Does a glossary stop the AI drift? Try it yourself.

Material for episode 2 of the channel «Stephan Fischli · Agentic Orchestration» · as of November 2026

In the video I measured four test series of five runs each: the AI models the same process, once without a glossary and three times with a glossary that can do more from series to series. With this package you have the same material: the documents of the fictional company 4Health, the glossary as it was confirmed at the end of the experiment, and the verb rule. With it you can try out what a glossary does in your own tool, or simply look at how a curated glossary is built.

In the package

- `01_Unterlagen_4Health/`: the 28 documents of the project "Sales & Order Management": process descriptions, terms and conditions, an interface specification, interview notes, e-mails, Excel lists, an old training course. English, German, French and Italian, and as in every real company they contradict each other, above all in their words.
- `02_Glossar_4Health.csv`: the glossary from the experiment, 487 confirmed terms, with the term per language, synonyms, rejected words, definition and sources. Column titles in German and English. Open with Excel.
- `03_Verbliste.csv`: the verb rule of series B2 and B3: one binding verb per kind of work, plus the business verbs that stay.
- `04_Task_EN.txt`: the request to the AI for series A (without glossary) and series B (with glossary and verb list).
- `05_Guide_EN.pdf`: this document, with the score sheet and the values from the video.
- `06_Score_sheet_EN.csv`: the score sheet to fill in with Excel.
- `07_Artikel_DE.pdf` and `08_Article_EN.pdf`: the article for the episode, with all numbers and sources, in German and English.
- The files without _EN are the same in German.

All documents are fictional; 4Health, people, partners and processes are invented. Marks such as "confidential" in some documents are part of the story. The glossary was obtained from all 122 documents of the case; the 94 documents of the first 4Health project (introducing BPM) are not in the package, so some sources point to documents that are missing here. The glossary is as it was used in the experiment, not polished afterwards: you will also find the AI's mistakes in it, such as synonyms with the wrong language or terms from contracts that have little to do with ordering.

How the glossary came about

Nobody writes a glossary by hand from over a hundred documents in four languages. So it took three steps, as shown in the video:

1 **AI suggestions**
The AI read the documents and collected terms with meaning and source: 2072 suggestions, one per word, not per term.

2 **Curation**
The AI assigns each suggestion to a term and checks whether it belongs to the project's focus. Five rows for the credit note become one term with one name per language; the other words become synonyms.

3 **Confirmation**
A term only applies once a domain expert confirms it. Rejected words are marked explicitly. Added by hand at the end: the three core objects of the process, which the AI had not found.

How to read a row

Term (EN)	DE	Synonyms	Rejected
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)	–

Synonyms are allowed and recognised, but replaced by the binding term in the model. **Rejected** words must not appear at all. "ERP order" is such a word at 4Health: in the model it is always "sales order".

Try it yourself

1 Series A: five times without glossary

For every run, open a **new** conversation (no history, no project knowledge), attach the documents and send the request for series A from `04_Task_EN.txt`. Do not answer follow-up questions, do not correct anything. Save every answer.

2 Series B: five times with glossary and verb list

The same, but with the request for series B and additionally `02_Glossar_4Health.csv` and `03_Verbliste.csv` attached.

3 Count

Per series, list all step names of the five runs below each other and count:

- **different step names**: how many different names occur? "Check order" and "Validate order" count as two.
- **different verbs**: how many different first words?
- **same main steps**: how many main steps have the same name in all five runs?

Enter the numbers in the score sheet and compare series A with series B.

> **Too many files?** Some tools only accept a few attachments. Then use the same selection in both series, for example SO-001, SO-008, SO-010, SO-011, SO-014, SO-015 and SO-017. What matters is that A and B get the same documents.

> **What to watch for.** Does the glossary arrive at all? A glossary the AI does not read has no effect; in the experiment, at first only 3 of 473 terms reached the AI. And where the names become the same: does the cut of the steps stay the same too? A glossary governs words, not what counts as a step of its own.

Values from the video

Process "B2C order-to-cash" · model gpt-oss · output in English · 5 runs each · in each run the AI rebuilds the analysis, then the model with all sub-processes (about 45–50 activities per run) · implemented in my lab tool MESTRO, November 2026.

Series	different activity names	different verbs	events the same in all 5 runs	flows the same in all 5 runs
A · without glossary	190	65	6	1
B · 484 confirmed terms	181	60	6	3
B2 · B, synonyms arrive, verb rule	135	13	7	4
B3 · B2 and the three core objects	**137**	**18**	**10**	**8**

Between B2 and B3, business verbs such as Ship, Produce, Pick and Pack were explicitly allowed (in B2, "Ship order" had often become "Create shipment"); that is why the number of verbs rises slightly again. The values from the video and those from a chat are not directly comparable: in the experiment a whole model with sub-processes was built, here only a list. The difference between A and B, however, can be compared.

Score sheet

Tool / model: _____ Date: _____

Run	Series A number of steps	Series B number of steps	Remark
1			
2			
3			
4			
5			
different step names			
different verbs			
same main steps			

> Tell me in the comments under the video what the glossary did in your tool. And: how many words does your company have for "order"?

bpmn.ch · Stephan Fischli · Material for episode 2 "Does a Glossary Stop the AI Drift?". Free to use for your own experiments and training, with attribution.