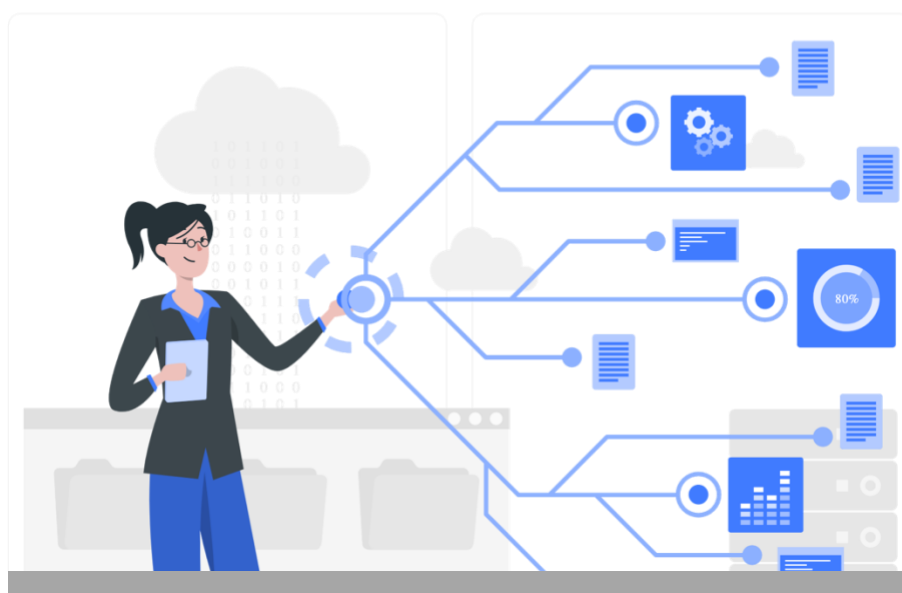


Pragmatic Approaches to Making LLMs Reliable and Cost-effective in Scientific Workflows

Technology Guide

Version 1.2



By:

Dr. Raminderpal Singh

Founder, HitchhikersAI

Global Head of AI and GenAI Practice, 20/15 Visioneers

raminderpal@20visioneers15.com

John Conway

Founder & Chief Visioneer Officer

20/15 Visioneers

Table of Contents

<i>Introduction</i>	5
<i>Section A: A Pragmatic Discussion on Using LLMs</i>	5
1. <i>About LLMs</i>	5
What can scientists do with LLMs today?.....	5
How do LLMs work?.....	5
How to reduce LLM hallucinations	6
How to use LLMs.....	6
How to keep data private	7
2. <i>Using LLMs to Enrich Animal Study Analysis</i>	7
How ChatGPT Enriched Animal Study Results	7
3. <i>Pragmatic Guidelines</i>	9
Pragmatic guidelines to getting the best out of LLMs.....	9
Pragmatic guidelines when using ChatGPT	9
Pragmatic guidelines when deciding what LLM system to use	9
<i>Section B: Suggested LLM Functions for Your Software Platform</i>	10
1. LLM Function: Organize, annotate, and explore legacy datasets to unlock new insights and make them usable in current research workflows.....	10
2. LLM Function: Protocol validation and error detection to ensure adherence to best practices and lab standards.....	10
3. LLM Function: Real-time protocol monitoring and suggestions for improvement.....	11
4. LLM Function: Tailored report customization for specific audiences (e.g., peer-reviewed journals, internal reports).	11
5. LLM Function: Automated hypothesis generation from literature and complex data analysis.	11
6. LLM Function: Real-time guidance and support for lab tasks, experimental designs, and data management.....	11
7. LLM Function: Automated literature search and summarization.	12
8. LLM Function: Literature gap identification.....	12
9. LLM Function: Streamline collaboration between team members by summarizing ongoing work, sharing key insights, and facilitating data handoffs.....	12
<i>Section C: Case Study - Revolutionizing Antibody Engineering with LLMs</i>	12
<i>Problem</i>	12
<i>Solution</i>	13

How Quine Biologics uses PythiaAI™ 14

Results 15

Conclusion 15

Introduction

Large Language Models (LLMs) offer much promise, with many of us already using LLM tools such as [ChatGPT](#) to help us through daily tasks. However, we have become cautious given the hype. And challenges such as operating costs, hallucination and data privacy are significant risks slowing down our implementation plans.

This Technology Guide provides pragmatic insights and learning on the opportunities and risks for using LLMs in scientific workflows - including actionable ideas and a real-world Case Study. Our goal is to give you the clarity, thinking, and confidence to take advantage of LLMs in advancing the capabilities of your software.

This document has three sections:

[Section A](#): A **Pragmatic Discussion** on the opportunities, risks and challenges of using LLMs in scientific workflows today. This discussion includes examples.

[Section B](#): A high level overview of the **LLM Functions** that would add value to data-driven scientific workflows.

[Section C](#): A real-life **Case Study**, showing the power of LLM functions - in this case, accelerating Antibody Discovery.

Section A: A Pragmatic Discussion on Using LLMs

1. About LLMs

What can scientists do with LLMs today?

Deep learning models have been around since the 1990s. [ChatGPT](#) opened the floodgates for LLMs to become mainstream. Because of a LLM's ability to process vast sets of data, there have been attempts to use them to accelerate insight generation and productivity gains across a range of industries. Drug discovery is no exception and many AI-led biotechs, including [Recursion](#) and [Insilico Medicine](#), are aggressively using LLMs in their workflows.

How do LLMs work?

[ChatGPT](#) says that an LLM is a type of artificial intelligence (AI) model designed to understand and generate human language. These models are based on deep learning architectures, particularly transformers, which enable them to process and generate text with a high degree of coherence and relevance. LLMs, like GPT-4, are trained on vast amounts of text data from the

internet, books, articles and other sources. This training allows them to learn the complexities of language, including grammar, context and even some level of reasoning and common sense. The latter is a large claim, so scientists are right to be wary. For those readers, who are interested in learning more, check out this [introduction](#).

It is important to note that:

- LLMs work by learning from user questions and iterations on the system. Essentially, they use and learn from your search terms, which is why some LLMs are cheap to access.
- LLMs are very expensive to build and train (and often to fine-tune).
- LLMs take advantage of known knowledge, and thereby are not great speculative tools (ie, when looking for insights in 'dark spaces').
- LLMs can be very useful for scientists. However, to make them effective, you must iterate the prompts and questions, which can be frustrating.

How to reduce LLM hallucinations

Hallucinations has made [news](#) many times. For scientists who have played around with LLMs, sometimes it is frustrating how unreal the answers can be. Hallucinations cannot be avoided when using non-curated knowledge. If scientific companies are able to, they should curate paper cohorts and upload them as the sole (or at least primary) source for analysis. Also, hallucination checks should be built into the workflow, using independent LLM calls.

How to use LLMs

Users of LLMs can be divided into two types:

1. Those who use what is available in a simple, and often private, way.
2. Those who build and train (or fine-tune) their own models.

This discussion focused on the former, mainly because the building of new models is currently too expensive for most of us.

The simple use of LLMs (point 1 above) can be done in two ways. The first is to go online and use an interactive system like ChatGPT or use software that includes wrappers around an LLM.

LLMs have their limitations and are not best-in-class for all modeling challenges. For example, AlphaFold 3 uses Diffusion Models (DMs) to get better results [[Devansh](#), [Sora](#)], as opposed to LLMs. Both DMs and LLMs come under the popular term Generative AI (GenAI). In addition,

there is the term Foundation Model, which is similar to, but more powerful than, application-specific [LLMs](#). As the GenAI field advances there will be more variants.

How to keep data private

Data privacy (confidentiality) is critical when using any external tools. When using LLM APIs, this can be achieved in several ways:

- Often, one can sign a BAA (Business Associate Agreement) with them to guarantee that they will not store search terms or any data. You can also request a 0 day retention policy (from the common default of 30 days).
- You have the option of running the LLM models locally (if you stick to the “open source” models, such as Llama and Mistral), using the [Ollama](#) tool. You can also download and run a wide range of HuggingFace open-source models using Ollama [instructions [here](#)], which opens up the opportunity for many new open-source models to be run locally.

2. Using LLMs to Enrich Animal Study Analysis

How ChatGPT Enriched Animal Study Results

The following is a simple case example to download and practice with ChatGPT, or other accessible LLM system. This shows the power that LLMs offer to improve scientists’ daily tasks, despite their caveats and challenges. You can download all the source files to use yourself ([see Simple ChatGPT Exercise](#)). Thank you to [Nina Truter](#) for her support in building this example.

About the example:

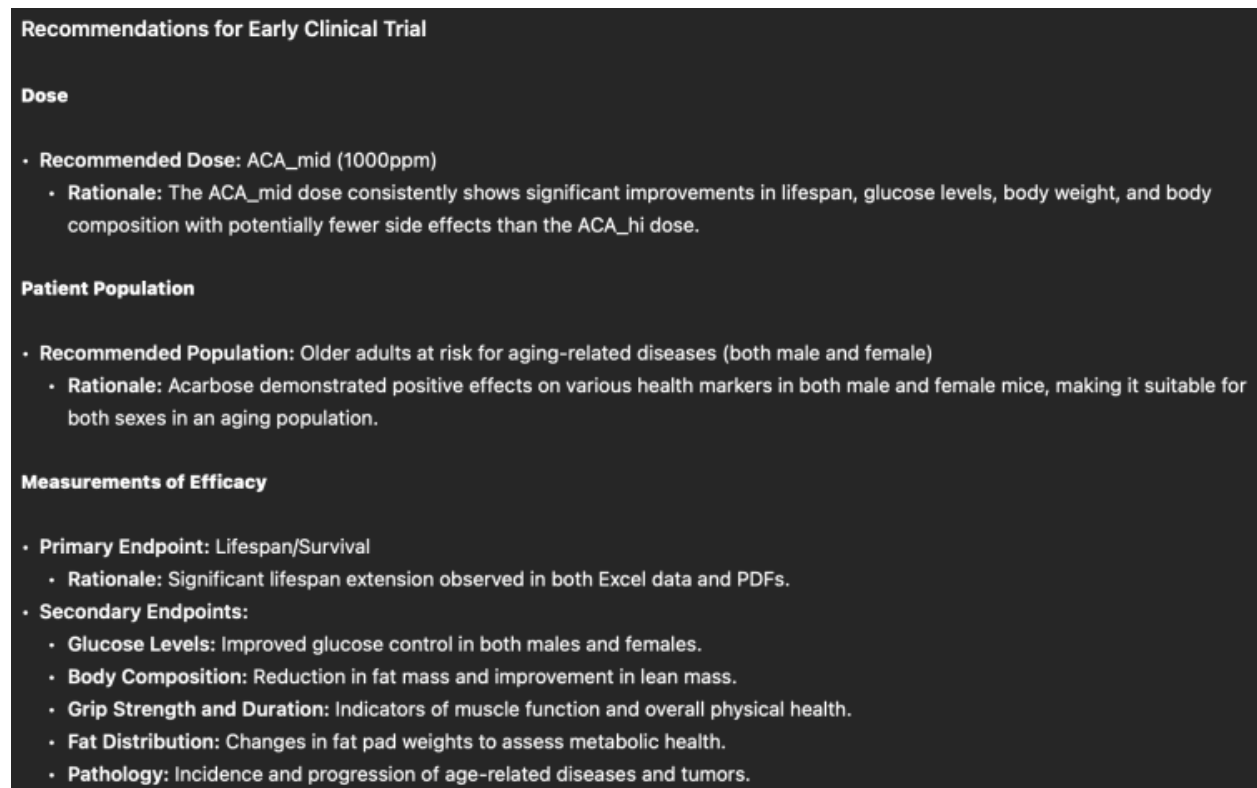
- Goal: Using extracted measurements from 10 papers on acarbose treated mice to improve the recommendations made from the results of the primary study.
- Key outputs required: Recommendation on dose, participants and measurements based on results from the primary study [[NCBI](#), [Phenome Jax](#)] and papers on acarbose treated mice, with supporting data points.
- Implementation challenges: Creating prompts to accurately extract information to support recommendations, accurately describing the content of multiple files and papers.

Importantly, you should be aware that commonly accessible LLM systems often share inputs you provide, so it is recommended that you not enter confidential information.

To help ChatGPT provide useful insights, there needs to be some 'prompt engineering'. This is a technical term for best-practices in the way prompts are written. As an example, the first prompt in this example is only to provide background and context to ChatGPT:

"You are a drug discovery scientist looking to make decisions on dose, participants and measurements when taking an existing diabetes drug into the aging-related diseases field. You have experimental results from a mouse study that show the effects of acarbose on lifespan, body weight, body composition, fat pads, glucose, grip strength, grip duration, rotarod and pathology. You also have several relevant scientific publications with studies investigating the effects of acarbose on different measurements in mice. You now want to interrogate your study results (which are in Excel files and images) and the publications separately for insights, and then together to get the best set of recommendations for your colleagues who are looking to perform early clinical trials with acarbose on aging-related diseases. To do this, you will now process a series of specific user-entered ChatGPT prompts."

The screenshot below shows the results from the last prompt. There are some nuances ChatGPT has not picked up on. For example, in female mice, the lifespan is not extended as much compared to male mice, but their physical measurements are improved. Improved prompts will aid the generation of more nuanced results.



Recommendations for Early Clinical Trial

Dose

- **Recommended Dose:** ACA_mid (1000ppm)
 - **Rationale:** The ACA_mid dose consistently shows significant improvements in lifespan, glucose levels, body weight, and body composition with potentially fewer side effects than the ACA_hi dose.

Patient Population

- **Recommended Population:** Older adults at risk for aging-related diseases (both male and female)
 - **Rationale:** Acarbose demonstrated positive effects on various health markers in both male and female mice, making it suitable for both sexes in an aging population.

Measurements of Efficacy

- **Primary Endpoint:** Lifespan/Survival
 - **Rationale:** Significant lifespan extension observed in both Excel data and PDFs.
- **Secondary Endpoints:**
 - **Glucose Levels:** Improved glucose control in both males and females.
 - **Body Composition:** Reduction in fat mass and improvement in lean mass.
 - **Grip Strength and Duration:** Indicators of muscle function and overall physical health.
 - **Fat Distribution:** Changes in fat pad weights to assess metabolic health.
 - **Pathology:** Incidence and progression of age-related diseases and tumors.

Figure 1, Results presented by ChatGPT (screenshot)

3. Pragmatic Guidelines

Pragmatic guidelines to getting the best out of LLMs

The guidelines are separated into two sections: those specific to using ChatGPT, and those to use when deciding on which LLM system to use. The former section allows the reader to begin now, with the recommendation not to share any confidential information on the tool. The latter section allows the reader to think ahead about an affordable, private and trustworthy system for the next couple of years.

Pragmatic guidelines when using ChatGPT

Note that many of these recommendations reflect the need to upload your own .pdf and .csv files:

- Where measurement data is used in your uploaded documents, label them as variables in your queries.
- In your queries, do not embed statements within statements – bring everything out in simple single purpose statements.
- Where ChatGPT stops because of long answers, break up the prompts and ask it to export to .csv or .pdf instead of printing on the screen. If it still stops part way through, type “Please continue from where you left off and finish the answer.”
- When the results are inaccurate and you need to re-run the query, try switching to a new ChatGPT window/chat. Another trick is to clear short-term memory within a prompt workflow, using this command “From now on, assume [new context] without considering our previous conversation.” For example, “From now on, assume we are talking about human-only clinical trials without considering our previous conversation.”

Pragmatic guidelines when deciding what LLM system to use

- LLM systems are made up of LLM models (for example, OpenAI and Claude) and software frameworks (for example, Autogen Studio and Open Web UI). Both these technology types are evolving rapidly, with multiple offerings. Any system you select needs to be adaptable to technology changes, for example if an effective drug discovery LLM model is offered as open-source next year. Note: There is a risk in using any “middleware” software to do the LLM work for you (especially when using [Agentic Workflows](#)), it is often recommended to keep your code as vanilla as possible.

- In addition to a rapidly evolving technology landscape, it is important to understand the changing balance between commercial and open-source offerings. As with other industries, these words of wisdom stand (relatively) true: “Wait a year, and someone will offer it for free (or at least it will be dirt cheap)!” It is important to be agile and respond to changes in this balance.
- Your LLM system may need to support both research-intensive activities, such as extracting new insights from large amounts of (.pdf) research publications, and workflow-driven tasks, such as designing a lab experiment. These are different problems that need to be solved with different approaches. A useful tip here is to design your workflow pretending that a human will perform the proposed LLM functions, and then swap the LLM technology in for the human once the design is ready.

Section B: Suggested LLM Functions for Your Software Platform

Below is a list of useful functions that LLM-focused workflows can perform today. They augment traditional data-driven software workflows.

1. LLM Function: Organize, annotate, and explore legacy datasets to unlock new insights and make them usable in current research workflows.

- How It Works: The LLM can parse and restructure legacy data (e.g., older formats, fragmented datasets) into a modern, usable structure. It automatically annotates the data with relevant metadata (e.g., experimental conditions, sample information) and highlights inconsistencies or missing data. The LLM can also identify connections between datasets, making it easier to reuse old data for current experiments.
- Value: Facilitates the integration of legacy data into modern research, reduces time spent on data preparation, and enhances the discovery of insights from historical datasets.

2. LLM Function: Protocol validation and error detection to ensure adherence to best practices and lab standards.

- How It Works: The LLM reviews experimental protocols for accuracy, completeness, and consistency. It compares the protocol to relevant standard operating procedures (SOPs) or industry standards, ensuring no critical steps are missed. It can also flag potential sources of error or inefficiencies in the experimental design.
- Value: Increases protocol reproducibility, reduces errors, and enhances overall experimental quality.

3. LLM Function: Real-time protocol monitoring and suggestions for improvement.

- How It Works: As experiments are conducted, the LLM continuously monitors the steps being taken, cross-referencing them with the written protocol. It flags deviations from the original design and suggests course corrections or improvements based on best practices and previous successful experiments.
- Value: Reduces deviations, ensures consistency in lab procedures, and minimizes human error, thus increasing the accuracy and reliability of experimental outcomes.

4. LLM Function: Tailored report customization for specific audiences (e.g., peer-reviewed journals, internal reports).

- How It Works: Researchers can specify the target audience (e.g., academic journal, regulatory body, internal stakeholders), and the LLM adjusts the report's complexity and focus. It tailors the language, highlights key insights relevant to the audience, and formats the document according to submission guidelines.
- Value: Speeds up the process of adapting reports for different purposes, enhancing communication and collaboration across departments or with external partners.

5. LLM Function: Automated hypothesis generation from literature and complex data analysis.

- How It Works: By analyzing literature (such as bioRxiv) and datasets from past experiments or public databases, the LLM identifies correlations, trends, or anomalies. Based on these patterns, it generates hypotheses for new experiments or research directions. It can also provide reasoning behind each hypothesis by linking back to the data or relevant literature.
- Value: Drives innovation by suggesting unexplored areas of research and validating these suggestions based on existing data patterns.

6. LLM Function: Real-time guidance and support for lab tasks, experimental designs, and data management.

- How It Works: The LLM acts as a digital assistant that provides immediate support for lab activities. It can answer specific lab-related questions (e.g., "What is the next step in this protocol?" or "What buffer should I use for this experiment?"). The assistant can also retrieve and interpret previous experimental data, suggest modifications to protocols, and track experimental progress in real time.
- Value: Enhances lab efficiency by providing researchers with on-demand support, reducing time spent searching for information or troubleshooting.

7. LLM Function: Automated literature search and summarization.

- How It Works: The LLM can search scientific databases (e.g., PubMed, bioRxiv) based on specific research queries and return summaries of relevant papers. It highlights key findings, experimental methods, and conclusions, condensing large volumes of literature into a digestible format.
- Value: Saves researchers significant time when conducting literature reviews and helps them quickly identify the most relevant papers for their work.

8. LLM Function: Literature gap identification.

- How It Works: After reviewing a body of literature, the LLM can identify areas that have been under-explored or where conflicting results exist. It generates suggestions for future research based on these gaps, helping scientists focus on unexplored or high-impact areas.
- Value: Increases research efficiency by focusing efforts on understudied areas, potentially leading to novel discoveries.

9. LLM Function: Streamline collaboration between team members by summarizing ongoing work, sharing key insights, and facilitating data handoffs.

- How It Works: The LLM integrates with shared digital workspaces and collaboration tools (e.g., Microsoft Teams, Slack, Asana) to provide regular updates on project progress. It can automatically generate summaries of work done by different team members, flagging important milestones or issues, and suggesting tasks that need attention. The LLM can also translate complex scientific data into more accessible formats for non-technical stakeholders.
- Value: Enhances teamwork and communication by ensuring that all collaborators have access to up-to-date information, reducing duplication of effort, and improving transparency across multi-disciplinary teams.

Section C: Case Study - Revolutionizing Antibody Engineering with LLMs

Problem

Antibody engineering, particularly the design of antibodies that target diseases with therapeutic potential, involves exploring a vast array of unstructured and structured data from scientific articles, internal research, and clinical studies.

Researchers face two main challenges:

- **Scientific Complexity:** Designing effective antibodies requires deep knowledge of protein targets, such as epidermal growth factor receptors (EGFRs), and identifying novel mechanisms of action that might lead to therapeutic breakthroughs.
- **Business Considerations:** Scientists also need to make strategic decisions regarding (for example) which diseases to target—whether to pursue well-known diseases that larger pharmaceutical companies focus on or to explore rare diseases and overlooked protein targets.

Traditional research methods, which involve manually analyzing literature and data, are time-consuming and may limit the breadth of inquiry. There's a growing need for tools that can assist in narrowing the research focus, identifying actionable insights, and accelerating the development of antibody therapies.

Solution

To address these challenges, a new AI + LLM platform ([PythiaAI™](#)) was developed by 20/15 Visioneers, which offers researchers the ability to quickly narrow down large search spaces by asking hypothesis-based questions. The platform's ability to process unstructured data from diverse sources allows scientists to focus their expertise on the most promising aspects of antibody design, while the AI aids in identifying novel disease targets and therapeutic possibilities.

PythiaAI™ is a reliable and low-cost agentic system that operates as follows:

1. **Hypothesis-Driven Search:** Scientists can input domain-specific knowledge and ask targeted questions, such as identifying novel methods for EGFR antibody design. The LLM refines the search space by curating relevant data from sources like PubMed, bioRxiv, as well as proprietary paper archives.
2. **Dataset and database extraction:** By connecting, extracting, and curating relevant datasets from external and internal sources, the system uses traditional data science and machine learning approaches to extract key insights. For example, the system ranks targets based on structured data available.
3. **Business Logic Integration:** By combining scientific queries with business considerations, such as determining which diseases to target based on market size or strategic opportunity, the LLM helps researchers make informed decisions. For example, should the scientist focus on rare diseases with high unmet needs, or pursue broader targets that are already under investigation by larger pharmaceutical companies?

4. Curation of Unstructured Data: LLMs convert unstructured data into actionable insights. For instance, in the context of EGFR antibody design, the system provides an output **epitope mapping** — identifying novel protein regions that haven't been explored or offering new perspectives on existing ones.
5. Augmenting Scientific Expertise: The LLM acts as a virtual Research Assistant, assisting scientists in areas where they may not have deep expertise. For example, if the researcher is focused on antibody engineering but lacks the medical knowledge to identify the most promising disease targets, the platform can fill that gap by sifting through clinical data and providing suggestions on lesser-known protein targets with therapeutic potential.
6. Predicting Future Science: By extracting, aggregating, curating, and analyzing large amounts of unstructured and structured information, the system can next apply pattern matching and cross-correlation LLM searches to see into the future. For example, what targets are yet to be studied in industry but are worth taking an early look at now.

How Quine Biologics uses PythiaAI™

Here's a specific example from Brett Spurrier, Founder of Quine Biologics, on how Quine recently used [PythiaAI™](#):

- We focus on generating novel antibody sequences designed to bind specific epitopes on protein targets. We take a structure-based design approach, meaning we analyze the structure of the antigen for things like potential binding properties as well as function. We engineer antibody CDRs to optimally bind with the highest affinity specifically to those epitopes of interest.
- One of the open questions that we need to answer in this process, however, is: where is the optimal location for the function that we're trying to achieve? We fed PythiaAI™ this request, asking for recent methods in EGFR antibody design as an example, along with a pre-curated knowledge base containing over 10 years of antibody engineering information pertaining to research groups and accompanying data.
- The system then alerted us to a recent 2023 paper describing a process for epitope mapping that was ultimately very useful for us in identifying new and unreported epitopes on EGFR. These epitopes, if blocked by an antibody antagonist, would potentially inhibit aberrant signaling pathways in the cell. There are no known antibodies targeting this particular region, so it ultimately became clear that this was the epitope we wanted to continue to investigate and develop new antibodies towards. That became the basis for our next project.

For more info, see this [webinar discussion](#) with Brett.

Results

The results for the investigation are shown in Figure 2. They show how LLMs can be used to identify epitope mapping finds based on an EGFR query.

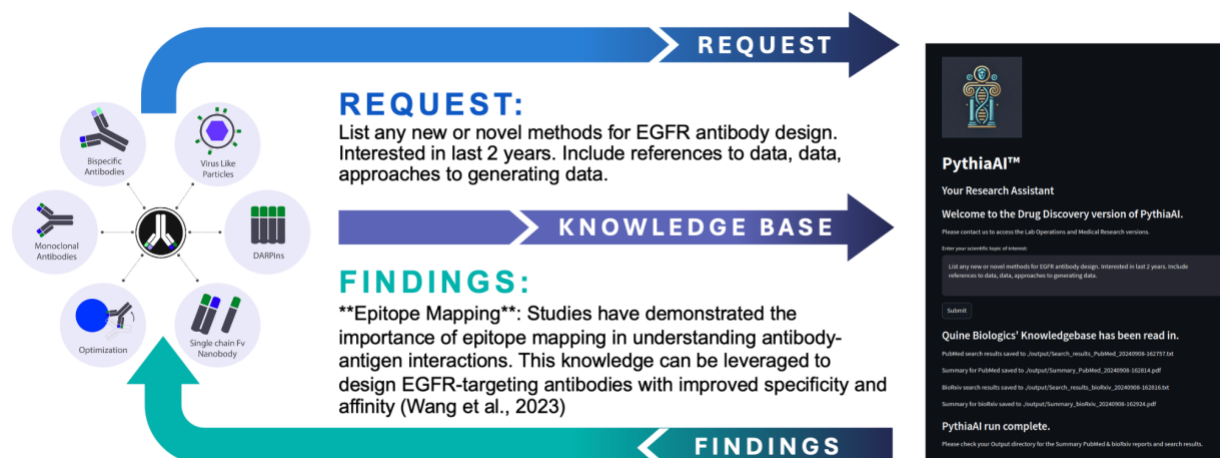


Figure 2: Description of high-level workflow using PythiaAI™ (right) with the Quine computational system (left), including the scientific question (Request) and output (Findings)

Conclusion

This Case Study demonstrates the transformative potential of LLMs in augmenting scientific workflows, particularly in antibody engineering. By narrowing down the search space, curating unstructured data, and aligning scientific discovery with strategic business goals, LLMs enable researchers to accelerate the development of novel therapeutics. The ability to ask hypothesis-based questions and receive highly curated, actionable insights allows scientists to focus on innovation while the system handles the heavy lifting of data analysis and target discovery.

