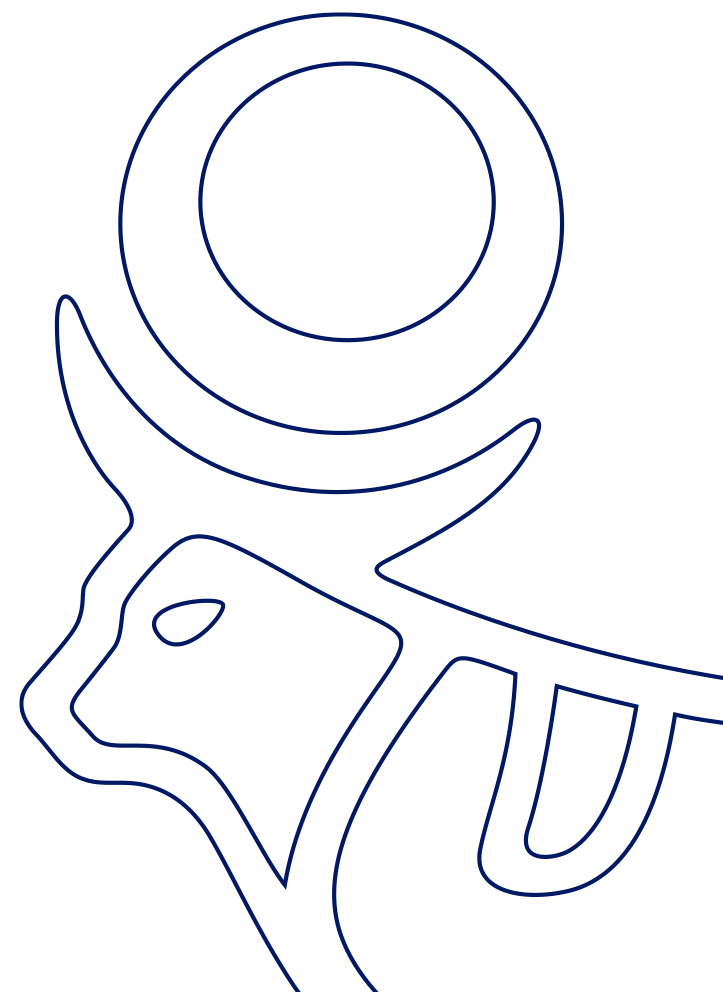


Automated Mock TFL Generation Using RAG and LLMs

Leveraging generative AI

Nicolaj Thostrup

18 November, 2025



Views and opinions expressed are those of the speaker and not necessarily Novo Nordisk

Agenda Overview

Challenges and current manual processes

Exploring automation with a RAG system

System presentation

Demonstration of automated Mock TFL generation

Summary of efficiency and accuracy improvements in clinical workflows

Streamlining Mock TFL Generation in Clinical Studies



- Mock TFLs are essential for early consensus among study stakeholders.
- They must adhere to protocols but allow flexibility for study design.
- Multiple iterations are often needed under tight timelines.
- Simpler studies may deprioritize Mock TFLs due to workload imbalance.
- Streamlined methods improve quality, alignment, and programming efficiency.

~300 outputs

Approximate outputs in a trial

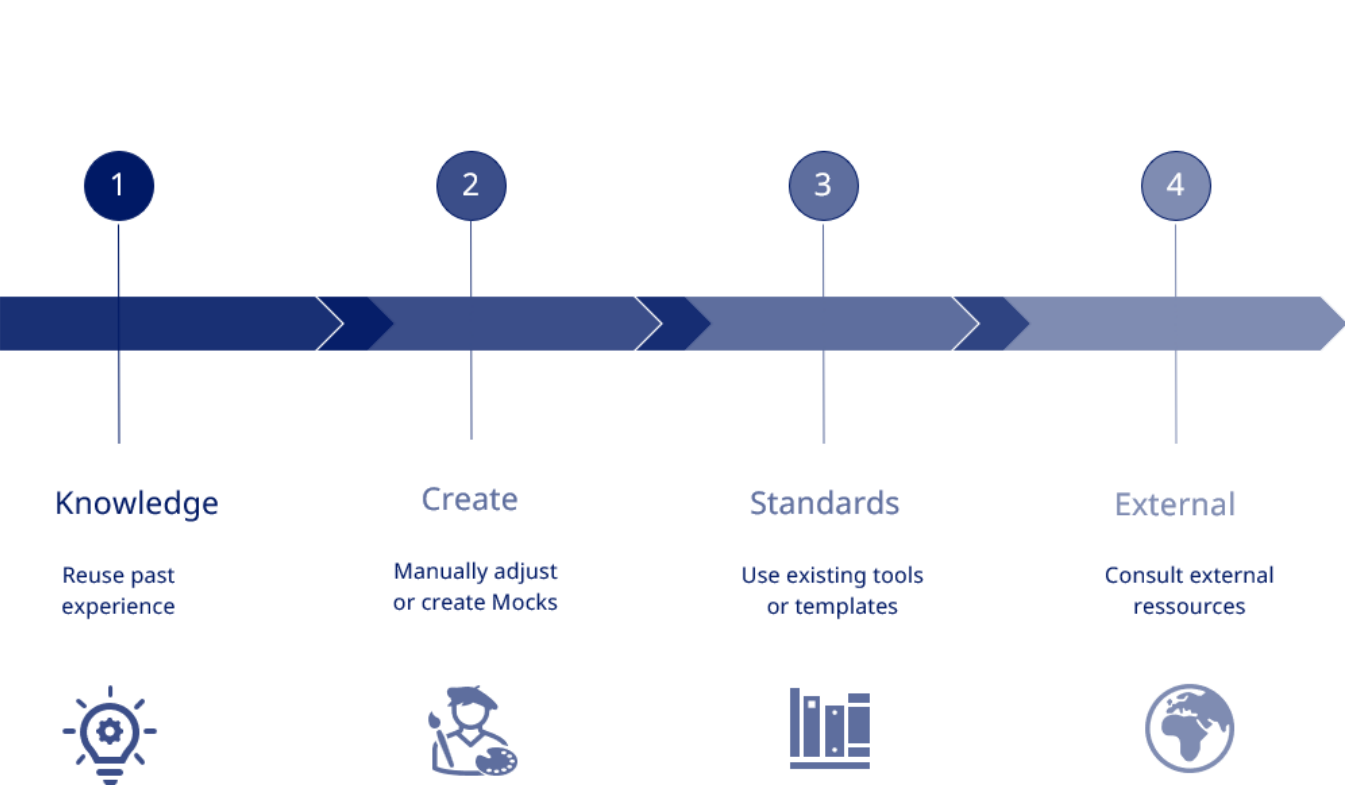
5 minutes

per Mock TFL

25 hours

of work

Current approach to Mock TFL generation

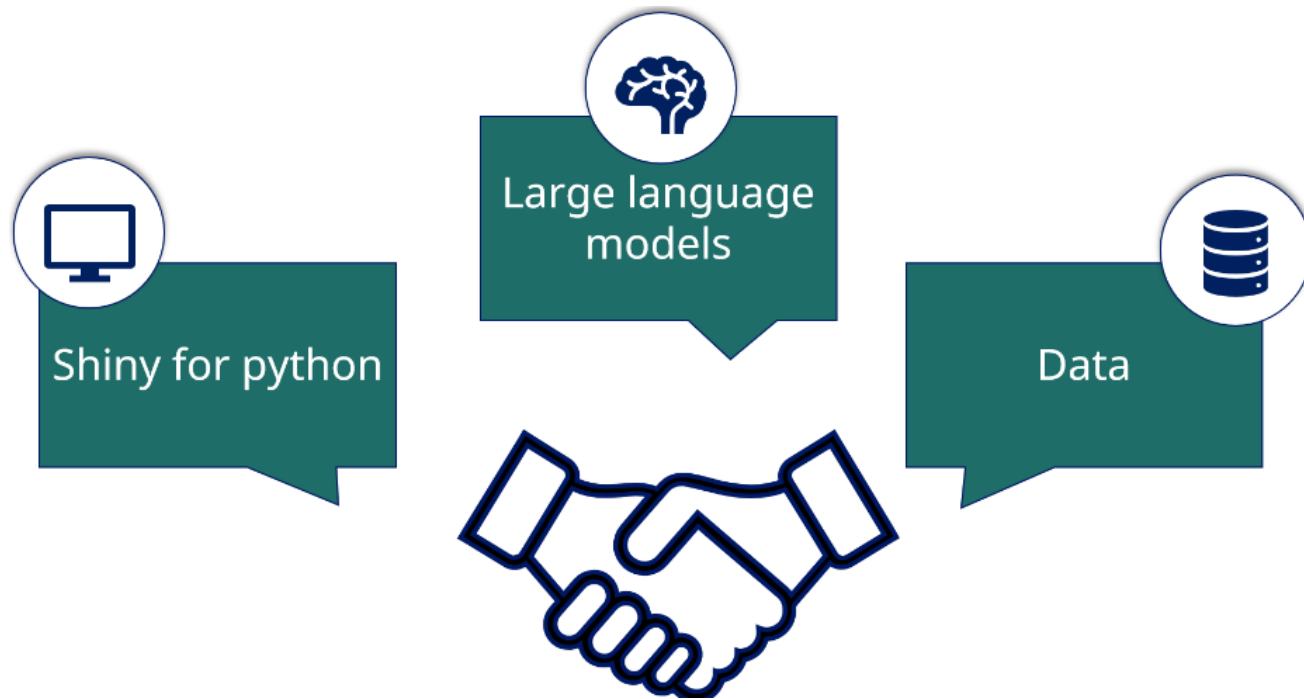


- Create Mocks manually in word or excel by reusing historical trials
- Lack of experience or knowledge within therapy area trials makes it difficult to identify relevant trials leading to inefficiencies
- Some companies might have generic templates or proprietary tools to streamline the process
- Use external sources like ClinicalTrials.gov for inspiration
- Despite tools, manual intervention remains necessary for refining formatting and structure.

500
HOURS

saved in 100 trials if 1 minute is saved per output

Addressing Inefficiencies



Generative AI can efficiently tackle immediate inefficiencies without requiring new technology breakthroughs.

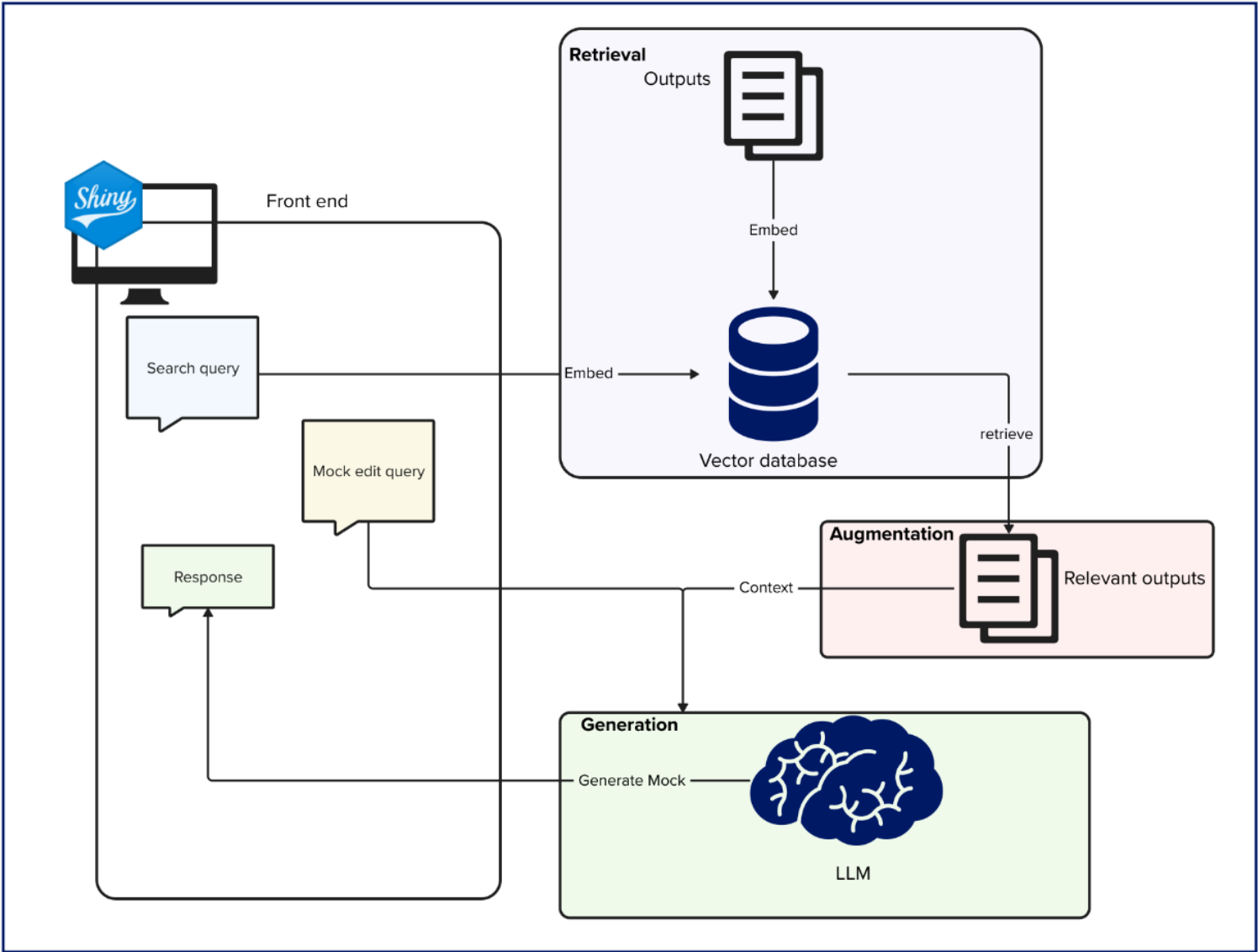
Systematic Data Use

Organizing existing data and applying established methods enables practical improvements in process efficiency.

Proof of Concept Development

Using Shiny for Python and large language models creates a PoC demonstrating automation-driven efficiency gains.

System architecture



User Input Submission

Users submit search or mock edit queries via an interactive frontend interface for processing.

Embedding and Vector Search

Input queries are converted into embeddings and used to retrieve relevant context from a vector database.

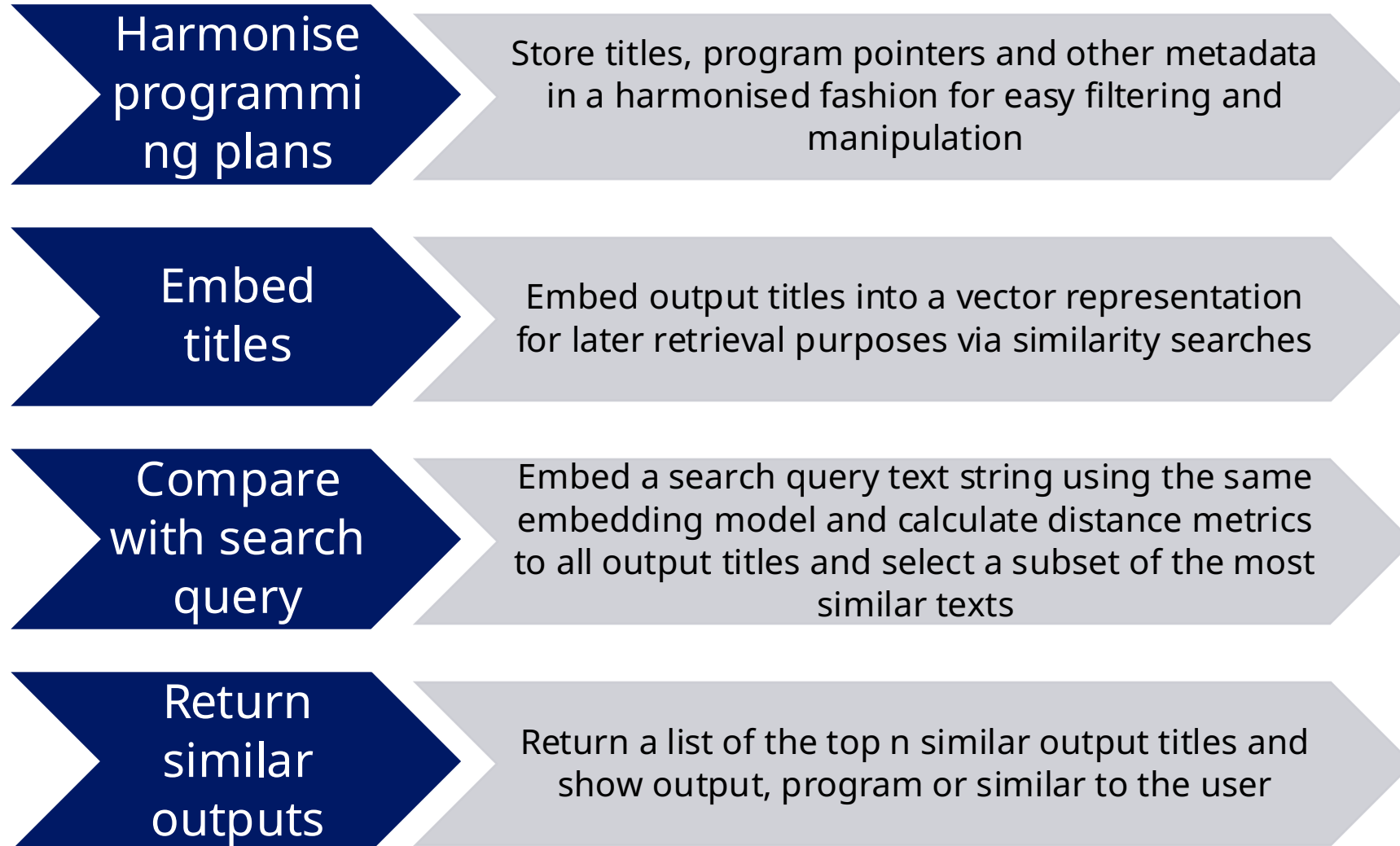
Context Injection and Mock Generation

Retrieved contextual data is fed into a language model to generate or update mock outputs automatically.

Response Display and Validation

Generated mock outputs are presented on the frontend for user validation or further editing.

Retrieval – using database of historical outputs



Augmentation and Generation

Concatenate	Obtain	Generate
Concatenate retrived outputs into one coherent context	Obtain instructions for request or edits	Task LLM with generating a Mock based on context and instructions

Examples of user-driven modifications

Adding Treatment Arms

Users can add treatment arms like 'Drug A' to customize clinical trial outputs.

Expanding Data Collection

Additional data points such as Bilirubin measurements at specific visits can be incorporated.

Renaming Visits Schema

Visits can be renamed systematically for clarity, such as 'Visit 1 day 1', 'Visit 1 day 2', etc.

Updating Headers and Footnotes

Headers and footnotes can be updated carefully for accuracy, including replacing abbreviations and aligning columns.

Search Outputs

Load Outputs

Mock generation

Settings

<

Search for outputs from title

Search text:

Baseline demographics t

Project and trial to harvest metadata from:

nnxxx-yyyy

Search

Search Results list

Search Results - Top match output

Comparative table of LLM models and their capabilities

MODEL	CONTEXT	GPQA SCORE (%)	MMLU SCORE (%)	SWE BENCH (%)	MULTIMODAL	RELEASE
GPT-5	400.000	85.7	92.5	74.9	Yes	Aug 2025
GPT-4.1	1.047.576	66.3	90.2	55.0	Yes	Apr 2025
GPT-4o	128.000	70.1	85.7	31.0	Yes	May 2024
GPT-3.5 turbo	16.385	30.8	69.8	-	No	Nov 2022
Claude Sonnet 4	200.000	75.4	-	72.7	Yes	May 2025
Deepseek R1	131.072	71.5	90.8	49.2	No	Jan 2025

Conclusion

Efficiency Improvement

Automation with RAG and LLMs significantly enhances the speed and efficiency of mock TFL generation.

Accuracy Enhancement

Using advanced models ensures higher accuracy and reliability in generating clinical trial outputs.

Transforming Workflows

Integration of these technologies promises to transform and modernize clinical trial workflows.

Thank you!

UNORDINARY
DRIVES
CHANGE



Nicolaj Thostrup
ncjt@novonordisk.com