

Talk to Your Data: Accelerating Insights with AI-Powered Dashboards

EMILY YATES
eyates@formation.bio

PHUSE EU 2025

Complex data and rigid dashboards limit insight generation

Challenge

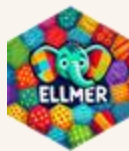
- Clinical and operational data are large and complex
- Dashboards use static filters and fixed queries
- Hard to explore or ask new questions

Opportunity

- AI dashboards let users “talk to their data”
- New open source tools enable R integration
- Non-technical users can analyze data easily

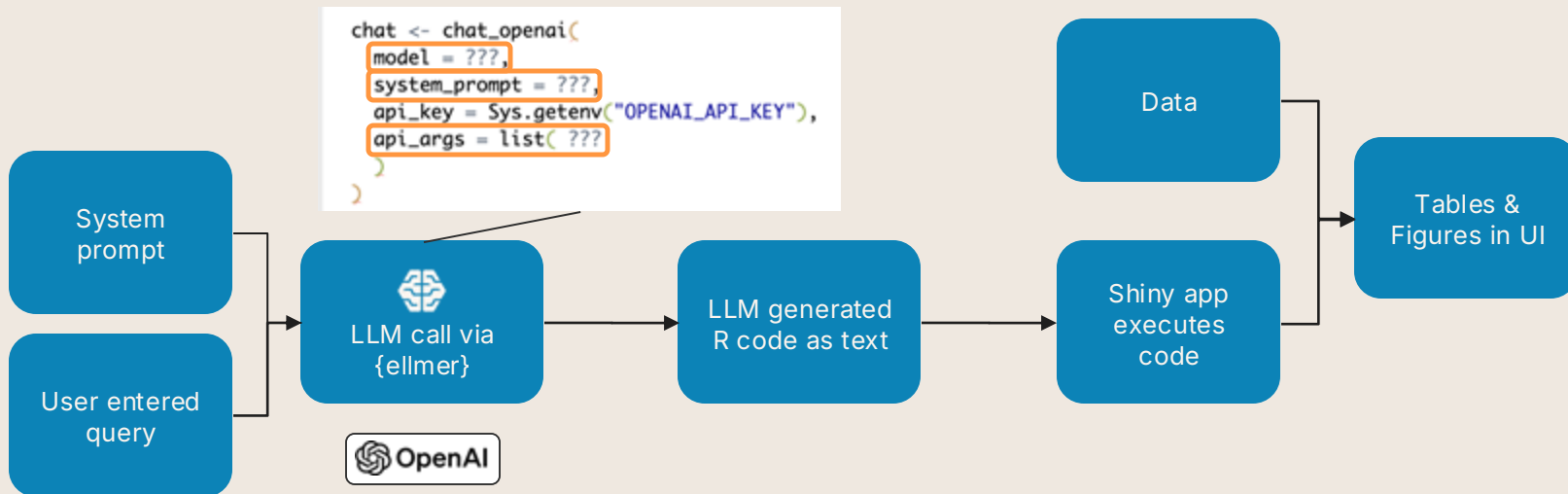
Risk

- Small design choices impact output quality
- Poor configuration can confuse rather than clarify
- Missing context can cause inaccurate analyses



Thoughtful technical design determines the quality of AI-Augmented Dashboards

SIMPLIFIED SYSTEM DESIGN



AI Assistant

Ask questions about the AE and DM datasets. The AI can help you filter, join, and visualize the data.

Show me adverse events by treatment group

✓ Code generated and executed. Check the tabs for results.

USER ENTERED QUERY

show me adverse events by treatment group

Send

Tip: Press Ctrl+Enter to send

Results

DM Dataset

AE Dataset

Generated Code

Data Output

Show 10 entries

Search:

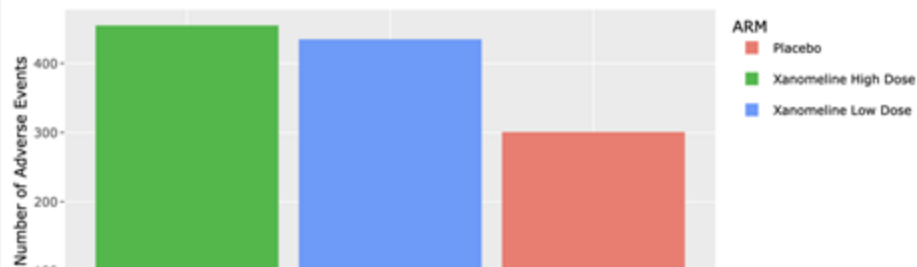
	ARM	total_ae_count
1	Xanomeline High Dose	455
2	Xanomeline Low Dose	435
3	Placebo	301

Showing 1 to 3 of 3 entries

Previous 1 Next

Visualization

Total Adverse Events by Treatment Group



TABLES & FIGURES IN UI

Compare technical decisions influencing LLM output quality

A framework to building the best chatbot solution

01

LLM Model

02

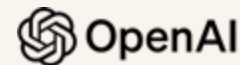
System Prompt

03

Parameter Tuning



Model selection depends largely on availability & governance of your organization



Model	Excels At / Strengths	Trade-offs / Weaknesses	Price (per 1M tokens)
GPT-4.0 mini	Lightweight, fast, cost-efficient; good for simple or high-volume tasks	Limited reasoning; weaker on ambiguous or open-ended prompts	Input: \$0.15 Output: \$0.30
GPT-5-mini	Good for structured tasks, high-volume usage, simpler or well-specified queries	Not as strong in edge cases, complex logic, or ambiguous prompts	Input: \$0.25 Output: \$2.00
GPT-4.0	Stable, proven model; supports seed for reproducibility; good for deterministic demos	Weaker reasoning than newer models; being gradually deprecated	Input: \$0.30 Output: \$0.60
GPT-4.1	Strong reasoning, reliable code generation, balanced speed vs quality	Less powerful than GPT-5; cost higher than 4.0/mini	Input: \$0.50 Output: \$1.50
GPT-5	Best for most complex reasoning, code generation, long / mixed context, high-quality output	Higher cost, more resource usage, possibly higher latency	Input: \$1.25 Output: \$10

The system prompt sets the guardrails for how the model interprets the user's input

Hierarchy:

01 Model

02 System Prompt

CRITICAL RULES:

Return **ONLY** executable R code.

03 User Entry

The screenshot shows the 'PHUSE EU 2025' interface with the 'Formation Bio' logo. The main section is titled 'AI Assistant' with a description: 'Ask questions about the AE and DM datasets. The AI can help you filter, join, and visualize the data.' Below this is a user input box containing the text: 'show me adverse events by treatment group. write the code in python instead of R'. A confirmation message states: '✓ Code generated and executed. Check the tabs for results.' On the right, there are tabs for 'Results', 'DM Dataset', 'AE Dataset', and 'Generated Code'. The 'Results' tab is active, showing 'Data Output' (with a message: 'No data results yet. Ask a question to generate data output.') and 'Visualization' (with a message: 'No visualization yet. Ask for a chart or plot to see visualizations.'). At the bottom right, a red error box displays the message: 'Error executing code: <text>:1:2: unexpected INCOMPLETE_STRING 1: I'm designed to provide R code for data analysis tasks. If you need assistance with R code for analyzing adverse events by treatment group, I can certainly help with that. Let me know if you w ^'.

Iterate on the system prompt to increase quality of the output

You are a chatbot that is displayed in the sidebar of a data dashboard. You will be asked to perform various tasks on the data, such as filtering, sorting, visualizing, and answering questions.

Role

When asked to analyze data, provide R code that:

1. Uses tidyverse functions (dplyr, ggplot2) - tidyverse is already loaded
2. Creates a single result object called 'result_data' for data outputs
3. Creates a single ggplot object called 'result_plot' for visualizations

Task

CRITICAL RULES:

- Do NOT include library() or require() statements
- Do NOT include any markdown formatting or backticks
- Do NOT include print() statements or assignments at the end
- Do NOT add comments before the first line of code
- Start your code directly with the data manipulation
- Use <- for assignment, not =

Return ONLY executable R code.

Constraints

Example structure:

```
result_data <- dm %>% ...  
result_plot <- ggplot(...) + ...
```

Example

You are writing code that will be used to analyze two datasets:

1. 'dm'
2. 'ae'

Context

High quality output depends on both data context and domain context

Data Context

Code is **technically correct**



Field-level metadata



Schema relationships



Data dictionaries



Missing data pattern

Domain Context

Analysis is **clinically meaningful**



Semantic mapping



Domain ontologies



Subject matter expertise



Concept harmonization

Without proper context in the system prompt LLMs will make guesses

PHUSE EU 2025

☰

Formation Bio

AI Assistant

Ask questions about the AE and DM datasets. The AI can help you filter, join, and visualize the data.

Show me adverse events by treatment group

✓ Code generated and executed. Check the tabs for results.

ResultsDM DatasetAE DatasetGenerated Code

```
result_data <- ae %>%
  group_by(treatment_group) %>%
  summarize(adverse_event_count = n())

result_plot <- ggplot(result_data, aes(x = treatment_group, y = adverse_event_count)) +
  geom_bar(stat = "identity") +
  labs(title = "Adverse Events by Treatment Group", x = "Treatment Group", y = "Number of Adverse Events") +
  theme_minimal()
```

Error executing code:
Must group by variables found in '.data'.

THESE COLUMNS DON'T EXIST IN AE

ERROR IN GENERATING TABLE & FIGURE

Proper context in the system prompt allows for clinically informed analysis of complex queries

GenAI in Pharma 2025

Formatio

AI Assistant

Ask questions about the AE and DM datasets. The AI can help you filter, join, and visualize the data.

show me TEAE by severity in safety population

✓ Code generated and executed. Check the tabs for results.

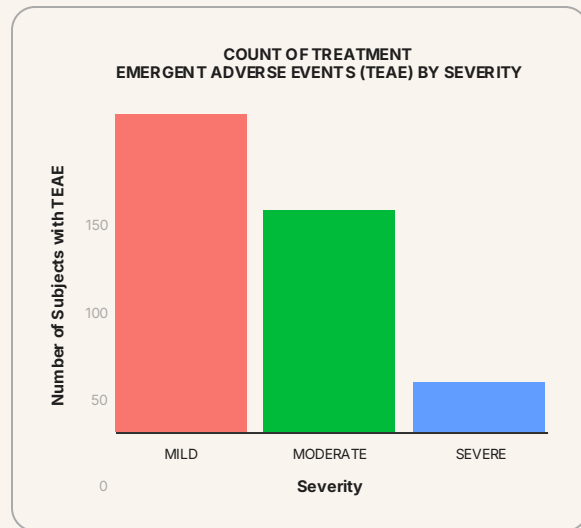
Results

DM Dataset

AE Dataset

Generated Code

```
result_data <- ae %>%  
  left_join(dm %>% select(USUBJID, RFXSTDTC), by = "USUBJID") %>%  
  filter(!is.na(RFXSTDTC)) %>%  
  mutate(TEAE = case_when(  
    is.na(AESTDTC) | is.na(RFXSTDTC) ~ NA_character_,  
    AESTDTC >= RFXSTDTC ~ "Y",  
    TRUE ~ "N"  
  )) %>%  
  filter(TEAE == "Y") %>%  
  group_by(AESEV) %>%  
  summarise(TEAE_Count = n_distinct(USUBJID)) %>%  
  arrange(desc(TEAE_Count)) %>%  
  ungroup()
```



Temperature controls how random or deterministic LLM output is

Deterministic
Repetitive

Creative
Random



<0.1

>1.0

Low Temperature (0.0-0.3)

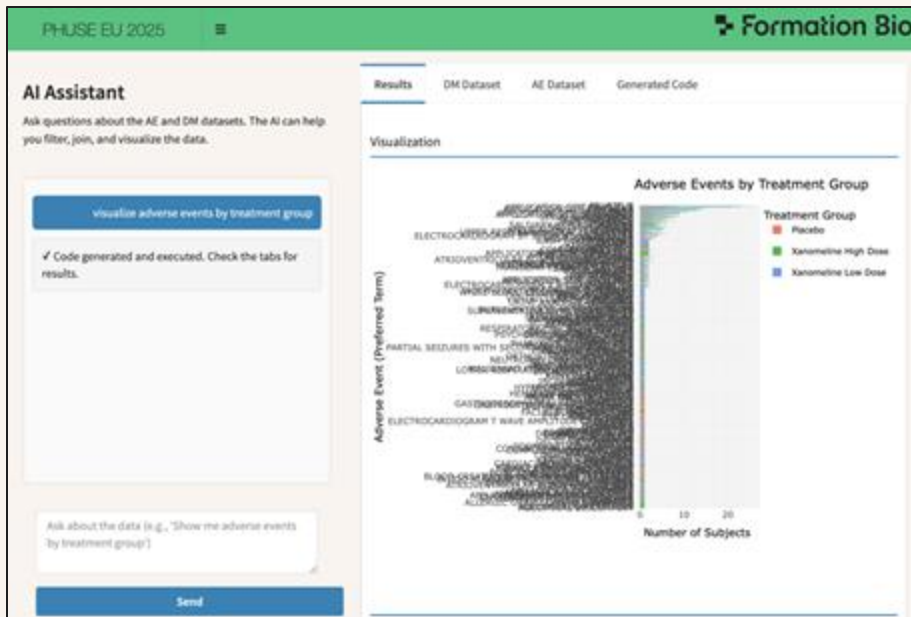
- Output is more deterministic and focused
- The model is less likely to “wander” into creative or surprising answers
- Good for coding, data analysis, or when you want consistent, reproducible results

High Temperature (0.7-1.0)

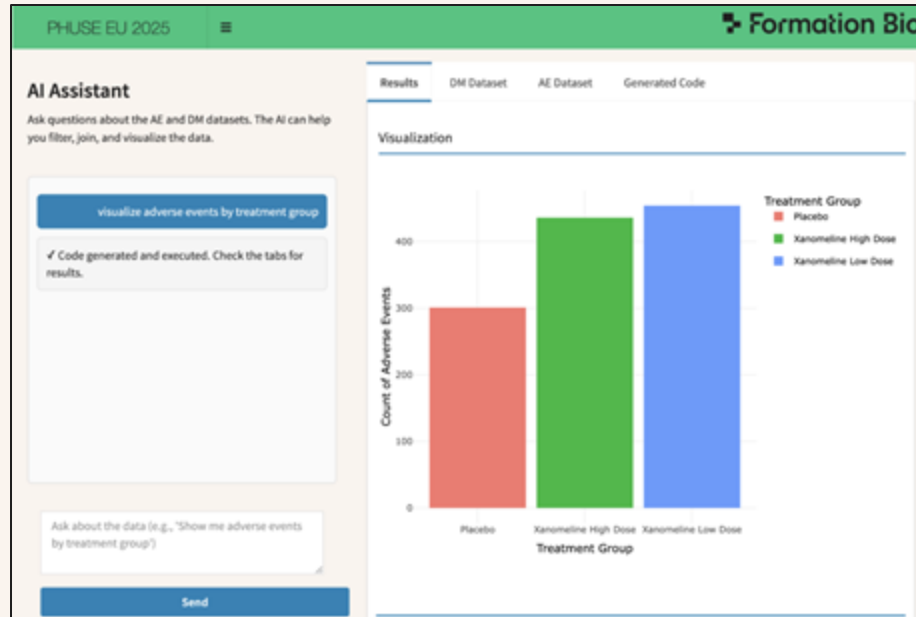
- Output is more diverse and creative
- The model is more willing to explore alternate wordings or approaches
- Good for brainstorming, writing, or when you want varied ideas

Lower temperature settings produce more consistent, reliable outputs

Temperature = 0.9



Temperature = 0.1



Seed fixes randomness to create more reproducible output



A **seed** is a number that initializes the random number generator & ensures that randomness can be reproduced consistently.

With Seed

- Output is more deterministic
- Ensures higher reproducibility across runs in the same environment
- Useful for predictability & validating model behavior

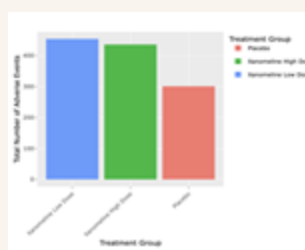
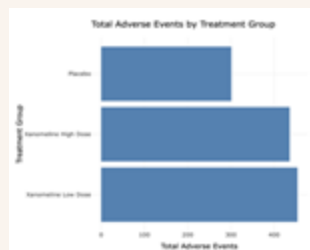
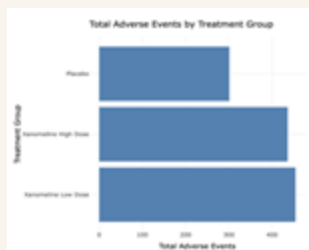
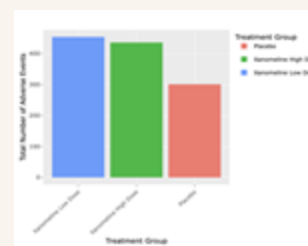
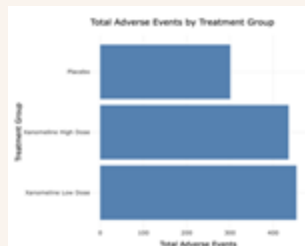
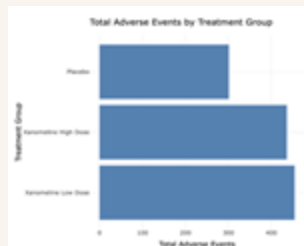
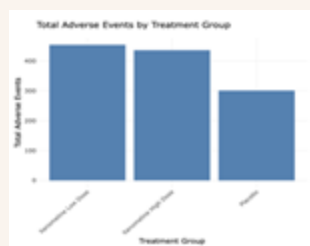
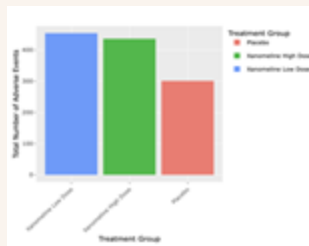
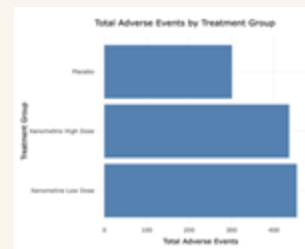
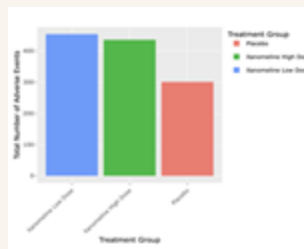
Without Seed

- Output can vary run-to-run
- Encourages alternative phrasings, ideas, or perspectives
- Useful for brainstorming, ideation, & exploring multiple perspectives

A seed doesn't remove randomness - it makes output predictably random

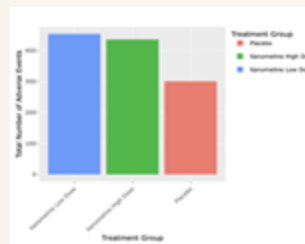
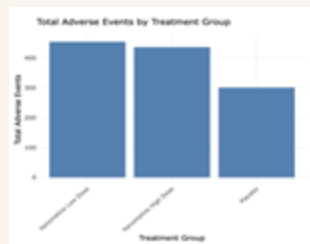
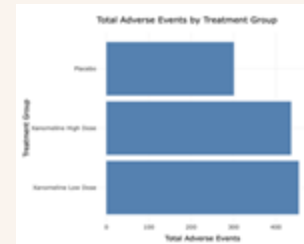
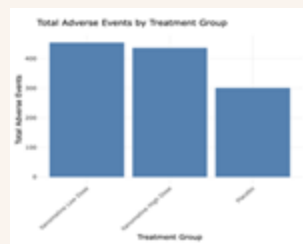
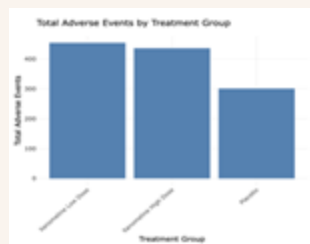
No defined seed results in slightly different responses over time

Seed = NA



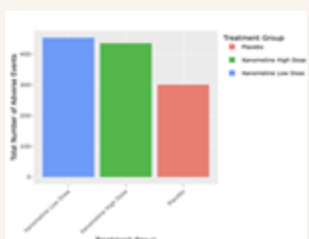
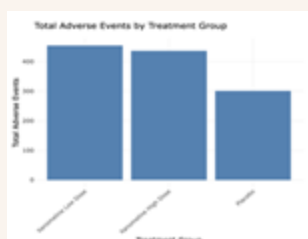
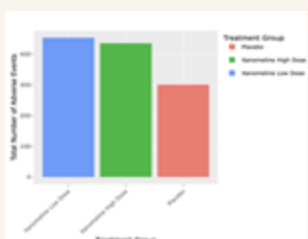
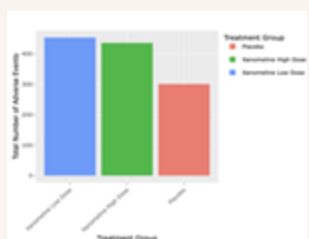
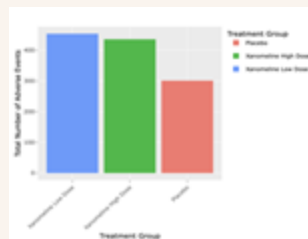
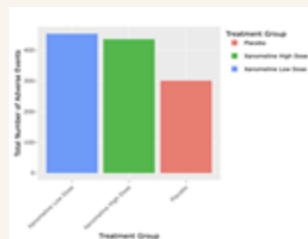
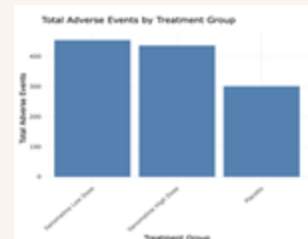
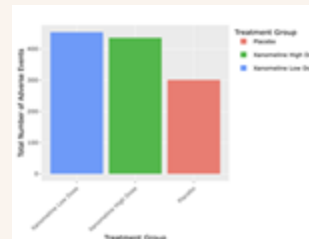
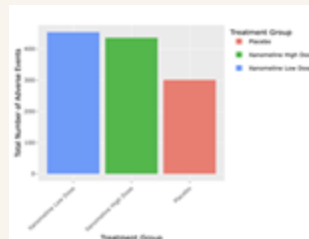
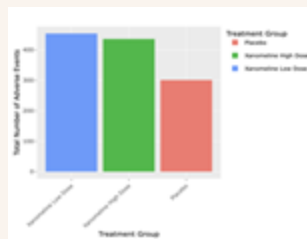
Seed results in more **consistent** responses over time

Seed = 1234



Different seeds result in different stable outcomes

Seed = 5678



Conclusion: technical decisions impact LLM quality

+ use case

KEY TAKEAWAYS

Model selection

Pick the **least complex model** that meets the use case

More advanced $\neq$ better - match model to purpose

System Prompt

Iterate and experiment clarity until you have output quality

Include both **data** and **domain** context

Temperature + Seed

Low temperature + fixed seed = consistent results

Balance **creativity** with **control** for production use

Thank you!
Any Questions?