to-cash» · Modell gpt-oss · Ausgabe Englisch · je 5 Läufe · pro Lauf baut die KI die Analyse neu, danach das Modell mit allen Unterprozessen (rund 45–50 Aktivitäten pro Lauf) · umgesetzt in meinem Labor-Werkzeug MESTRO, November 2026.

Reihe	verschiedene Aktivitätsnamen	verschiedene Verben	Ereignisse in allen 5 Läufen gleich	Flüsse in allen 5 Läufen gleich
A · ohne Glossar	190	65	6	1
B · 484 bestätigte Begriffe	181	60	6	3
B2 · B, Synonyme kommen an, Verbregel	135	13	7	4
B3 · B2 und die drei Kernobjekte	**137**	**18**	**10**	**8**

Zwischen B2 und B3 wurden fachliche Verben wie Ship, Produce, Pick und Pack ausdrücklich erlaubt (in B2 war aus «Ship order» oft «Create shipment» geworden); darum steigt die Zahl der Verben wieder leicht. Die Werte aus dem Video und die aus einem Chat sind nicht direkt vergleichbar: Im Versuch entstand ein ganzes Modell mit Unterprozessen, hier nur eine Liste. Vergleichen lässt sich aber der Unterschied zwischen A und B.

Messblatt

Werkzeug / Modell: _____ Datum: _____

Lauf	Reihe A Anzahl Schritte	Reihe B Anzahl Schritte	Bemerkung
1			
2			
3			
4			
5			
verschiedene Schrittnamen			
verschiedene Verben			
gleiche Hauptschritte			

> Schreibt mir in die Kommentare unter dem Video, was das Glossar in eurem Werkzeug bewirkt hat.
> Und: Wie viele Wörter hat eure Firma für «Bestellung»?

bpmn.ch · Stephan Fischli · Begleitmaterial zur Folge 2 «Hilft ein Glossar gegen den KI-Drift?». Frei verwendbar für eigene Versuche und Schulungen, mit Quellenangabe.