

Taking the Plunge:

Getting Large Language Models to Reliably and Cost-efficiently Work for you

Dr Raminderpal Singh

raminderpal@hitchhikersai.org

<https://www.hitchhikersai.org/>

A non-profit impact community, accelerating the adoption of AI/ML and data in drug discovery & development.

**DON'T
PANIC!**
THE ANSWER IS 42



DRUG DISCOVERY WORLD
DDWTM
turning science into business



LLMs and Agents

Examples

Getting started

LLMs and Agents

LLM in Science

Text Processing

Literature
Summarization

Key Concept
Extraction

Paper Synthesis

Scientific Analysis

Pattern
Recognition

Hypothesis
Formulation

Method Critique

Content Creation

Technical
Writing

Experiment
Protocols

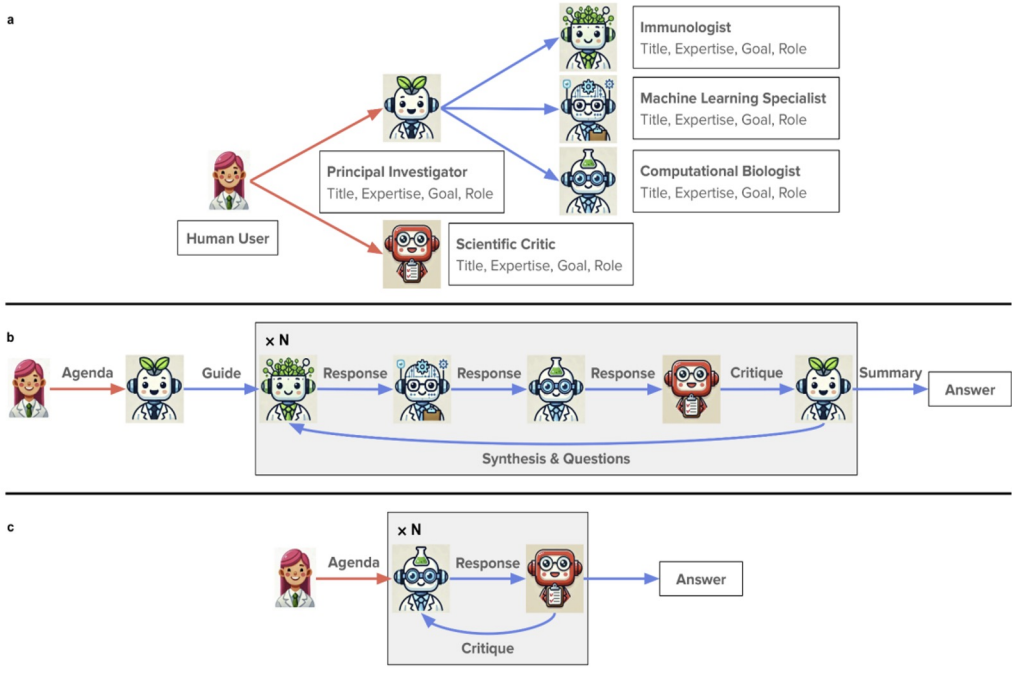
Documentation

Code Support

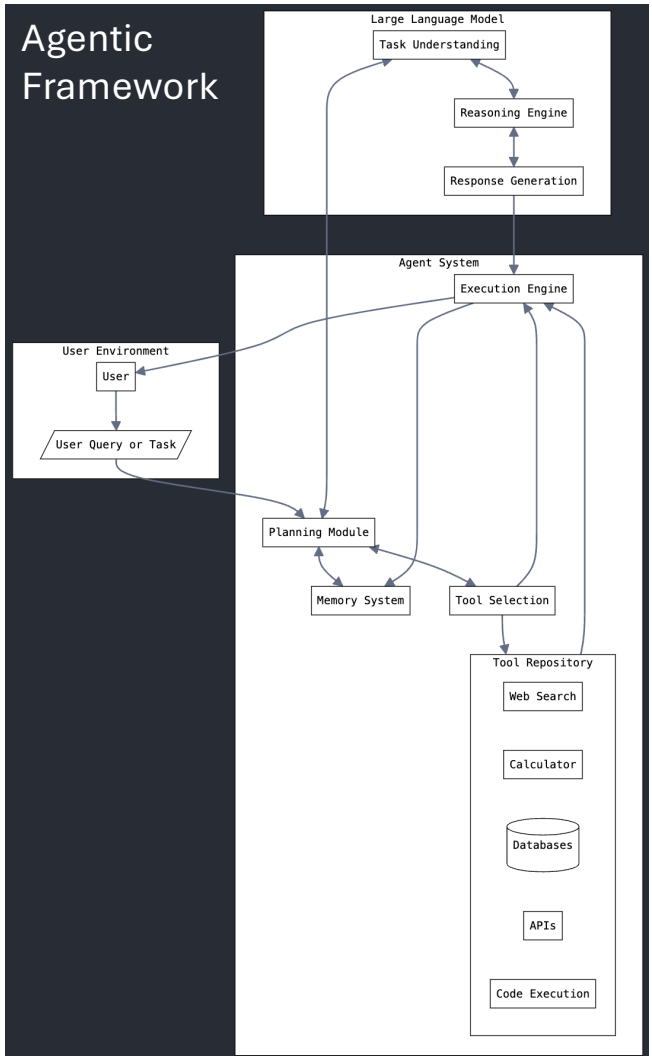
Data Analysis
Scripts

Visualization
Code

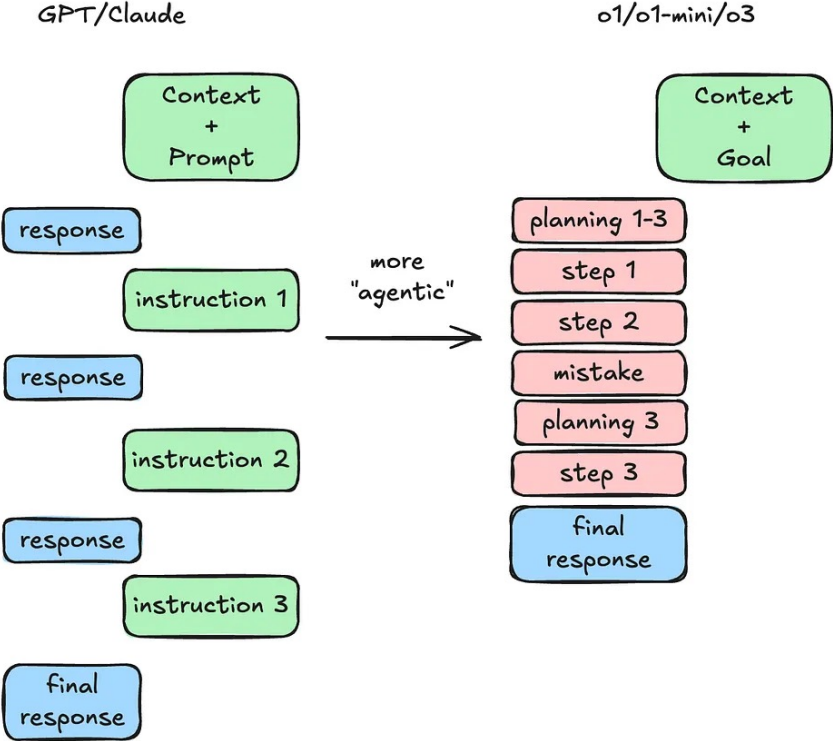
LLM Agents are cool



The Virtual Lab: AI Agents Design New SARS-CoV-2 Nanobodies with Experimental Validation
<https://www.biorxiv.org/content/10.1101/2024.11.11.623004v1>

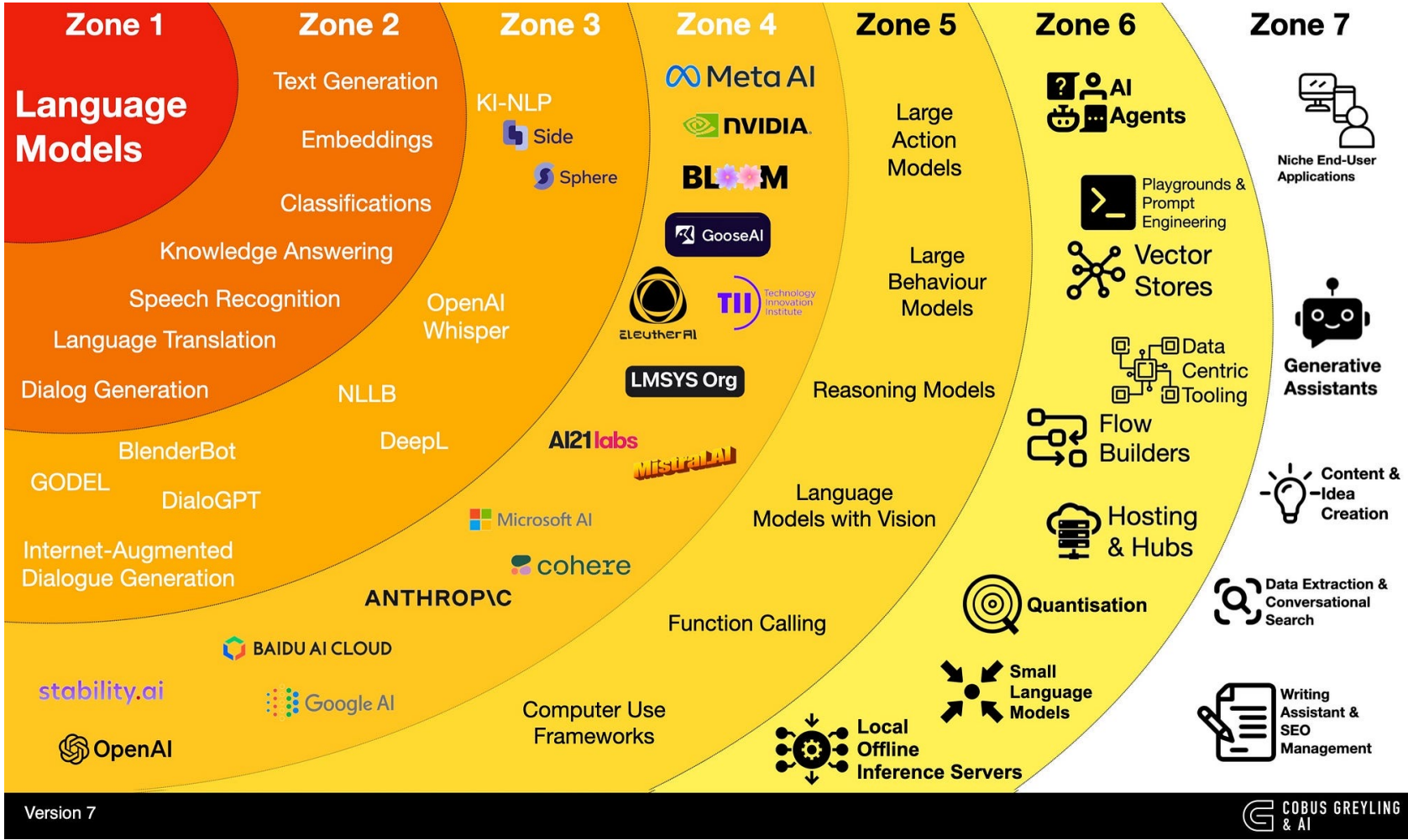


LLM themselves are becoming agentic



<https://www.latent.space/p/o1-skill-issue>

Lots going on here ...



COMPARISON: Core Characteristics and Relationships

Aspect	Artificial Intelligence (AI)	Natural Language Processing (NLP)	Large Language Models (LLMs)
Definition	The broader field of creating systems that can simulate intelligent behavior	A subfield of AI focused on enabling computers to understand and process human language	Neural network architectures specifically designed to process and generate human language at scale
Scope	Encompasses all forms of machine intelligence, including vision, robotics, expert systems, and decision-making	Focuses specifically on computational linguistics and language-related tasks	Specializes in language understanding and generation using transformer-based architectures
Historical Development	Emerged in the 1950s with early computing; evolved through symbolic AI, expert systems, and machine learning	Developed in the 1960s, initially rule-based, then statistical, now neural-based	Emerged in 2017 with the transformer architecture; rapid evolution through BERT, GPT series, etc.
Key Technologies	Includes machine learning, neural networks, expert systems, genetic algorithms, and robotics	Includes tokenization, parsing, word embeddings, and various ML algorithms	Primarily based on transformer architecture with self-attention mechanisms
Primary Applications	Autonomous systems, game playing, pattern recognition, problem solving, and general intelligence tasks	Language translation, sentiment analysis, text classification, information extraction	Text generation, conversation, code writing, creative tasks, and complex reasoning
Data Requirements	Varies by application; can use structured or unstructured data	Typically requires annotated linguistic data and text corpora	Requires massive amounts of text data for pre-training
Processing Approach	Can be rule-based, statistical, or neural, depending on the application	Combines linguistic rules with statistical and neural methods	Primarily neural, using attention mechanisms and deep learning
Strengths	Versatile problem-solving capabilities across domains	Specialized language understanding and task-specific performance	Powerful language generation and transfer learning abilities
Limitations	May struggle with common-sense reasoning and generalization	Can be brittle and domain-specific	Resource-intensive, potential for bias, lack of true understanding
Relationship to Others	Parent field encompassing both NLP and LLMs	Broader field that includes LLMs as one implementation approach	Specialized implementation that advances both AI and NLP goals

Examples

These are apps I have recently written. Please contact me for more details.

EXAMPLE 1: LeadgenAgent

Free app now available: **LeadgenAgent**

- ✓ Accelerating product & consulting sales in the life sciences industry
- ✓ Free for commercial and non-commercial use, for Win / Mac / Linux
- ✓ Implemented with a reliable LLM agentic framework
- ✓ Runs with \$0 operating costs

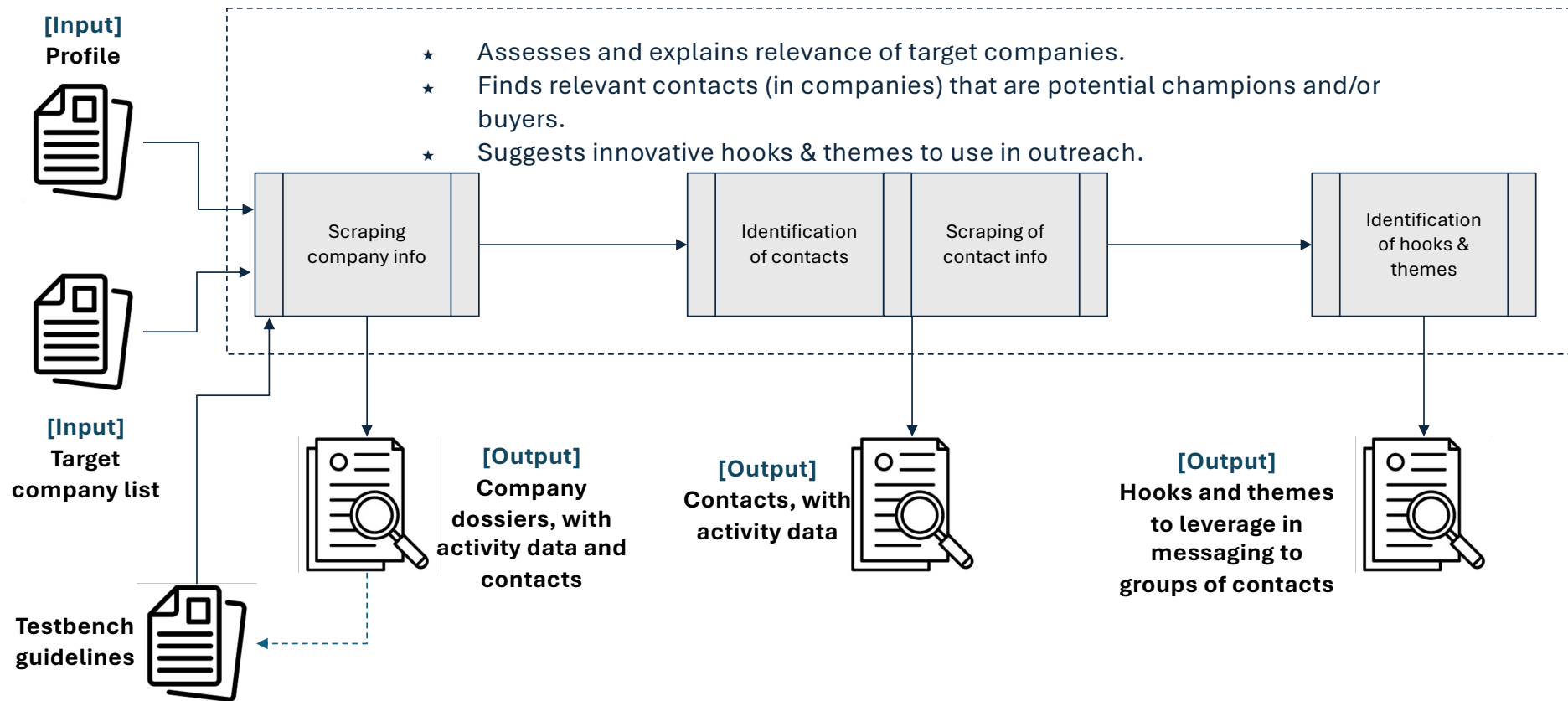
**DON'T
PANIC!**
THE ANSWER IS 42

To learn more: <https://www.hitchhikersai.org/leadgenagent>

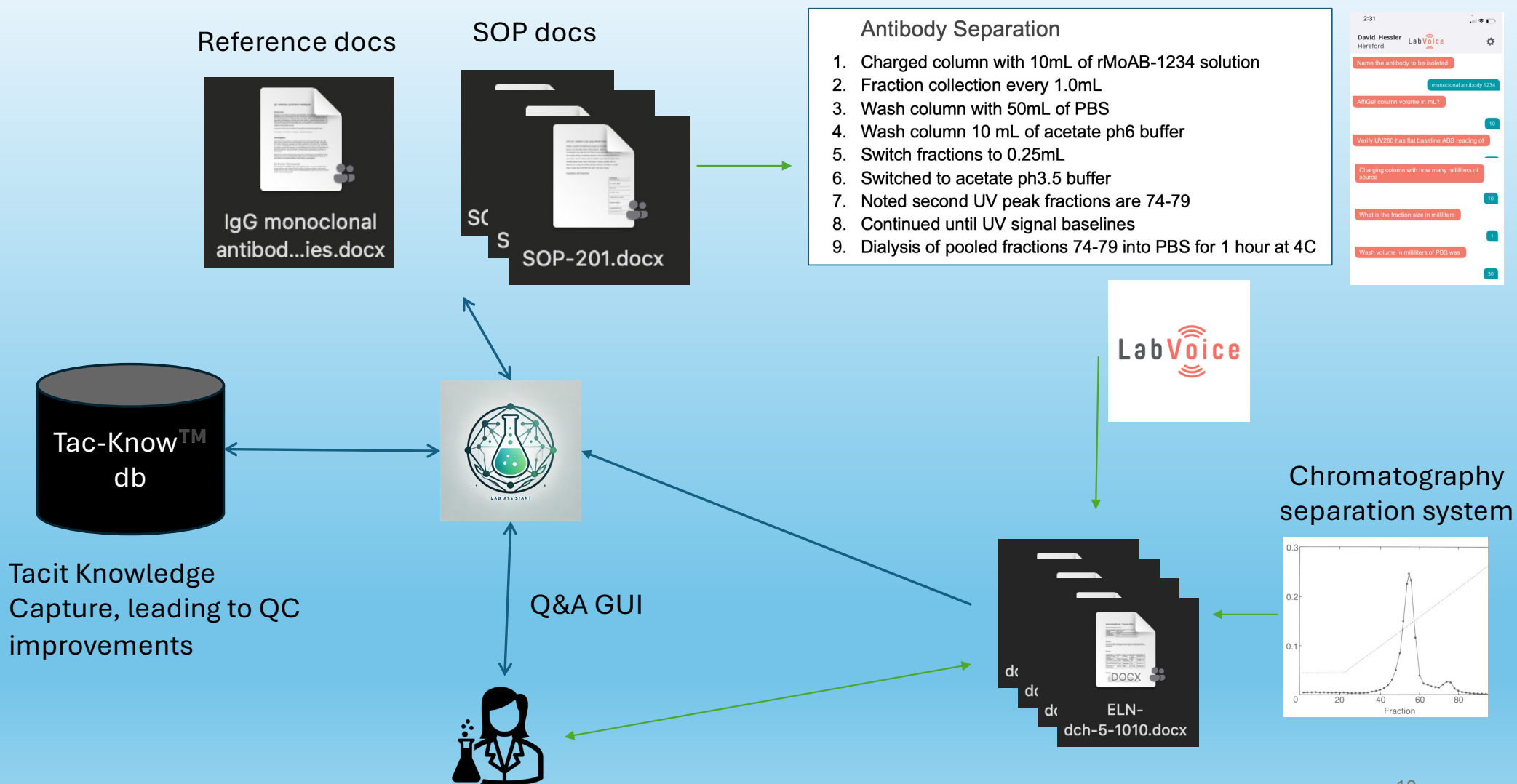
Finding and engaging clients & contacts is a continuous challenge, and the approach in LeadgenAgent is showing actionable results. Vendors like Ontologic need tools such as LeadgenAgent that provide accurate context that are relevant to our offerings.

- Dr Eli Pollock, co-Founder, Ontologic

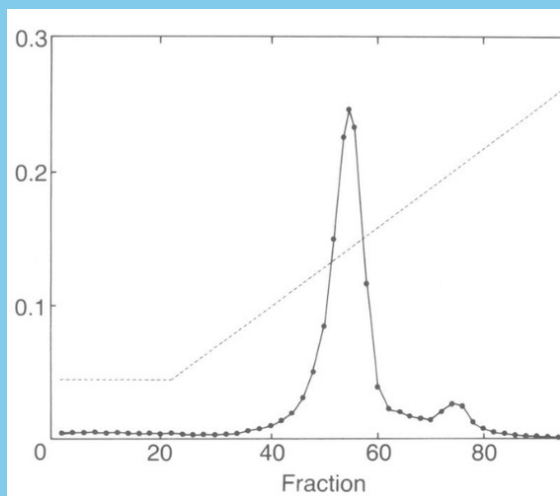
Which contacts would buy my offering? Who is the right entry point into an account? What “hooks” will work for a particular contact? What messaging themes will resonate with sub-groups of clients?



EXAMPLE 2: TAC-KNOW LAB ASSISTANT™



ELN



Conclusion

Separation of IgG from general lysate was achieved however there was insufficient separation between the waste peak and the IgG peak so the target is likely to be lower purity than desired. Will send fractions 55 (waste), and pooled 74-79 off for analytical purity analysis by PAGE.

Tac-Know db

=== Tacit Knowledge Entry ===

Date: 2025-01-07 13:37:10

Type: Lab Process Recommendation

Context: IgG Purification

Source: Expert System Analysis

Based on the laboratory conclusion, it appears that the separation of IgG from the general lysate was achieved, but there was insufficient separation between the waste peak and the IgG peak, resulting in a lower purity than desired.

To address this issue, I recommend re-evaluating the washing and elution steps in the Affinity Chromatography (LC Protein-A) process. According to the reference document, "Washing the column with PBS, and then extracting waste at pH 6, must be done with care. The UV280 detector will show when the column is truly clear of eluted proteins. Assuming that a given number of column volumes of wash will work is a mistake: use the effluent detector to indicate when a stage is complete." (Reference: IgG antibody purification strategies)

Specifically, I suggest re-washing the column with 2X column volumes of PBS or until UV280 detection baseline is achieved, as stated in SOP-201: Isolation of IgG using AffiGel Protein-A. Additionally, the pH 6 acetate buffer washing step should be repeated until the UV280 detector indicates a stable baseline, ensuring that the column is truly clear of eluted proteins.

Furthermore, the reference document highlights the importance of running the column slowly enough to allow binding time for the IgG. Therefore, I recommend verifying that the column flow rate is set at no faster than 10% column volume per minute, as specified in the reference document.

If the issue persists, consider implementing additional polishing steps, such as Ion Exchange Chromatography (LC using DEAE media) or Precipitation, as described in the reference document. These methods can provide additional purification and help achieve the desired purity levels.

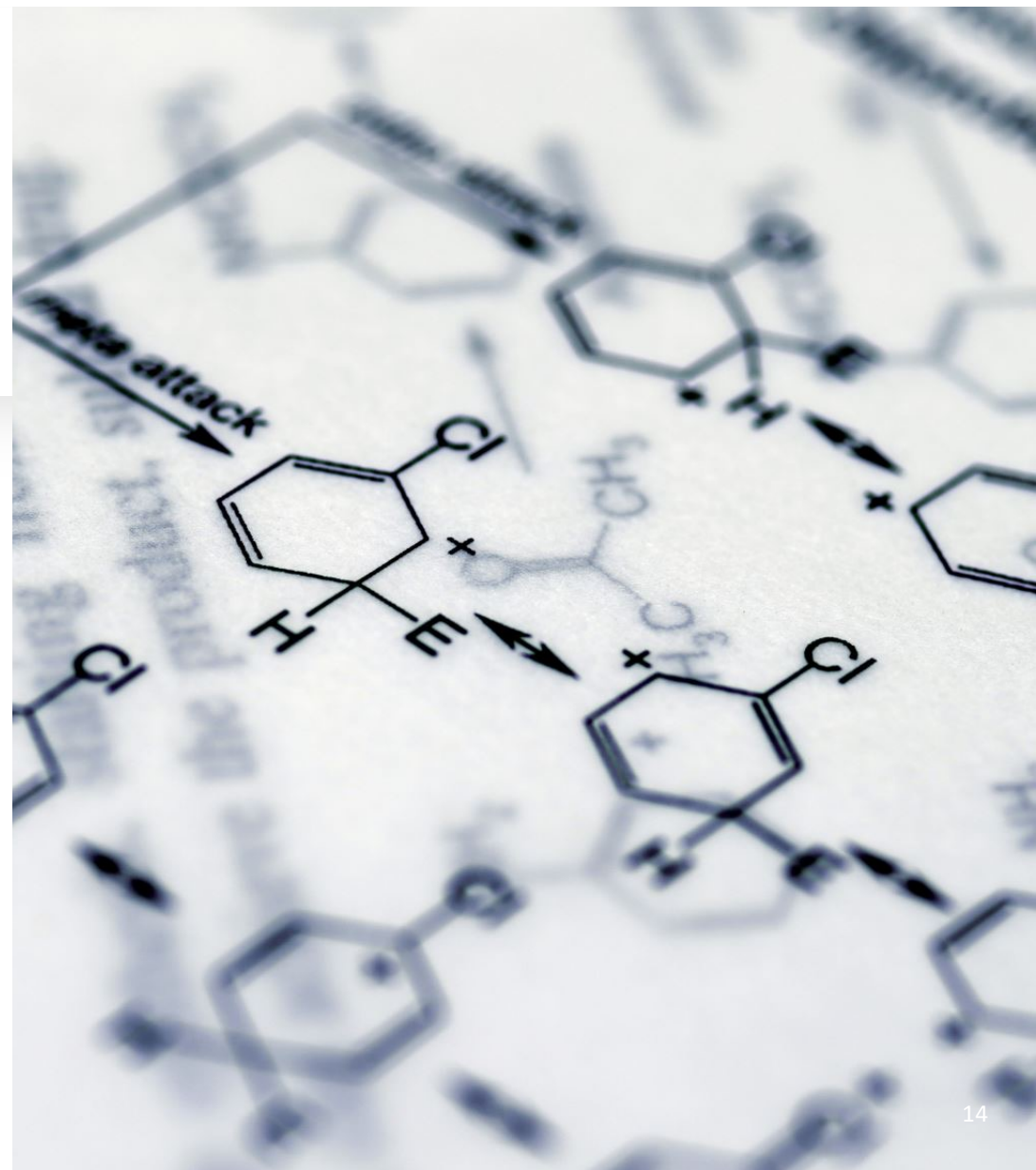
Lastly, it is essential to ensure that the analytical purity analysis by PAGE is conducted correctly to verify the purity of the IgG fractions. This will help determine the effectiveness of the revised purification strategy and identify any further adjustments that may be necessary.

EXAMPLE 3:

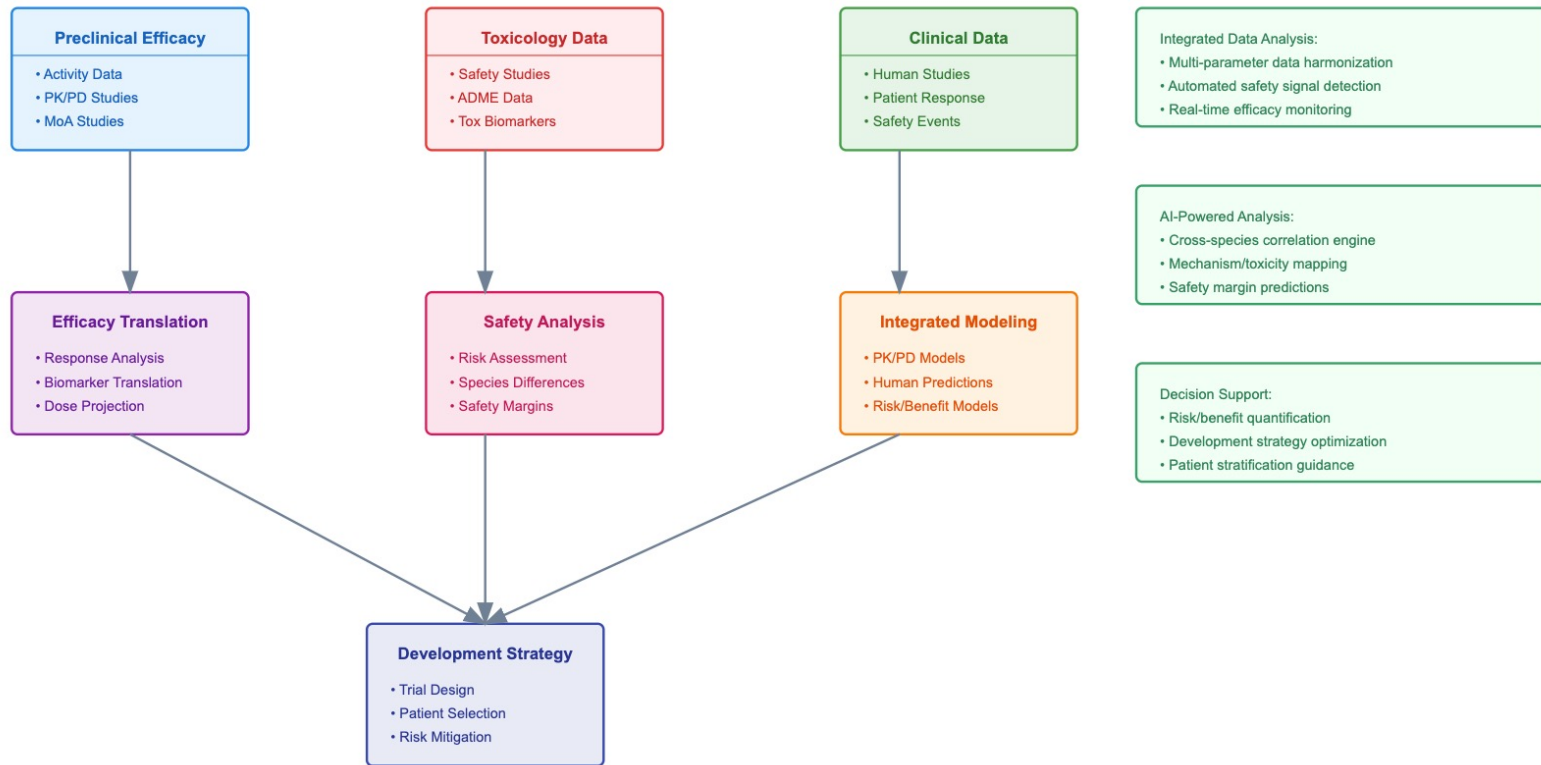
Accelerating data understanding & efficient decision making, during translation from late pre-clinical to clinical

- Scientists need a way to quickly analyze / understand findings in their data, connect these findings to existing scientific literature, and generate evidence-based hypotheses within hours rather than weeks.

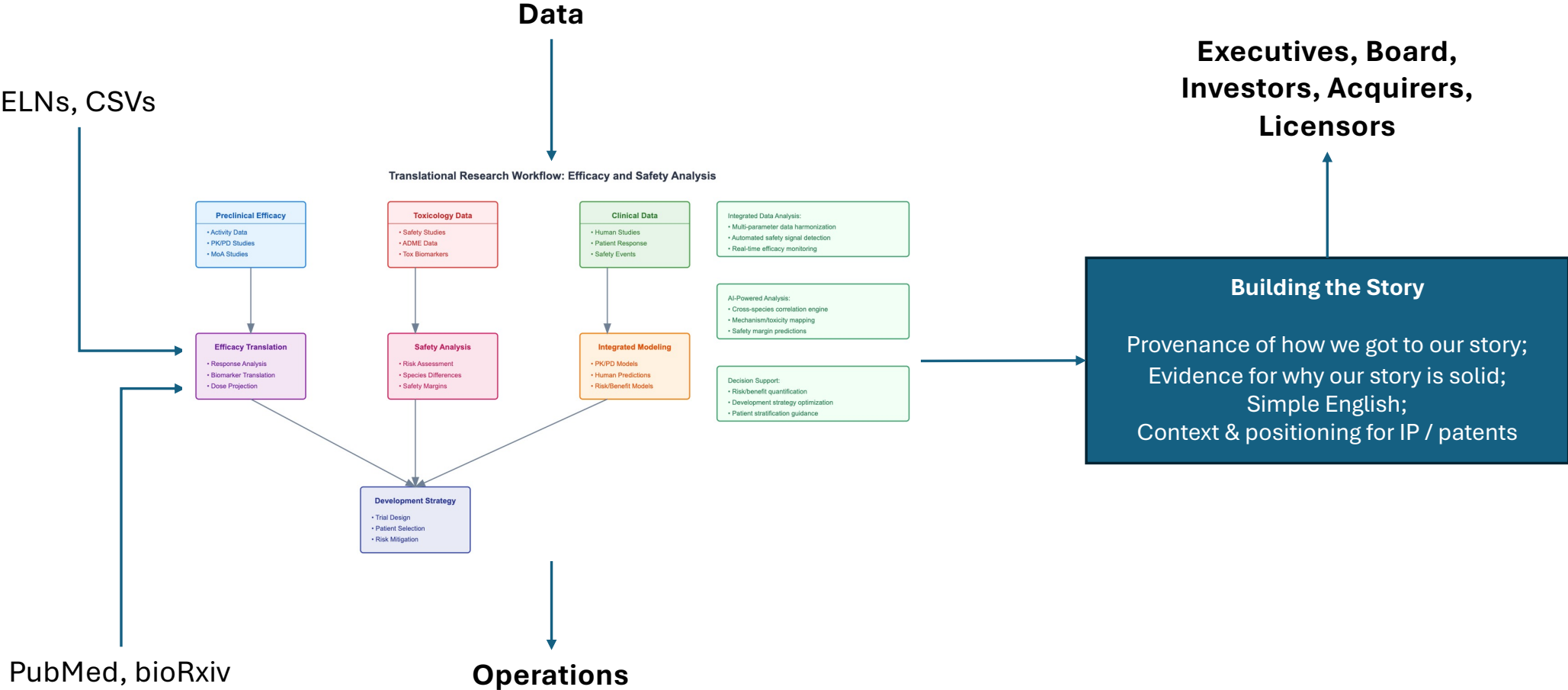
Action: If you are interested in participating as an early user, please let me know.



Translational Research Workflow: Efficacy and Safety Analysis

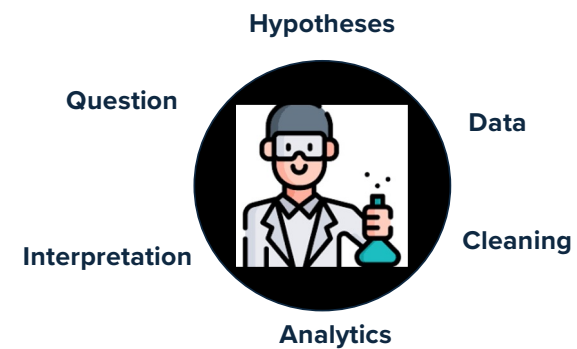
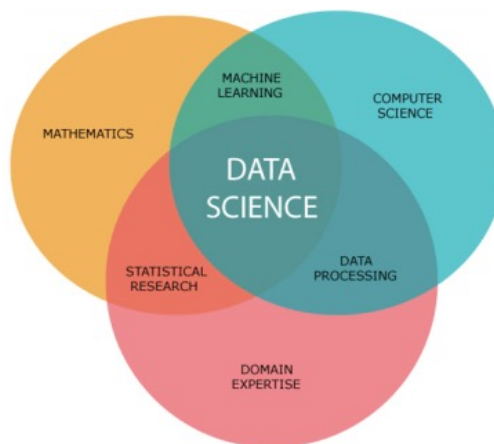


System design



Getting started

LLMs are force multiplier for how drug discovery is done



Pragmatic guidelines to building LLM systems

LLM systems are made up of LLM models (for example, OpenAI and Claude) and software frameworks (for example, Autogen Studio, HuggingFace smolagents, Pydantic AI, Langgraph). 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.

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!”** 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 system with a human performing the proposed LLM functions, and then swap the LLM technology in for the human once the design is ready.**

Pragmatic guidelines to building LLM systems (con't)

Model Options

PRINCIPLE: Use affordable models if you can. Life sciences info is heavily-typed, and your efforts should be on prompts and prompt chains to get the most out of the workflow.

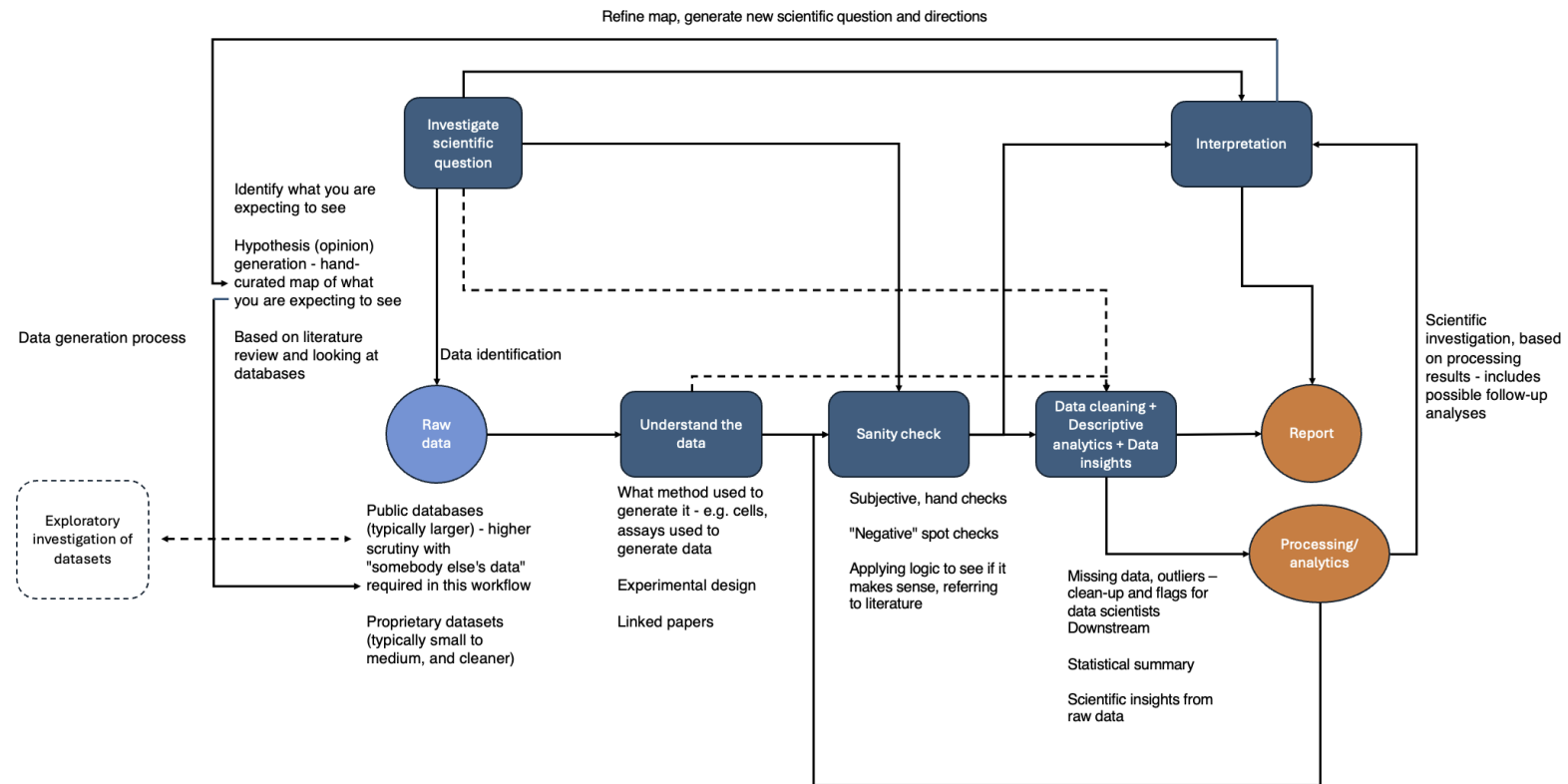
- Full function APIs: GPT4o / o1 and Claude 3.5 Sonnet are the go-to places, but can be expensive.
- Lesser function APIs: OpenAI and Anthropic offer older models at cheaper costs, but my preference is to go to open-source models through low-cost services such as GROQ CLOUD. For example, Llama3.3:70b is a great model.
- Local models: Ollama is great. It offers numerous native models as well as running GGUF files available on HuggingFace. There is a lot you can do with smaller models, e.g. Phi4 (14b) and Llama3.1 (8b).

Agentic Frameworks

PRINCIPLE: Be wary of black-box frameworks with 100ks of code in them. You need to know (roughly) what is going on, otherwise everything breaks when you try and scale.

- Cloud providers offer them and there are several open-source. For example, within Bedrock in AWS.
- My favorite is Pydantic-AI as it is open-source, light-weight, and a natural extension of Pydantic (which many of us already use in our coding). Also, it support numerous model providers.

LLMs are primed to the impact the Scientific Workflow



**PLEASE JOIN
OUR
COMMUNITY:**

hitchhikersai.org

Welcome to HitchhikersAI

A non-profit impact community, accelerating the adoption of AI/ML and data in drug discovery & development.

**DON'T
PANIC!**
THE ANSWER IS 42

HitchhikersAI aims to fix the disconnect between AI/ML and data and their practical application in early drug discovery by offering targeted non-profit consulting to biotech companies.

This involves helping scientists clearly define their research questions (killer questions) and designing customized plans that integrate educational resources, computational tools, and curated data to effectively use AI/ML technologies in their research.

The community is growing and currently consists of 300+ bench scientists, data scientists, mathematicians, business owners, executives, academics etc.

Drug TARGET REVIEW

ALSO, PLEASE
CHECK OUT MY
REGULAR
ARTICLES ON
AI IN DRUG
DISCOVERY:

<https://www.drugtargetreview.com/?s=raminderpal+singh>



ARTICLE

Part two: how ChatGPT enriched animal study results

15 July 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Part four: an industry leader's perspective on managing data quality

24 September 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Scientific workflow for hypothesis testing in drug discovery: Part 2 of 3

14 January 2025 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers, Nina Truter)



ARTICLE

Part two: the impact of poor data quality

26 August 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Bridging science and technology: a biotech CEO's perspective

14 October 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

An industry leader's perspective on the complexity of scientific data

24 October 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Kickstarting the use of AI for biotech: part three

20 June 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Part one: what can scientists do with LLMs today?

24 June 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Part three: pragmatic guidelines to getting the best out of LLMs

24 July 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Part one: an introduction to data quality

14 August 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Kickstarting the use of AI for biotech: part two

24 May 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)



ARTICLE

Kickstarting the use of AI for biotech: part one

15 May 2024 | By Dr Raminderpal Singh (Hitchhikers AI and 20/15 Visioneers)

THANK YOU

raminderpal@hitchhikersai.org