SPA-RAG:

A Smarter Way to Validate and Fix Terminology in AI Workflows

Pavel Soukenik

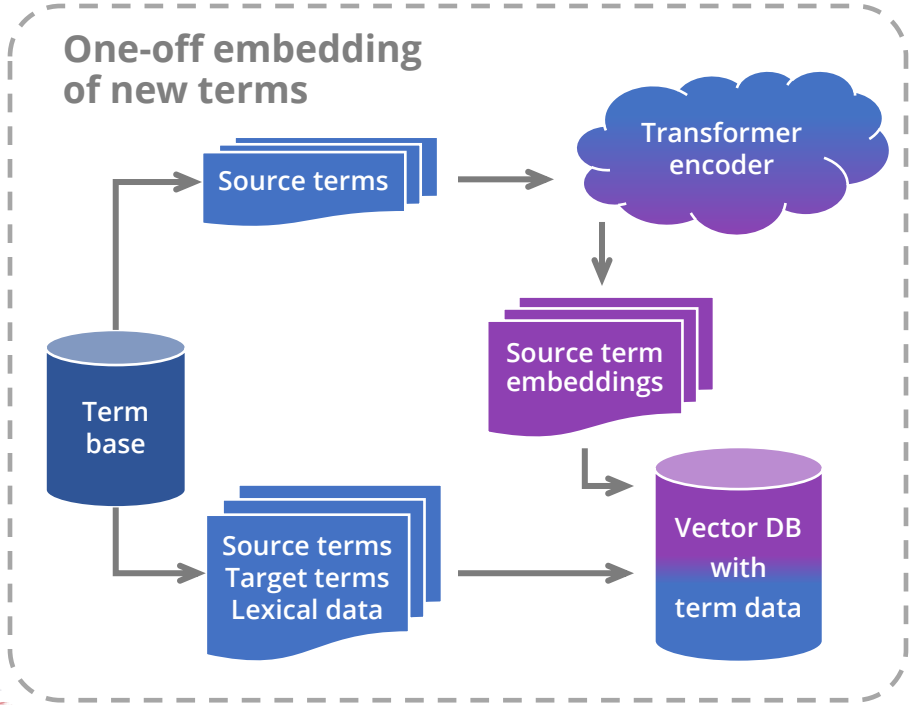
LOC WORLD

#LocWorld54 Monterey

Checking and fixing terminology with standard RAG is

slow, **costly**, and **unreliable**.

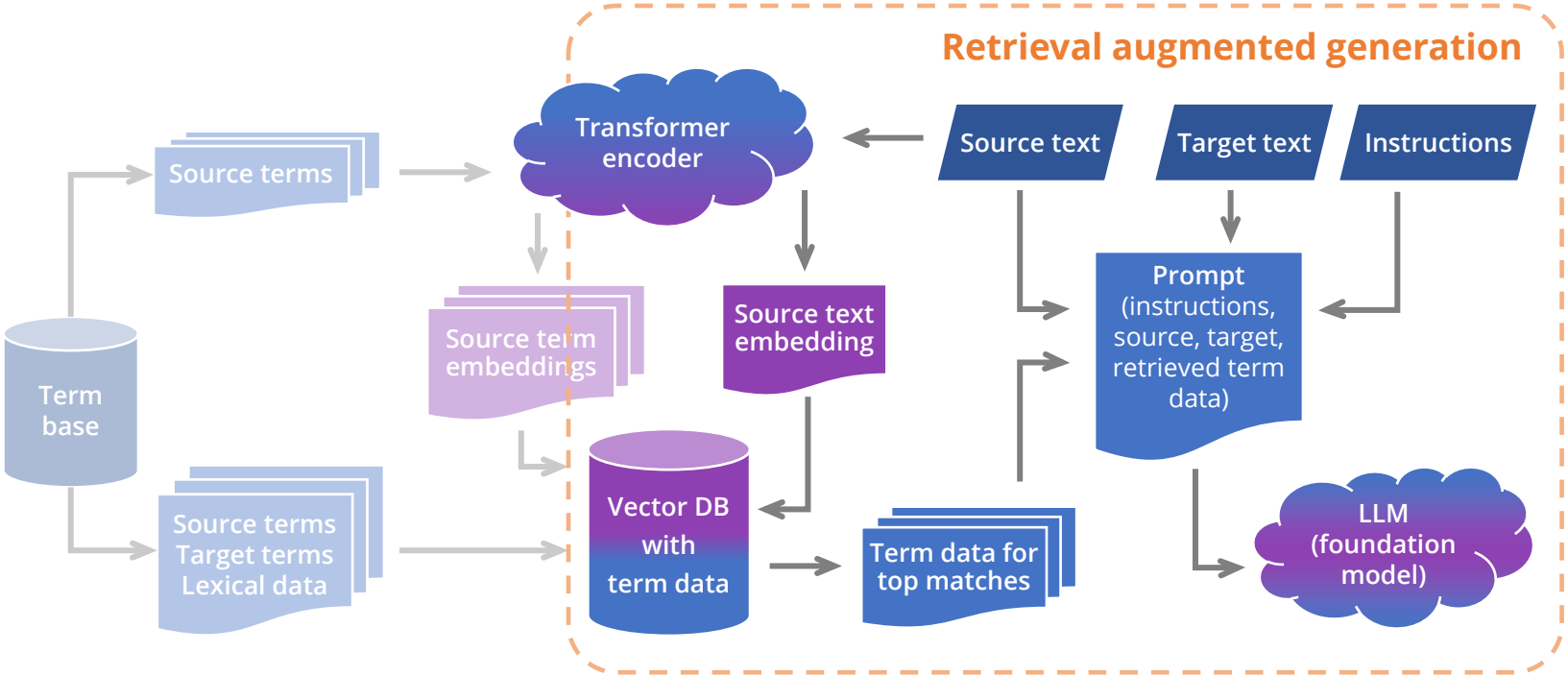
Baseline RAG for terminology



A typical example of a Retrieval Augmented Generation using a text embedding model.*

* Different concept than word embedding.

Baseline RAG for terminology



Issues with standard RAG

- **Quality:** Loosely relevant items → noisy prompts → inconsistent LLM output
- **Speed:** Retrieving + LLM for every segment is slow
- **Cost:** Consumes embedding and generation tokens for every segment

What is SPA-RAG?

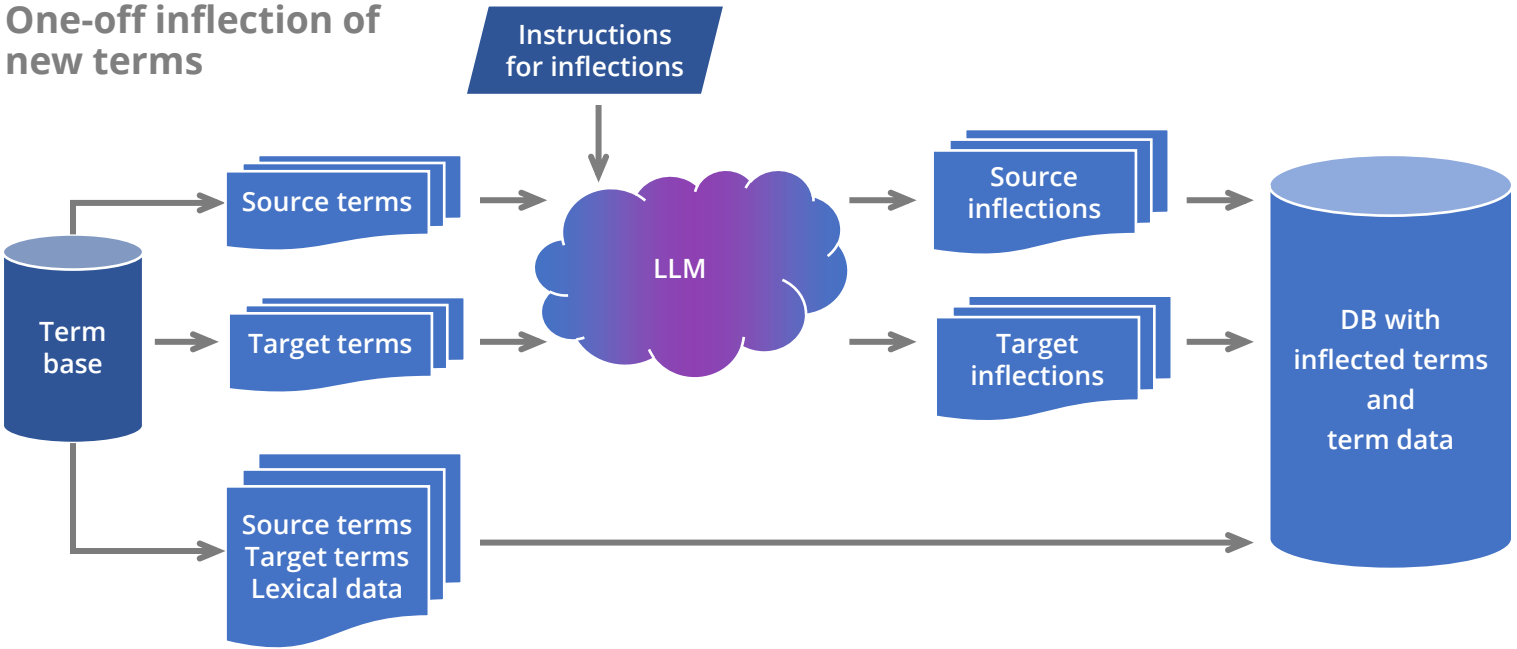
A Special-Purpose Algorithm
for
Retrieval Augmented Generation

(It can be any deterministic algorithm used to retrieve information for the generative AI step. Terminology is just an example used here.)

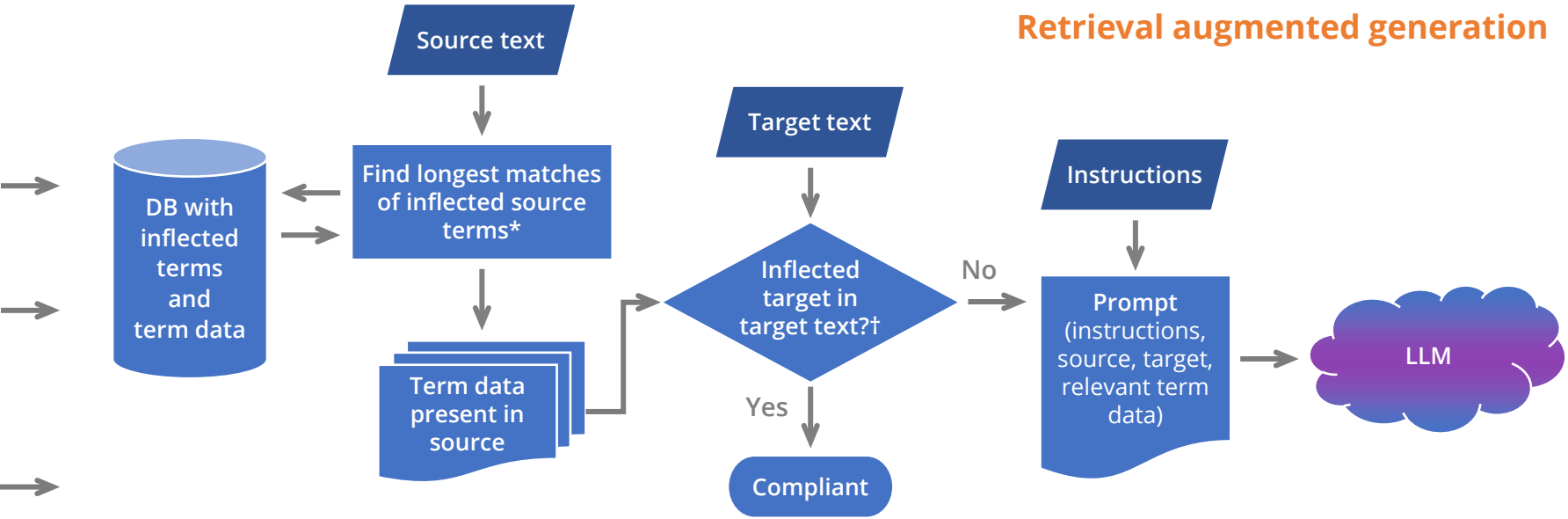
How is this SPA-RAG different?

- Adds inflections to source and target terms
- Precision retrieval of only relevant terms
- Fast deterministic filter
  (skips LLM for compliant segments)

SPA-RAG for terminology



SPA-RAG for terminology



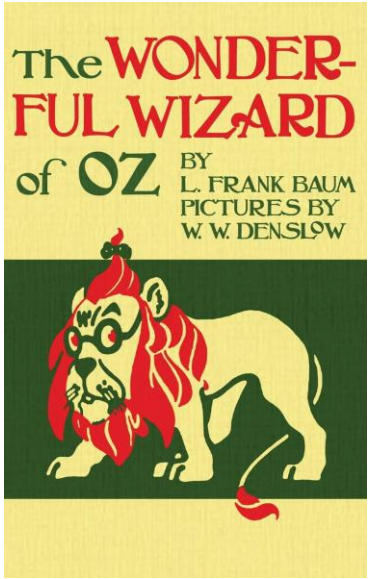
* Algorithm identifying all longest non-overlapping term matches.

† "Yes" if no term data or any inflected preferred term in target; always "No" for auto-detected polysemous terms.

Baseline RAG

v.

SPA-RAG



Test Case

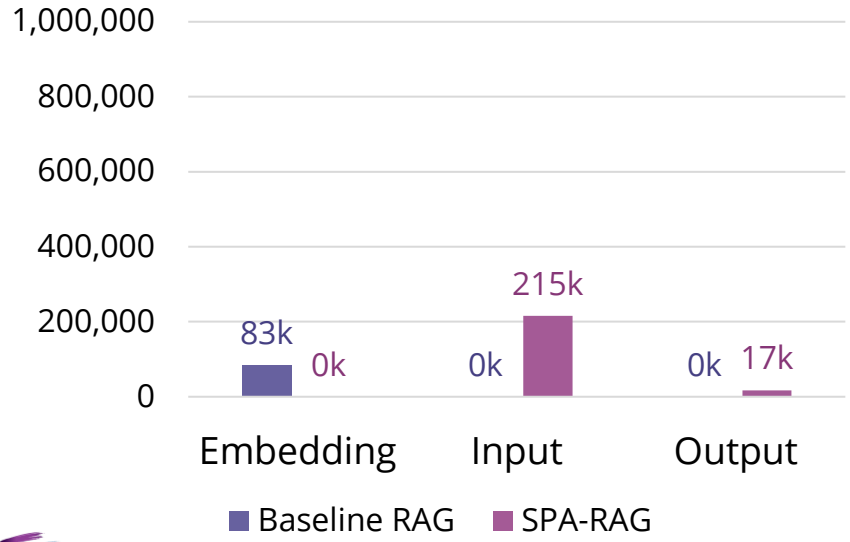
Detect and correct terminology errors in a German NMT translation, when given a glossary based on a human translation.

Glossary: **364 terms**
English: **594 words**
German: **530 words**

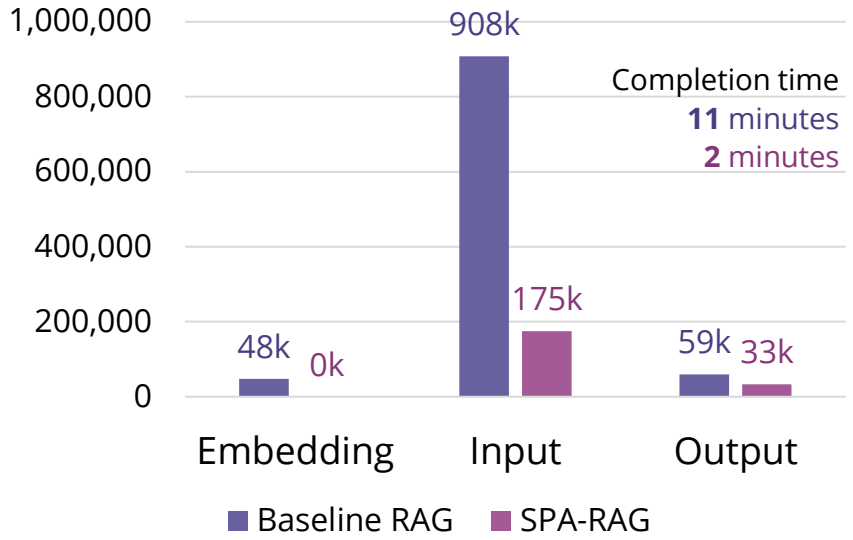
Token Usage

Document: **2,167 segments**
English: **39,502 words**
German: **36,798 words**

One-off preparation

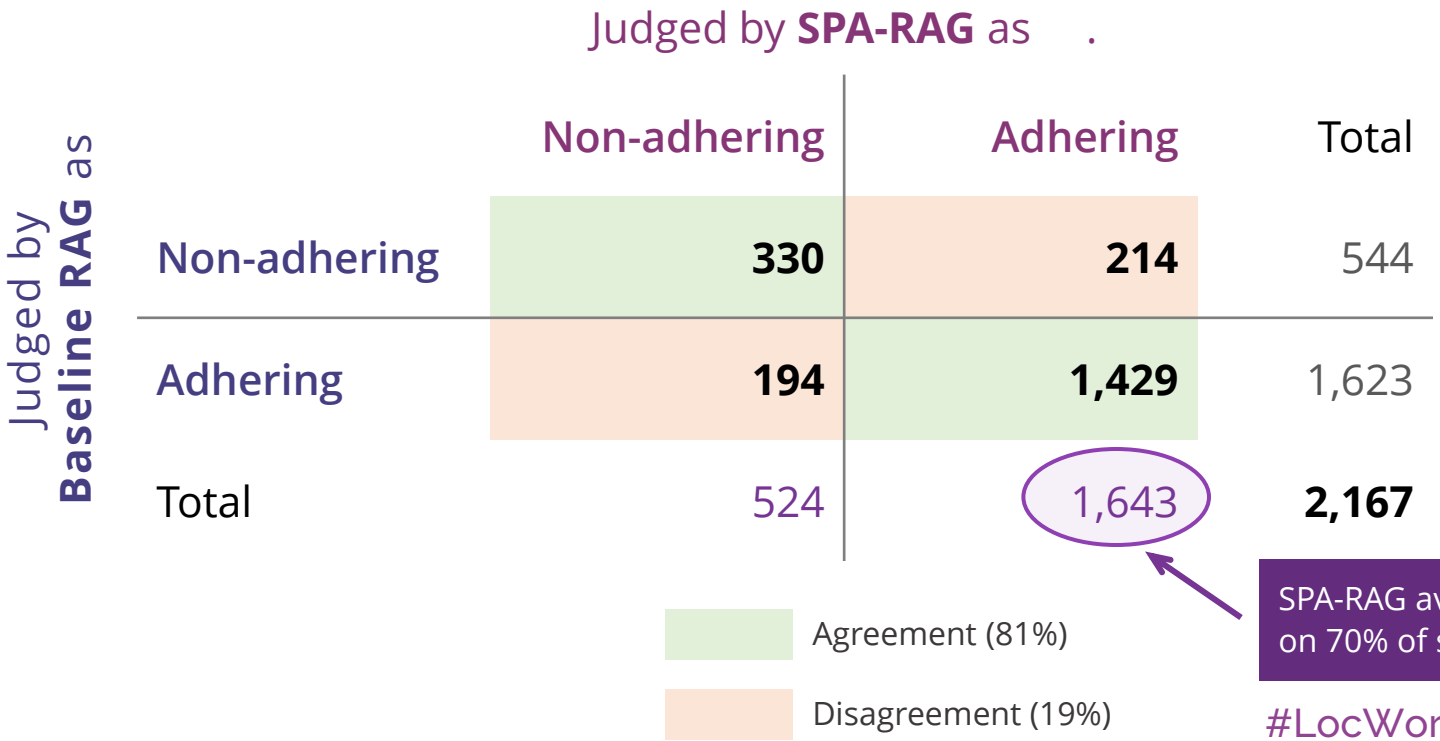


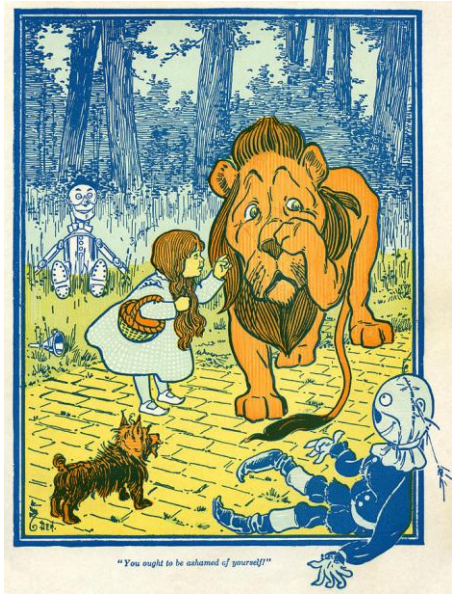
Detection and fixing



(SPA-RAG preparation could be significantly optimized with batching.)

Segment-level agreement





But which one was more **accurate**?

Let's look at **examples**.

Full results are available at:
https://github.com/psoukie/spa-rag-test-results

Baseline RAG false positives

I am fond of the Winkies, and if I could get back again to the Country of the West, …

Ich mag die Winkies, und wenn ich wieder ~~in das Land des Westens~~ ins Land der Winkies zurückkehren könnte…

Baseline RAG

land of the Winkies → *Land der Winkies*

Winkies → *Winkies*

…

SPA-RAG

LLM skipped (no errors detected)

Checked against Winkies and Country of the West.

Baseline RAG false negatives

Then she went back to the house, and having helped herself and Toto to a good drink of the cool, clear water, she set about making ready for the journey to the City of Emeralds.

Dann ging sie zurück ins Haus, wo sie sich und Toto einen kräftigen Schluck aus dem kühlen, klaren Wasser gönnte, um sich für die Reise in die Smaragdstadt vorzubereiten.

Baseline RAG

Good Witch of the North → …
Emerald City → *Smaragdstadt*
Witch of the East → …
The Wonderful Wizard of Oz → …

Dorothy → …
Glinda → …
land of the Winkies → …
castle of Glinda → …

SPA-RAG

City of Emeralds → *Stadt der Smaragde*

(also checked against house, Toto, and water)

Baseline RAG's semantic strengths

Where her lips touched the girl they left a round, shining mark,
as Dorothy found out soon after.

Dort, wo ihre Lippen das Mädchen berührten, hinterließen sie ein~~en~~ runde~~ns~~,
glänzende~~ns~~ ~~Abdruck~~ Zeichen, wie Dorothy bald darauf feststellte.

Abdruck ('imprint, physical mark')
Zeichen ('symbol, sign')

Baseline RAG:

mark of the Good Witch's kiss → *Zeichen des Kusses der Guten Hexe* (Note: protective mark on Dorothy's forehead)
Dorothy → *Dorothy* Glinda → *Glinda* mark upon your forehead → *Zeichen auf deiner Stirn* | ...

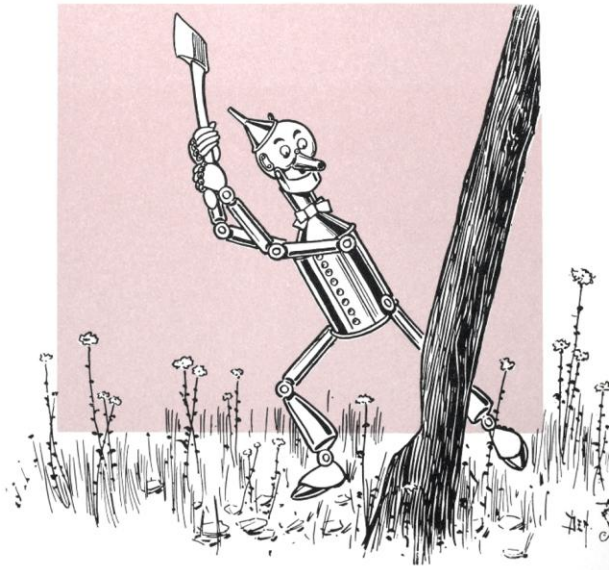
Retrievals can be too literal

Then Dorothy lost heart.

Dann verlor Dorothy ~~den Mut~~ das Herz.

(SPA-)RAG

heart → *Herz*



Takeaways

- Limit AI to where it is needed.
- Don’t dump glossaries into generic RAGs or knowledge bases.
- Invest in special-purpose solution or work with a specialist.
- Run a test and fine-tune the glossary, the retrieval, and the results.

/psoukenik

SPA-RAG

Questions & Answers

LOC WORLD

#LocWorld54 Monterey