

Using AI to find veterinary data sources in Europe

Join our learning curve!

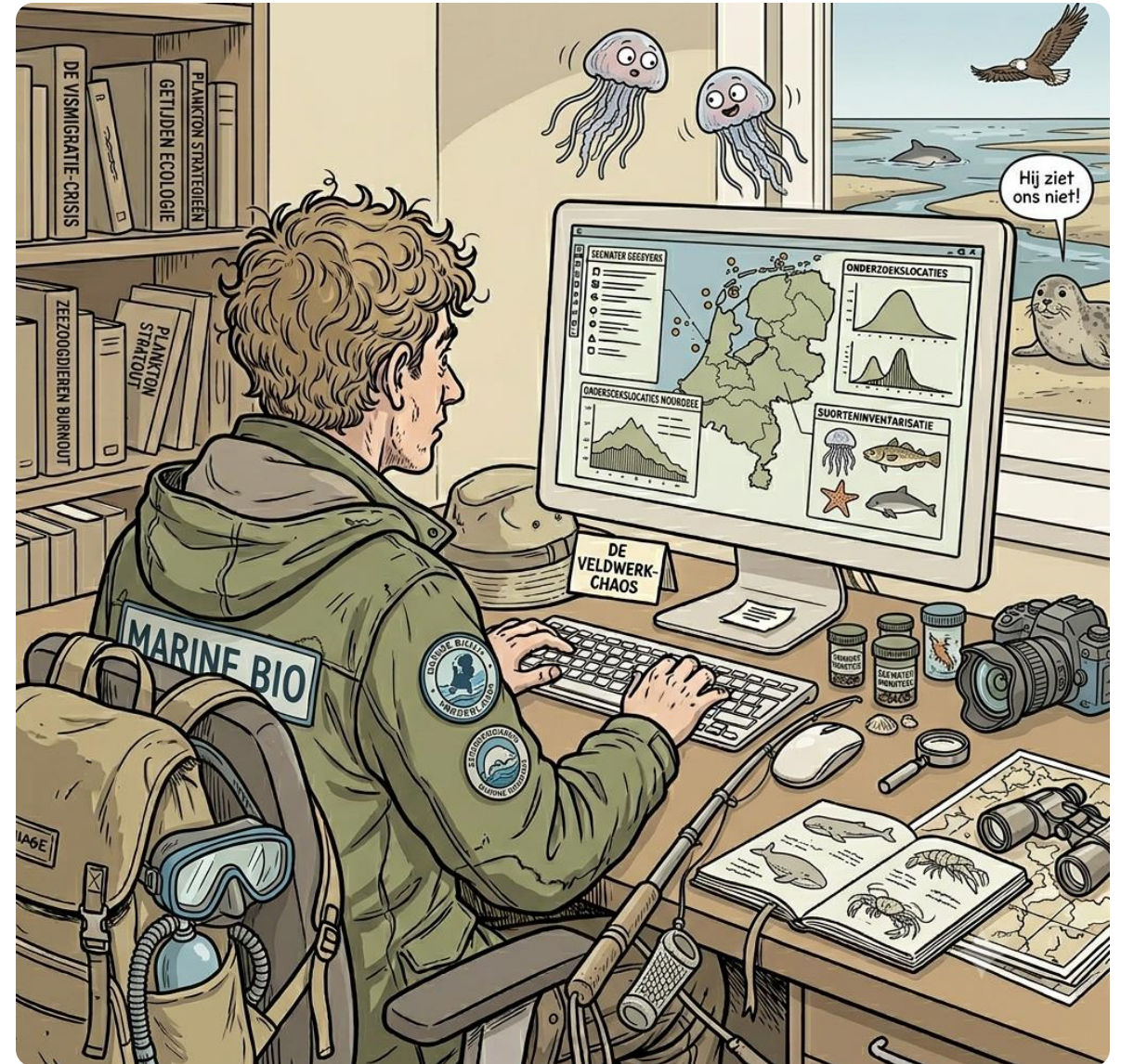
30-06-2026

Ronald Petie, Valerio Bianchi, Ewa
Pacholewicz

Learning curve?

From biologist to modeler/data scientist

This image is AI generated



Goal

- Find veterinary data sources.
- Details
 - European Medicine Agency (EMA)
 - Continuation project
 - EU + Iceland, Liechtenstein and Norway.
 - 10 data sources per country

General approach

Process in R, using gemi.R package.
NB: this is NOT the web browser but the API!



Prompt Structure

Context

- "You are a veterinary expert ..."

Specific task

- "List veterinary institutes ... "

Constraints and requirements

- "When you working, accuracy and repeatability are ..."

Examples

- "The types of institutes to include are: veterinary ..."

Desired outcome

- "Your output should be a table with the headers ..."

Prompt keywords

- Accuracy, repeatability
- Examples:
 - "Institutes to include/exclude are ..."
- Being specific:
 - "A "veterinary data source" is any record, information system, or ..."
 - "Country: the ISO 3166 alpha-3 three letter code for the country ..."
- Shouting:
 - "Do NOT add or remove columns or change column names."
 - "You MUST use these exact values for ..."

Iterate through countries

```
5 You are a veterinary expert from the Netherlands. You are working on a project with the aim
6
7 ## Specific task
8
9 I would like you to make a list of Veterinary institutes for the countries listed below:
10
11 country_names.
12
13 ## constraints and requirements
14
15 When you working, **accuracy** and **repeatability** are the most important values. When you
16
17 ## Examples
18
```

- Read Prompt (markdown file)
- Replace placeholder
- Call `gemini_search()` function, use updated prompt as input
- Collect response

Gemini API

Google AI studio

- <https://aistudio.google.com/app/apikey>
- Use paid tier!

Projects

Understanding projects

+ Create a new project

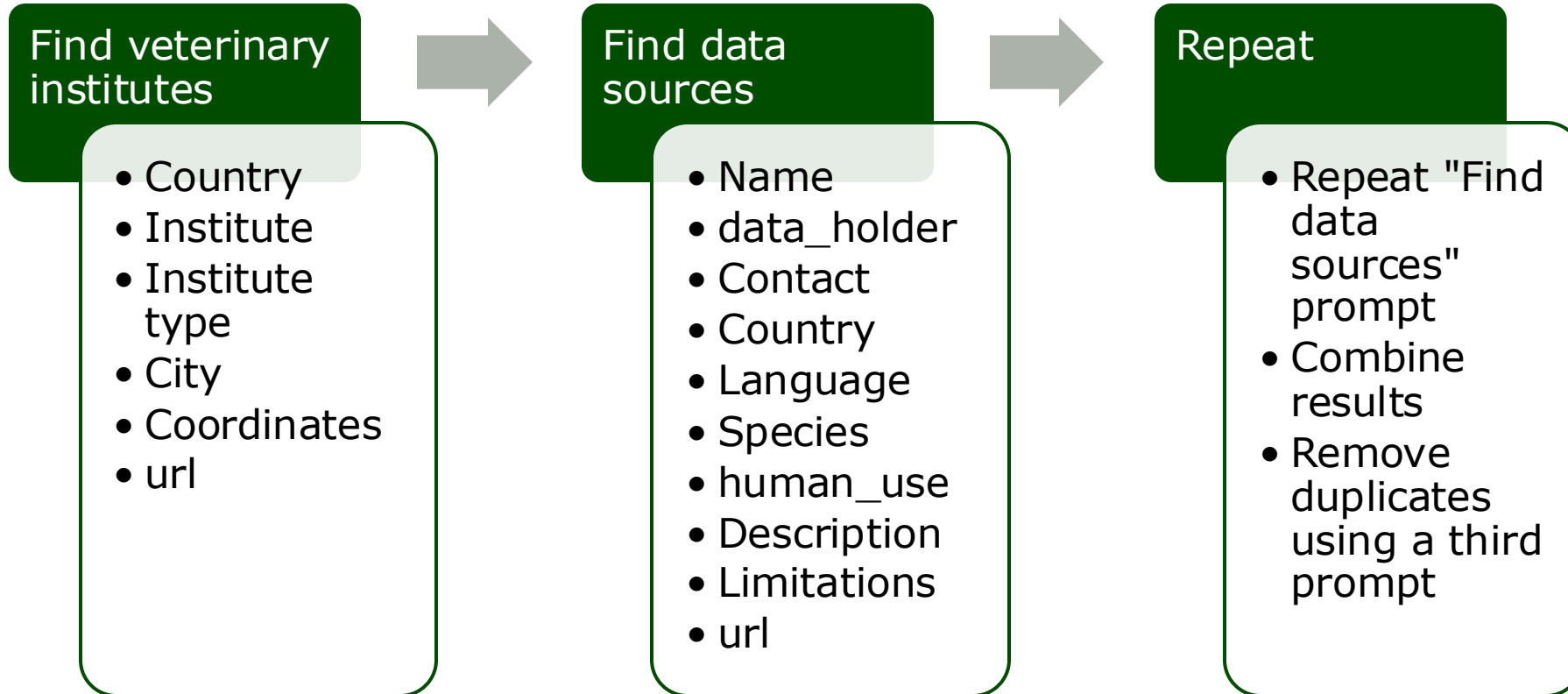
Import projects

Search for a project

Project	Keys	Created	Billing Tier	Status
vet-database-search gen-lang-client-0710:	1 key	Oct 16, 2025	My Billing Account Tier 1 · Postpay	   



Institute-based search



Broad search

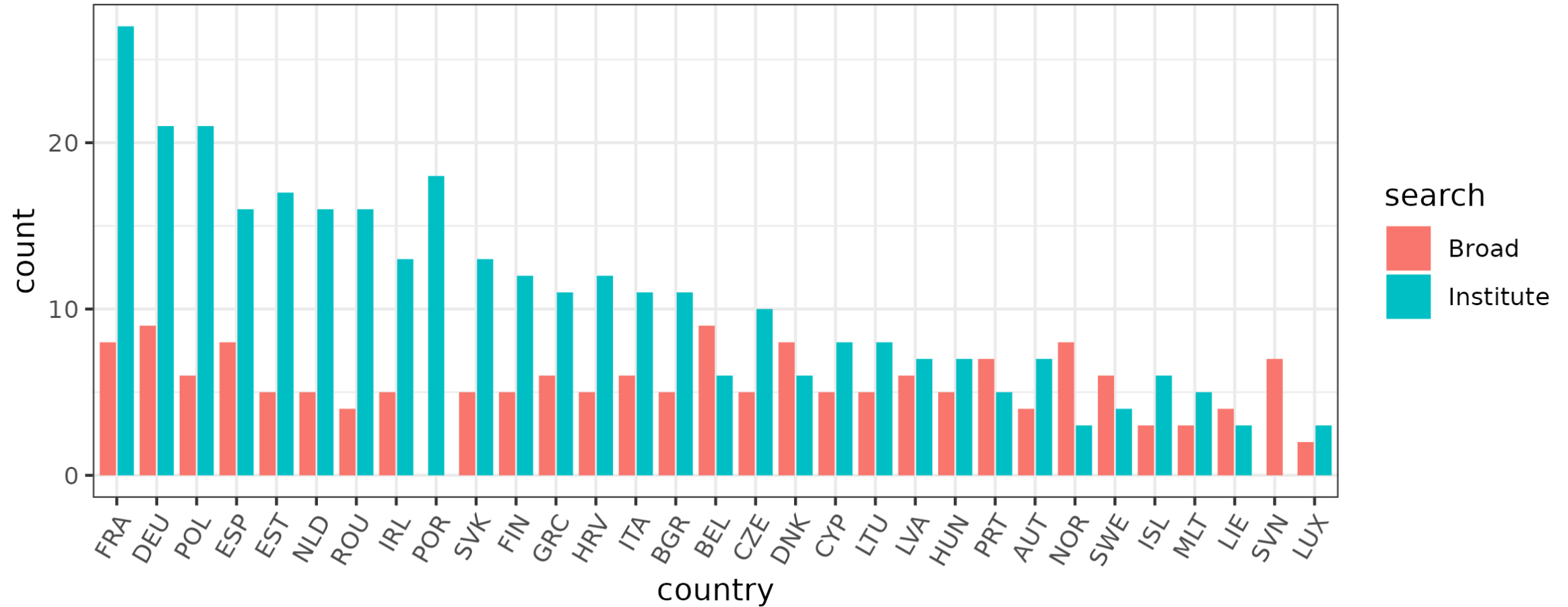
Broad search

- Country
- Data_holder
- Name
- Data_type
- Url
- Certainty
- description

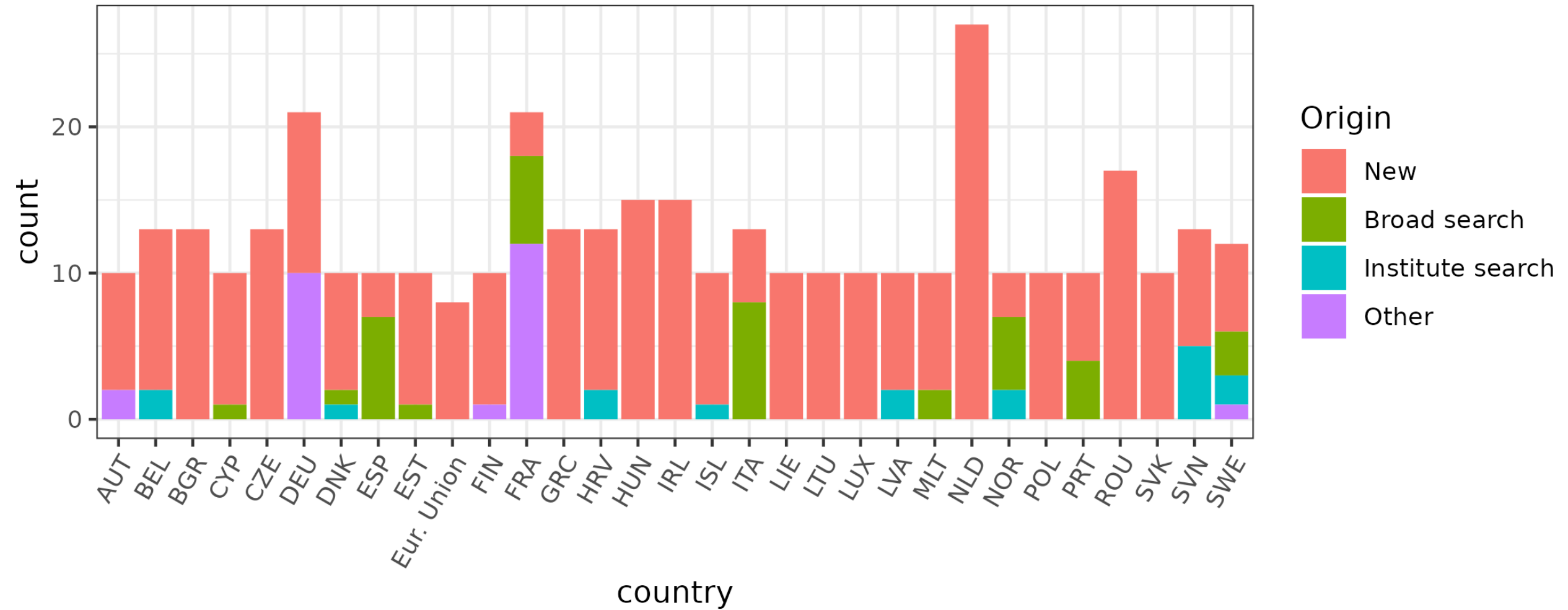
Experience thus far

- Output table could vary.
 - Therefore, switched to `gemini_structured()` function.
 - JSON schema dictates output.
 - No **grounding**!
- **Crashes** occur when asking for conflicting things, like being **accurate** & listing **at least** 10 data sources
- **Good info**: data holder, description & name of data source.
- **Really, really bad info**: URLs.
 - Base URL generally OK
 - Rest hallucinated
 - Google search **grounding** failed

Institute VS broad search



Data source origin



Prompt for web-interface Gemini

Danish Veterinary Data Sources

Please search in Danish for sites containing Danish data sources for the categories given below. Return your answer in English and return at least one link per category. The context for all categories MUST be "veterinary medicine". Return your answer in table for...

Denkproces tonen

Denmark has a robust digital infrastructure for veterinary and agricultural data, primarily managed by the **Danish Veterinary and Food Administration** (Fødevarestyrelsen) and **SEGES Innovation**.

The following table provides data sources and portals for veterinary medicine categories in Denmark.

Danish Veterinary Data Sources

Category	Data Source / Portal Name	Description	Link
Antimicrobial resistance	DANMAP	The Danish Programme for surveillance of antimicrobial consumption and resistance in humans, animals, and food.	danmap.org

Conclusions

- LLM prompt-based tasks can be automated using, for instance, the gemini.R package.
- Use the gemini_structured() function when you want reliable table output.
- Use the paid tier for speed.
- LLMs best suited for classification or summary tasks.
- Functioning URLs need **grounding** to prevent hallucinations.
 - Call Gemini API directly and set "tools" to "google-search".
 - **"Tavily"** is the real-time search engine for AI agents and RAG workflows"

Thanks! Questions?

Ronald Petie

Valerio Bianchi

Ewa Pacholewicz



JSON for gemini_structured()

```
140 ## JSON schema
141
142 ```{r}
143 vet_data_schema <- list(
144   type = "ARRAY",
145   description = "A list of identified public and private veterinary data sources associated with the specified country.",
146   items = list(
147     type = "OBJECT",
148     properties = list(
149       country = list(
150         type = "STRING",
151         description = "The English name of the country, for instance 'Austria'."
152       ),
153       data_holder = list(
154         type = "STRING",
155         description = "The entity that collects, processes, controls, or holds the data (e.g., government agency, univer:
156       ),
157       name = list(
158         type = "STRING",
159         description = "The primary name of the data source. Multiple names separated by '/'. Former names in parentheses:
160     ),
```