

PROMPT BATTLE ROYALE: SCIENTIFIC DATA EDITION

*Compete to design the most
powerful prompts for extracting and
prioritising data*



ADVENTURES
IN PHARMA

Where are we focusing today?



Today, we're designing prompts that produce a single, consistent output eg. upload one PDF (ie. a scientific paper) into ChatGPT to get one label or decision or 1-sentence summary.



Once built, this prompt can be applied at scale across multiple documents and tools to generate consistent, structured outputs that can be easily compiled into tables for analysis. This is useful for analysing larger datasets, such as congress abstracts, publications, clinical trials, HCP insights, competitor insights and more...

This type of prompting assumes you have the source files/dataset already. Searching for data is outside of the scope of this workshop

How can AI help us with scientific data mining?

1

AI TO CATEGORISE (TAG WITH METADATA)



2

AI TO PRIORITISE (IDENTIFY PRIORITY PAPERS)



3

AI TO SUMMARISE (BUILD SLIDES ON PRIORITY PAPERS)



WITH HUMAN REVIEW FOR ACCURACY*

***ALWAYS CHECK AI OUTPUTS**



Tips for advanced prompting for scientific data mining

AI can be powerful, but only if you tell it exactly how to think

1 TASK SPECIFICITY  Clearly define what you want the model to do <i><u>Categorise, prioritise, summarise</u></i>	2 CONSTRAIN THE OUTPUT  Limit what the model can return, and tell it what you explicitly DO NOT want <i>Fixed labels, structured fields, consistent format</i>	3 EMBED YOUR EXPERTISE  Turn your judgement into rules <i>Inclusion criteria, grey areas, decision logic</i>	4 PROVIDE CONTEXT  Tell the model what matters <i>Drug, population, outcomes, study types</i>
5 HANDLE UNCERTAINTY EXPLICITLY  Tell the model what to do when information is missing <i>"If unclear, label as..."</i>	6 OPTIMISE FOR REPEATABILITY  The same prompt should give the same structure every time <i>So outputs can be compared, copied, and scaled</i>	7 DESIGN FOR USE  Make results directly usable in your workflow <i>Tables, single labels, concise summaries</i>	

Prompting to prioritise



Based on the article content, categorise each study as “Include” or “Exclude” according to the following rule:

- Include if the study presents clinical efficacy and/or safety data for Product X, either directly (eg., Phase 1–3, real-world, case reports) or indirectly (eg., meta-analyses or reviews summarising clinical outcomes of Product X).
- Exclude if the study does not include clinical data on Product X efficacy or safety – for example, pre-clinical, pharmacokinetic, mechanistic, or animal/in vitro research, or studies of other compounds unrelated to clinical outcomes of Product X.

Objective and clear output options

Inclusion criteria

Exclusion criteria

Provide a 1-sentence rationale

Add reasoning



A more complex case



TRY IT YOURSELF

TURN THIS ALGORITHM INTO A COMPLEX PROMPT (USE GENAI TO HELP YOU)

HOW WOULD YOU ENSURE THE PROMPT SCALES (IE. REVIEW MULTIPLE PAPERS AT ONCE)?

Prompting to categorise

EXAMPLE OF
CAMINO PROMPT

Based on the provided title and abstract categorise this article using one of the following tags:

Meta-analysis
Review
Guidelines/consensus
Phase 3
Phase 2
Phase 1

Clinical study – unknown
Real-world study
Case reports
Pre-clinical/Research article
Study undefined.

Objective

Clear output options

Consider the following while categorising:

- If the article mentions an earlier phase trial but is actually a later phase trial, categorise it based on the later phase.
- If the trial is defined by a number followed by a letter, ignore the letter.
- If the article is a review but mentions clinical trials, categorise it as a Review.
- Use the “Clinical study – unknown” tag where you think it is a clinical study but the exact type is undefined. Use this tag for a randomised controlled trial where the phase is undefined, for an undefined multicentre study, or another study that reports patient outcomes such as efficacy, safety or long-term results.
- Use the “study undefined” tag where you are not sure what type of study this is. It could be because the abstract is missing and there isn’t enough information in the title alone to make an informed decision.

Rules

Grey area handling

Provide only the exact appropriate tag name described above, without changing the tag or any additional explanation.

Output constraint

THIS TYPE OF PROMPT:

- ✓ DEFINES A SINGLE CLEAR TASK
- ✓ RESTRICTS POSSIBLE OUTPUTS
- ✓ HANDLES AMBIGUITY
- ✓ PRODUCES CONSISTENT, REUSABLE RESULTS

Prompting to extract

EXAMPLE OF CAMINO PROMPT

You are a medical researcher tasked with summarising academic articles.

Role/context

Based on the title and abstract, create a concise summary of this article focusing on the key findings and conclusions.

Objective and source

Please include

- The main objective or research question
- The methodology used (if mentioned)
- The key results
- The main conclusions

Output requirements

Keep your summary concise, informative, and objective. Use a professional tone suitable for a medical audience.

Output constraint and format

THIS TYPE OF PROMPT:

- ✓ LIMITS WHAT SOURCE TEXT TO USE
- ✓ SPECIFIES EXACTLY WHAT INFORMATION TO PULL OUT
- ✓ REDUCES INVENTED CONTENT
- ✓ PRODUCES SUMMARIES IN A CONSISTENT STYLE

Workshop scenarios

SCENARIO 1: CATEGORISE



You are reviewing article titles and abstracts from a literature search.

Create a prompt that categorises each article based on the type of content.

Produce consistent, single-label outputs that clearly describe the primary focus of each article.

Tip: Use tags such as:

- Treatment
- Diagnostic
- Guidelines
- Epidemiology
- Safety
- Pre-clinical
- Health economics

SCENARIO 2: PRIORITISE



You are screening articles from a literature search on 'drug x' in 'therapy area y'.

Create a prompt that classifies each study as Include or Exclude based on relevance to your research question.

Tip:

- Focus on clinical relevance (eg. efficacy, safety, outcomes related to 'drug x')
- Exclude irrelevant content (eg. pre-clinical, mechanistic, unrelated topics or treatments)

SCENARIO 3: EXTRACT



You are reviewing included articles from your literature review.

Create a prompt that generates a structured summary of each study based on the title and abstract.

Produce summaries that are, concise, consistent and easy to compare across studies.

Tip: Include

- Study objective
- Methodology
- Key results
- Main conclusions

SCENARIO 4: COMPARE



You are reviewing multiple studies on the same drug.

Create a prompt that compares studies and identifies which provides the strongest clinical evidence.

Produce an output that:

- Compares studies consistently
- Highlights key similarities and differences
- Identifies the most robust evidence

Tip:

- Define what “strong evidence” means
- Avoid subjective or vague reasoning



ADVENTURES IN PHARMA