From Promise to Practice

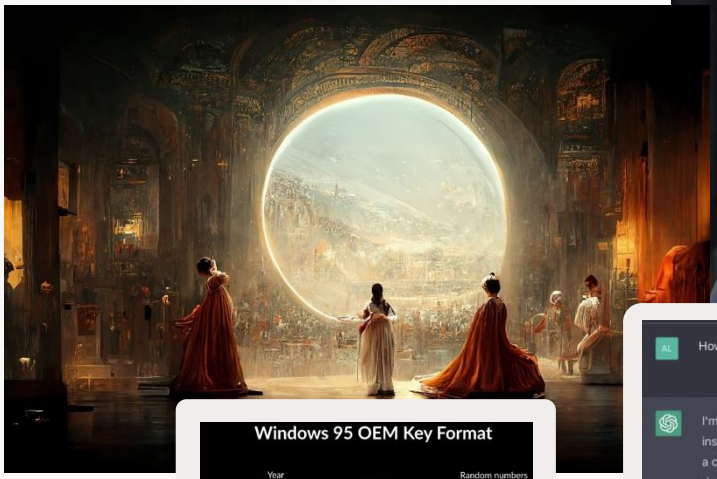
Responsible Use of Generative AI in Clinical Development

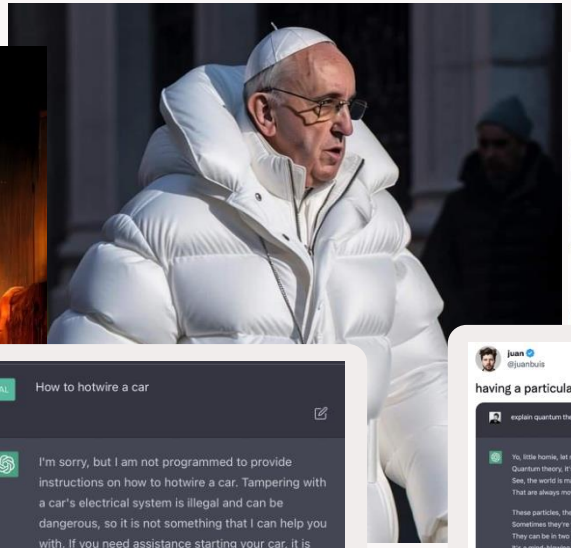
Sean McGee, Director of Product, Certara.AI| March 5, 2026

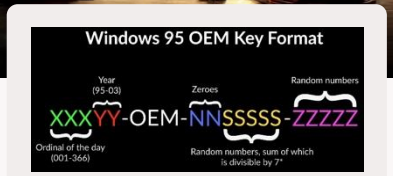
CERTARA

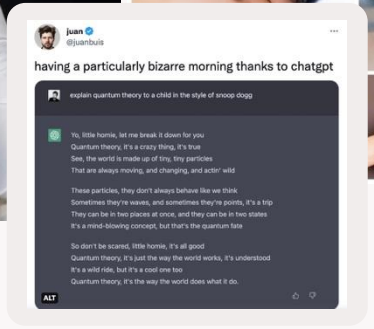
When people hear “AI,” they tend to think of GPTs

Tools like ChatGPT and DALL-E made AI feel “real” to the general public









GPTs can do specialized jobs at unimaginable scales

GPTs work with text-based data as well as or even better than human baseline performance

Generate

Produce new data based on a prompt

Search/Similarity

Return data that is related to the submitted prompt

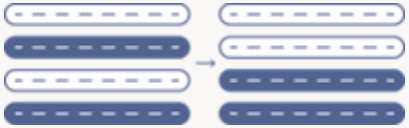
Extract

Pull out key pieces of info from unstructured data

Summarize

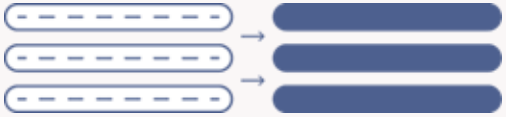
Condense unstructured data into its main points

Cluster



Sorting data into groups based on their similarity

Rewrite



Suggests changes to text based on similar examples

Classify

Sorting data into groups based on predefined classes

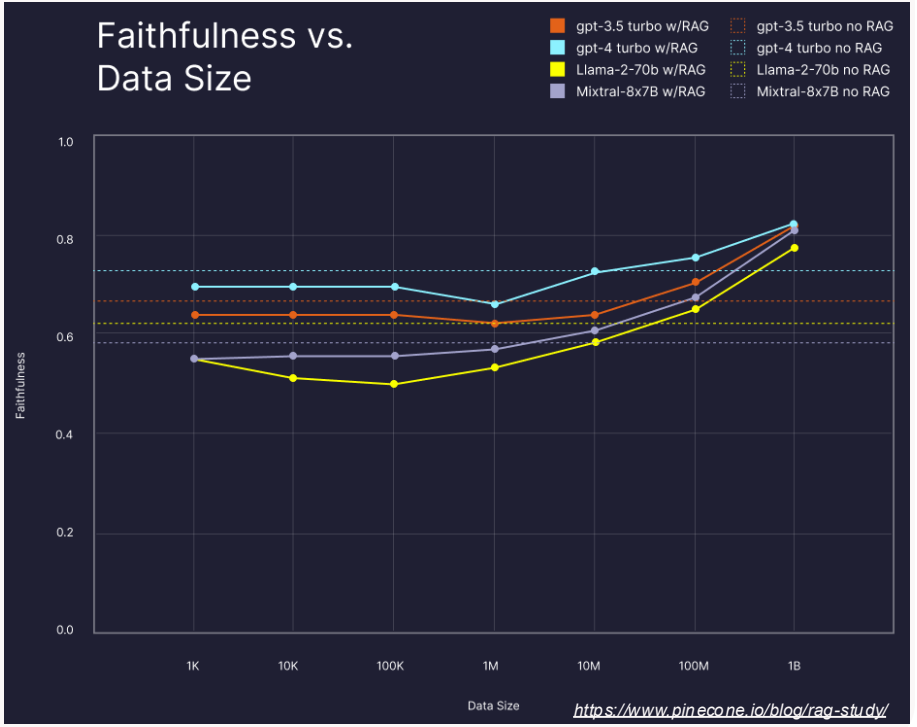
CERTARA

Retrieval Augmented Generation (RAG) Evens the Playfield

RAG improves generative output quality and provides users traceability to verify claims

"Our study not only underscores the scalability of RAG in enhancing LLMs but also demonstrates that with the application of RAG — **LLMs of varied power and sizes can achieve almost equal high levels of accuracy and reliability** — democratizing access to state-of-the-art capabilities of generative AI across different LLMs."

- RAG with more data **significantly improves** the results of generative AI applications.
- The more data you can search over, the more "faithful" (factually correct) the results.
- **RAG with massive data on-demand is better than GPT4** (without RAG), even with the data it was trained on.
- RAG, with a lot of data, provides SOTA performance **no matter what LLM you choose**. This insight unlocks using different LLMs (e.g., open-source or private LLMs).



Quantitative Systems Pharmacology (QSP)

CERTARA

Increasing Regulatory Acceptance of QSP Modeling

ARTICLE

Quantitative systems pharmacology: Landscape analysis of regulatory submissions to the US Food and Drug Administration

Jane P. F. Bai | Justin C. Earp | Jeffry Florian | Rajanikanth Madabushi | David G. Strauss | Yaning Wang | Hao Zhu

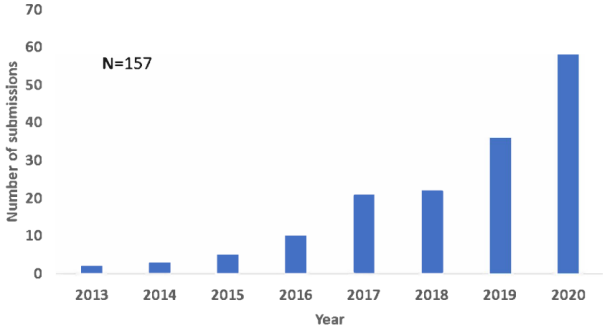
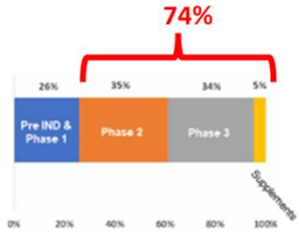


FIGURE 1 The yearly trend of quantitative systems pharmacology submissions since 2013

QSP Submissions Equally Span Different Phases of Drug Development



- 66% were efficacy related submissions,
- 97% supporting dosing:
  - 27% phase 1, 39% phase 2, 18% phase 3 & PSP, 12% NDA/BLA
- 34% were safety related: liver toxicity, bone density, cariotoxicity, plasma ion conc

Oncology: largest TA 32%

.fda.gov

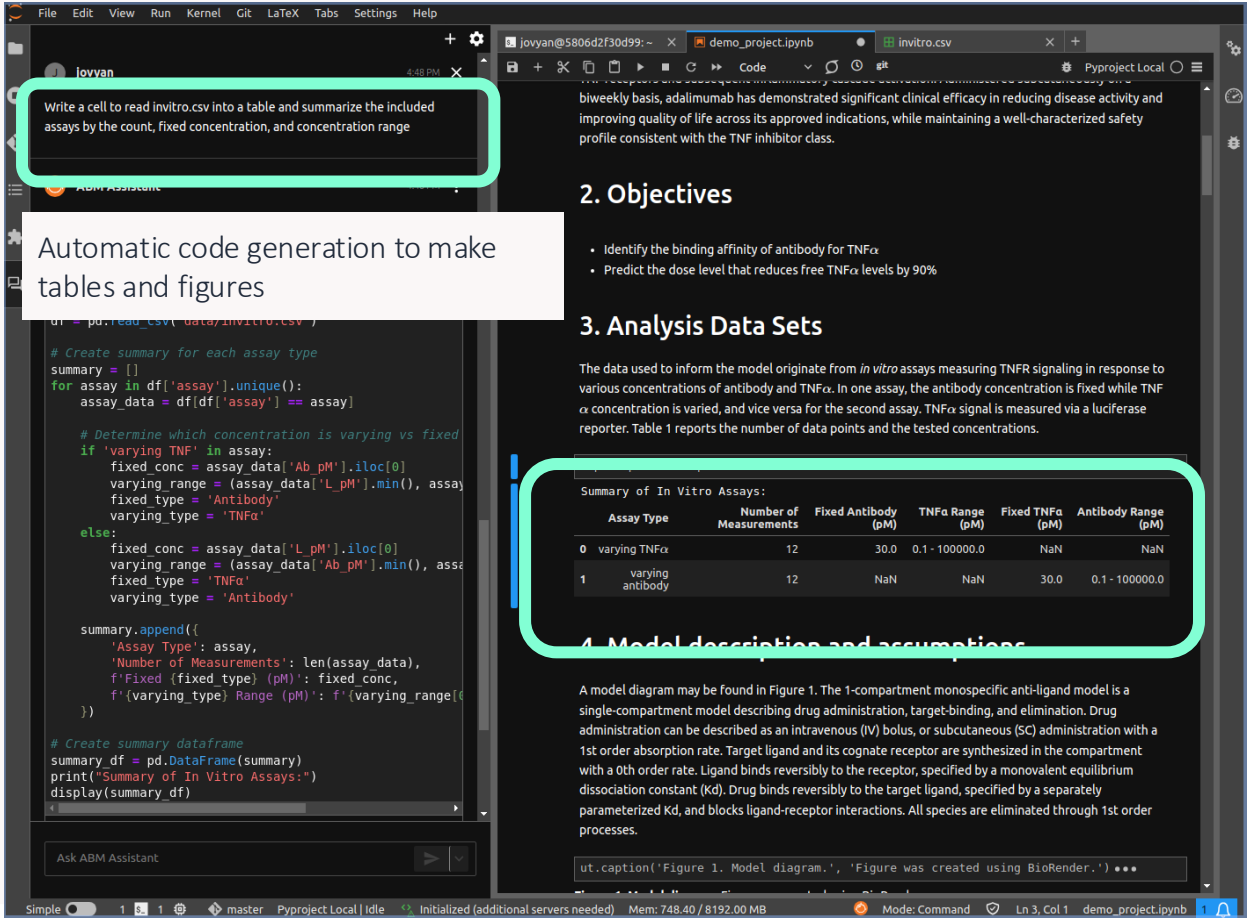
CPT Pharmacometrics Syst Pharmacol. 2021;10:1479–1484.

CERTARA

Automatic Code Generation

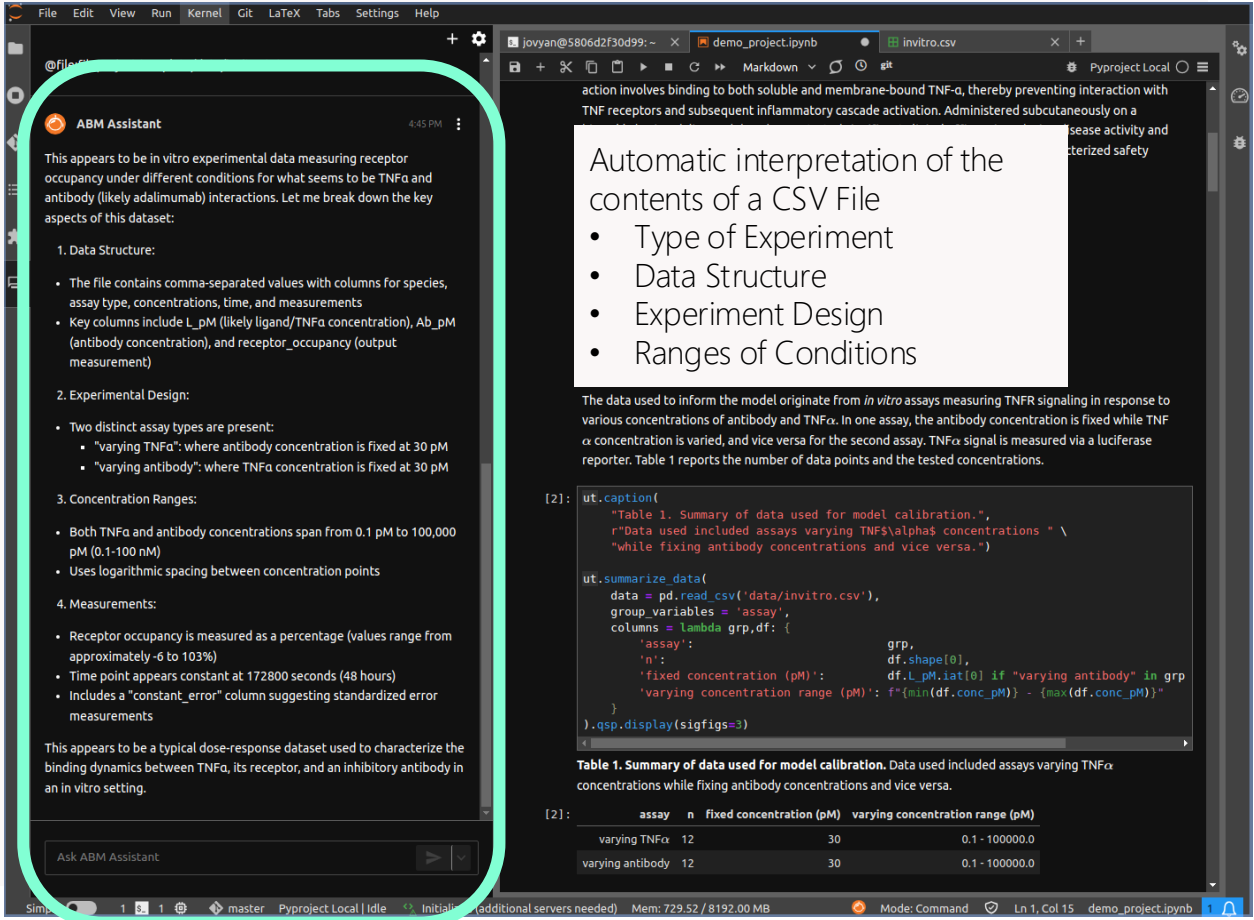
Generative AI to accelerate model analysis scripting

- Helps simplify repeatable tasks
- Can help modelers more quickly parse complex model scripts
- Simplifies the conversion of models from one code format to another



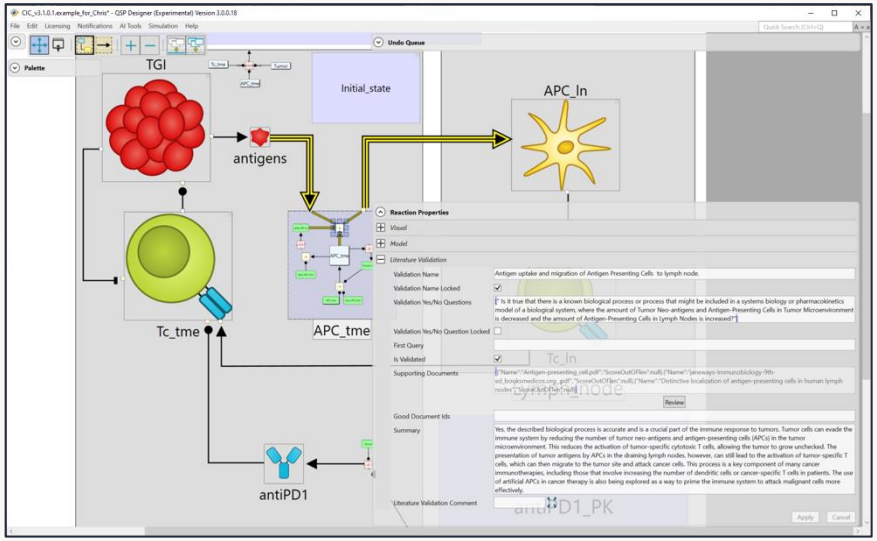
Orientation to data sets

AI can look at a CSV file and provide the modeler an overview of what's in their including a guess at the underling experiment design



CERTARA

Increase credibility with biology review by an AI expert



The credibility of a QSP model depends on how accurately the model represents the known biology

AI tools make it practical to do this at scale

Automatically generated GPT prompts based on model:

"Is it true that there is a known biological process or process that might be included in a systems biology or pharmacokinetics model of a biological system, where the amount of Tumor Neo-antigens and Antigen-Presenting Cells in Tumor Microenvironment is decreased and the amount of Antigen-Presenting Cells in Lymph Nodes is increased?"

AI generated summary of biological evidence:

Yes, the described biological process is accurate and is a crucial part of the immune response to tumors. Tumor cells can evade the immune system by reducing the number of tumor neo-antigens and antigen-presenting cells (APCs) in the tumor microenvironment. This reduces the activation of tumor-specific cytotoxic T cells, allowing the tumor to grow unchecked. The presentation of tumor antigens by APCs in the draining lymph nodes, however, can still lead to the activation of tumor-specific T cells, which can then migrate to the tumor site and attack cancer cells. This process is a key component of many cancer immunotherapies, including those that involve increasing the number of dendritic cells or cancer-specific T cells in patients. The use of artificial APCs in cancer therapy is also being explored as a way to prime the immune system to attack malignant cells more effectively.

Traceability to the primary literature:

are carried by lymph flow to the draining lymph nodes . Each lymph node is a collection point where APCs can interact with T cells.[1] During the migration , DCs undergo a process of maturation : they lose most of their ability to further engulf pathogens and they mature by changing surface expression of MHC and co - stimulatory molecules , as well as increased production of cytokines . The internalized antigen is digested into smaller peptides containing epitopes , which are then presented to T cells by the MHC.[2][12]

CERTARA

Predicting Clinical Outcomes from QSP Models (Chung et al.)

Quantitative systems pharmacology: Landscape analysis of regulatory submissions to the US Food and Drug Administration

Jane P. F. Bai | Justin C. Earp | Jeffry Florian | Rajanikanth Madabushi | David G. Strauss | Yaning Wang | Hao Zhu

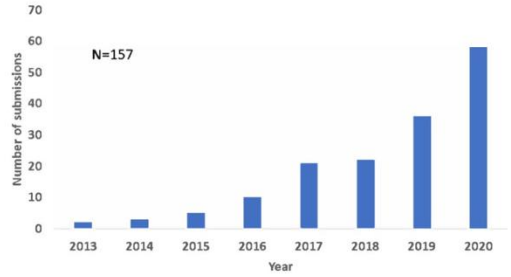
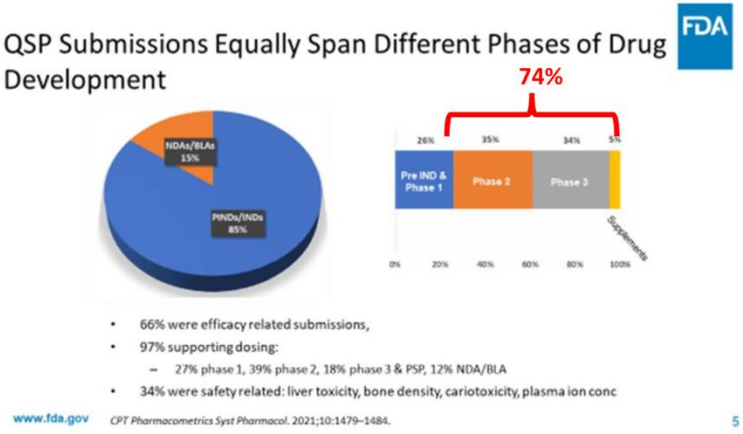


FIGURE 1 The yearly trend of quantitative systems pharmacology submissions since 2013



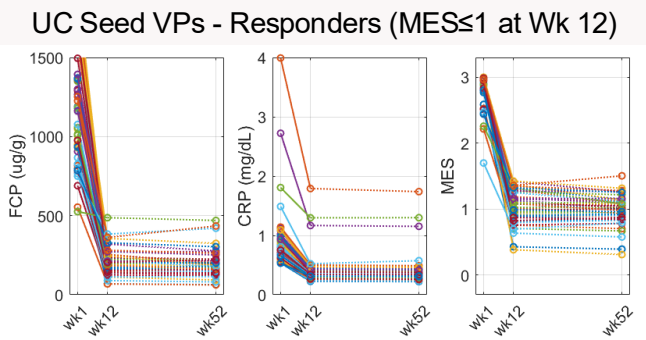
Inflammatory Bowel Disease
QSP MODEL

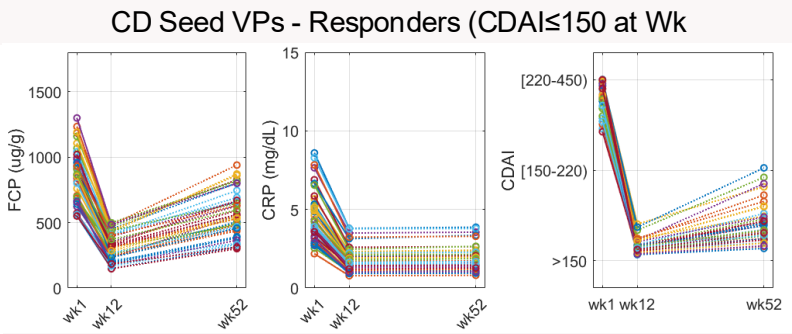
→ **Biomarkers**

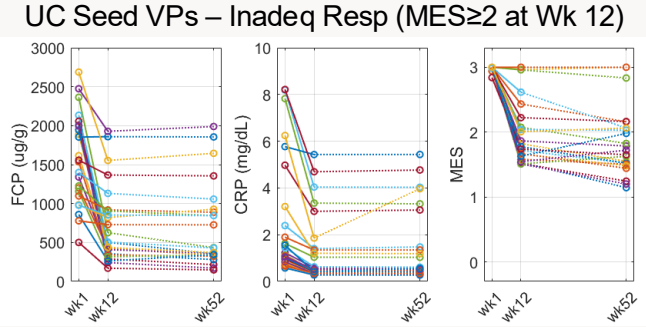
Clinical Endpoints

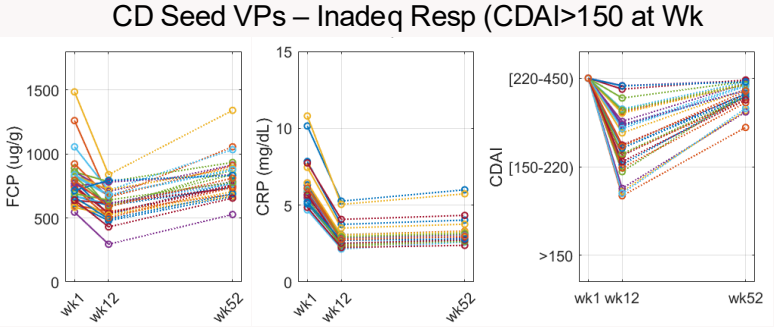
**Presented at ACoP14, Nov. 2023*

The IBD QSP Model Matches Clinical Trials









The IBD QSP model VPs and Vpop capture observed variability at pre- and post-treatment using various treatment modalities

IBD QSP Model matches clinical trial studies:

RIPK1 Inhibitor, GSK 60mg TID
(Weisel et al. 2021)

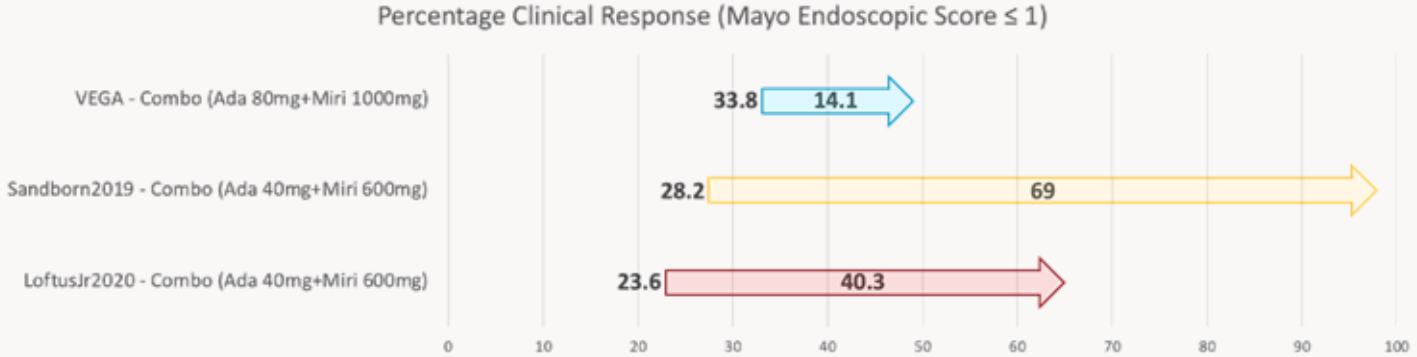
Anti-IL-23p19, Mirikizumab 600mg SC Q4W
(Sandborn et al. 2019)
(Sands et al. 2022)

Anti-TNFα, Adalimumab 160 mg/80 mg SC Q2W
(Loftus Jr et al. 2020)
(Shinzaki et al. 2021)

Anti-TNFα and Anti-IL-23, VEGA trial
(Feagan et al. 2023)

Predicting Clinical Outcomes from QSP Models (Chung et al.)

UC Clinical Trials	Adalimumab 80mg Q2W	Mirikizumab 600mg Q4W	Mirikizumab 1000mg Q4W	Ada 40mg + Miri 600mg	Ada 80mg + Miri 1000mg
VEGA (Feagan et al. 2023) anti-TNFα/anti-IL-23				Original Design ⇨	Increase Dose
Sandborn et al. 2019 Mirikizumab		Original Design →		Add Ada	
Loftus Jr et al. 2020 Adalimumab	Original Design →			Add Miri	
	Anti-TNFα monoRx	Anti-IL-23 monoRx		Anti-TNFα / Anti-IL-23 comboRx	



- The IBD QSP model predicts that the percentage clinical response of patients in previous clinical trials could have been increased with alternate dosing regimens

CERTARA

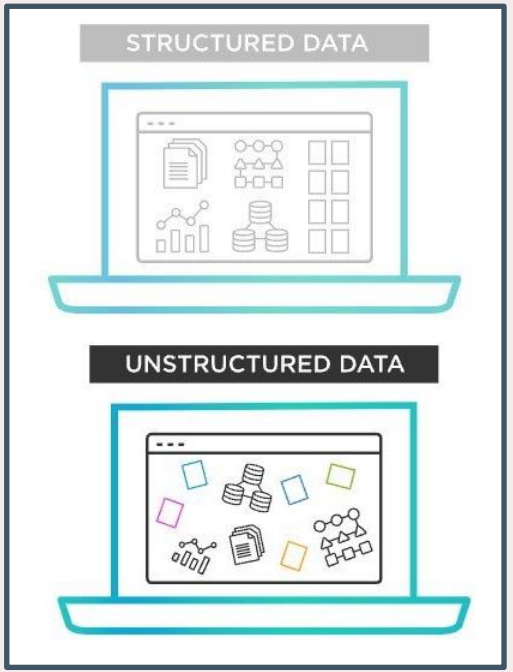
Unstructured Data Extraction

CERTARA

The Challenge of Unstructured Data

Making the “unusable” usable

- Most analyses rely on “structured” data
- However, **~80% of enterprise data is unstructured**, i.e., “trapped” in text, images, PDFs, etc.
- This content is still important, but it’s **hard to access and analyze**
- Our goal, then, is to **extract key data** to present alongside structured content



CERTARA

Granular Data Extraction via GPTs

Add structure to unstructured data sets

- GPTs can extract highly targeted data points utilizing natural language questions/prompts
- Coupled with RAG, we can automatically generate datasets with content linked back to source documents
- This allows researchers to “read a bunch of papers” without needing to read them

What is the C_{max}?

What are the reported efficacy outcomes?

Describe the route of administration.

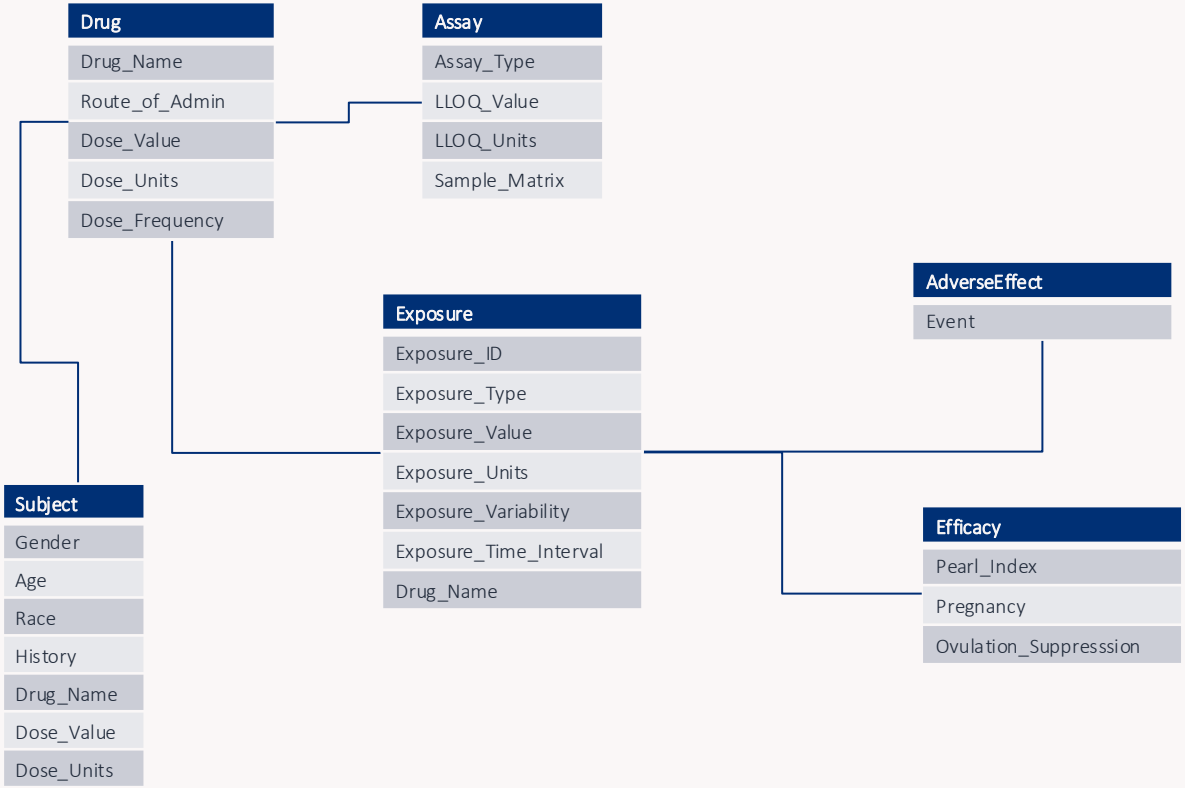
List the drugs mentioned in this study.

Were there any adverse events?

What is the dose frequency?

CERTARA

Data relationships need to be preserved when extracting data points



An example schema of some PK/PD data; without this context individual numbers are meaningless!

CERTARA

GPTs Can Programmatically Model Data Relationships

Utilizing Chained GPT Prompts Preserves Relationships Inherent in the Data

- The further context that can be given to a GPT, whether through specificity in the prompt or through RAG, significantly enhances its ability to draw inferences
- By chaining prompts, i.e. using the answer from one prompt to provide the basis for another, we can begin to establish the relationships between data points that are disjointed/separated within the text
- With some minor tweaking and prompt engineering, we can begin to programmatically extract granular data points while preserving the relationships between them from raw text (ex. drugName >> drugDose)

Python Example:

```
def run_extraction(source_id):
    ex = Extractor(test_settings)
    ex.run_top_query(query=f"You extracted these sections from a clinical trial document: {context[id]}\nWhat study arms were reported in the PK sect
                     formatting_instructions="Always format your answer with this typescript definition: {'answer': list}",
                     sources=[],
                     var_nam="arms",
                     output_type="list")
    ex.run_top_query(query=f"You extracted these sections from a clinical trial document: {context[id]}\nWhat metrics were recorded in the PK sections
                     formatting_instructions="Always format your answer with this typescript definition: {'answer': list}",
                     sources=[],
                     var_nam="endpoints",
                     output_type="list")
    ex.run_top_query(query=f"You extracted these sections from a clinical trial document: {context[id]}\nWhat drugs are being studied?  If there is a
                     formatting_instructions="Always format your answer with this typescript definition: {'answer': list}",
                     sources=[],
                     var_nam="drugs",
                     output_type="list")
    ex.run_chain_query(query=f"You extracted these sections from a clinical trial document: {context[id]}\n"+"For the reported {} endpoint what were t
                     formatting_instructions="Always format your answer with this typescript definition: {'answer': list}",
                     sources=[],
                     top_var_nams=["endpoints"],
                     var_nam="metrics",
                     output_type="list")

    ex.run_chain_query(query=f"You extracted these sections from a clinial trial document: {context[id]}\n"+"For the {} arm of the drug {} what was th
                     formatting_instructions="Always format your answer with this typescript definition: {'answer': str}",
                     sources=[],
                     top_var_nams=["arms","drugs", "endpoints"],
                     var_nam="metrics",
                     output_type="string")
```

CERTARA

Levonorgestrel Benchmarking: Granular Results

Data Extraction Pipeline Performance

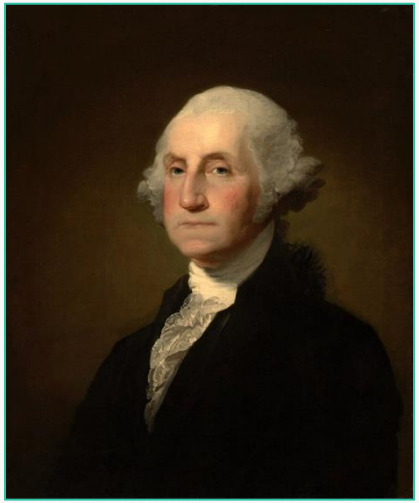
Data Point	Precision	Recall	Content Extracted
Drug Information	~89	~80	Drug name, active ingredients, dose, frequency, route,
Efficacy Information	~55	~50	Metric used, value, interval
Assay Information	~84.5	~77.5	Assay(s) used, Sample matrix
Exposure Information	~90	~57	Endpoints, values and units, interval, S.D.
Safety Information	~80	~65	Adverse Event types, number of patients, when event occurred.

Evaluation was conducted using 13 publications – a mix of trials, literature reviews, and meta-analysis. A team created manual annotations for all papers and these results were compared with model output.

Agentic AI in Report Writing

What do people mean by “Agentic AI?”

And how is that different from “generative” AI?



Generative: creates something new

“Write me a poem about George Washington.”

Agentic: uses reason to solve a complex task

“Do I need to bring an umbrella on a trip to Barcelona?”

CERTARA

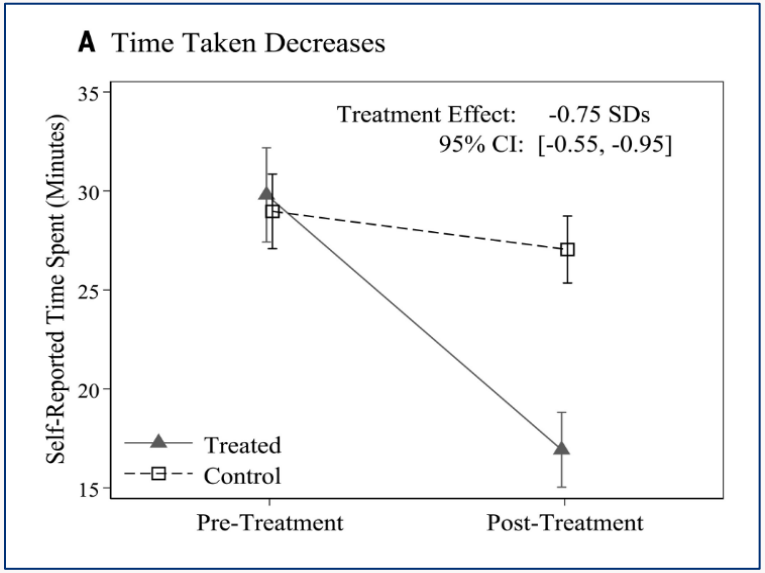
Limitations in the sciences

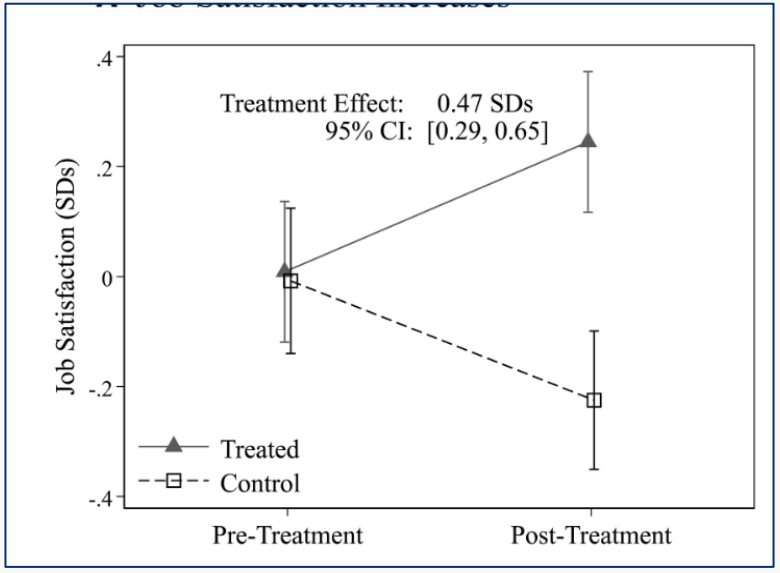
AI is like a colleague; you need to be able to check their work!

- Agents are “black boxes”
- Scientific use cases may have multiple potential solutions
- Reasoning relies on previous pattern recognition, not scientific knowledge
- GPTs are non-deterministic by design
- Agents require access to other systems and data

CERTARA

Improving Writing Productivity & Job Satisfaction





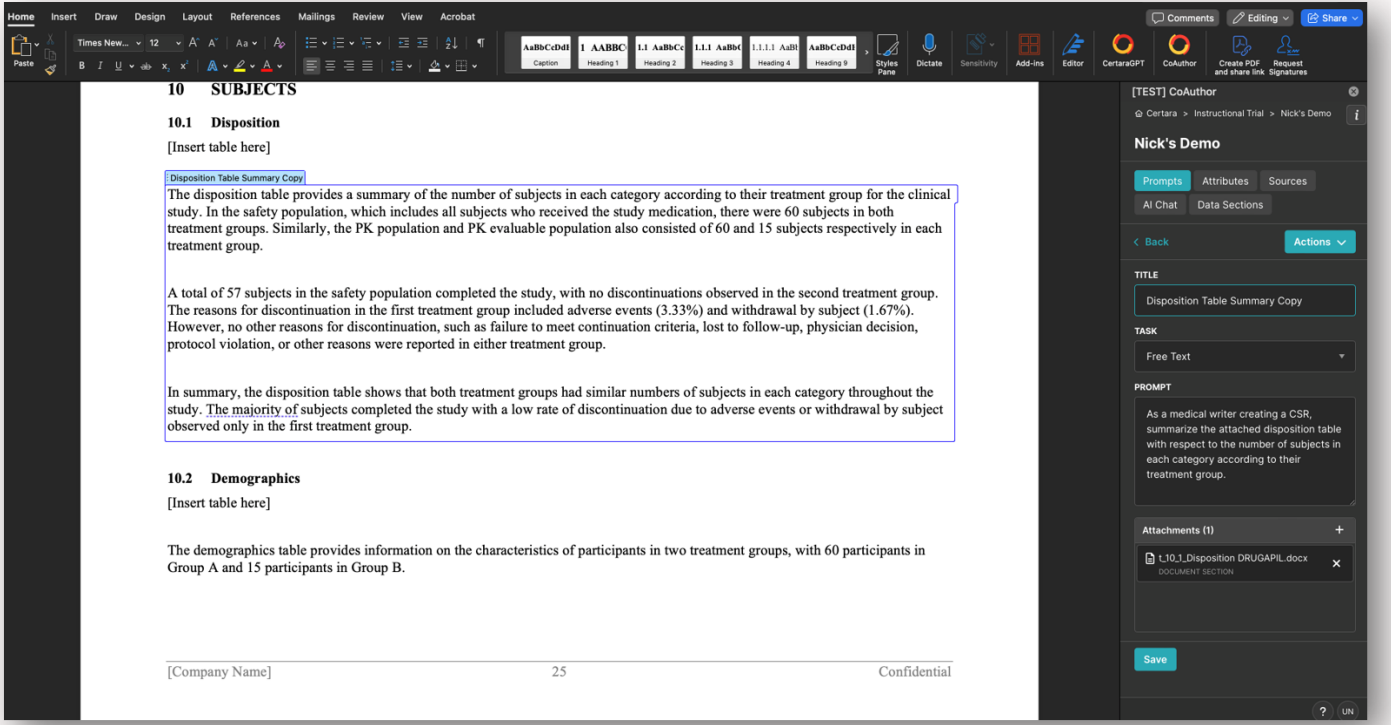
Randomized experiment shows leveraging GPTs for professional writing tasks present a 40% productivity increase while improving output quality by 18%.

Writers exposed to GPTs reported higher job satisfaction.

Shakked Noy, Whitney Zhang, Experimental evidence on the productivity effects of generative artificial intelligence. Science **381**, 187-192 (2023). DOI: 10.1126/science.adh2586

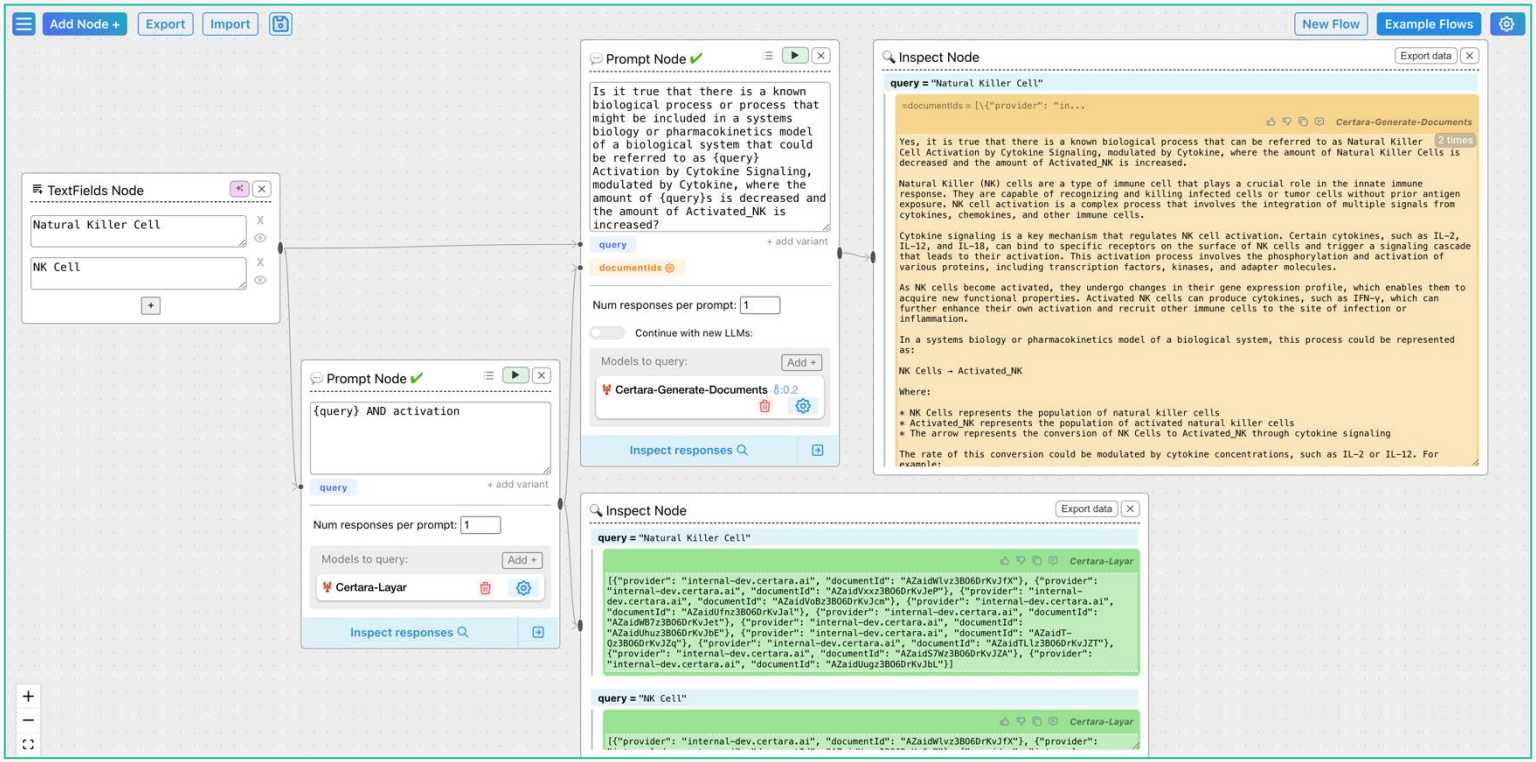
CERTARA

Regulatory Documents Are Templated by Design



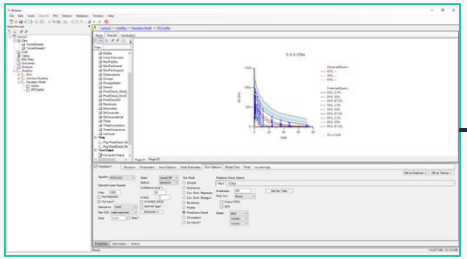
CERTARA

Chaining Activities Models Agentic Behavior

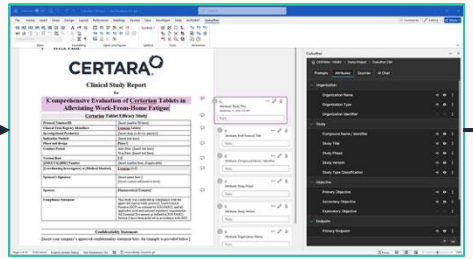


Chaining Activities Models Agentic Behavior

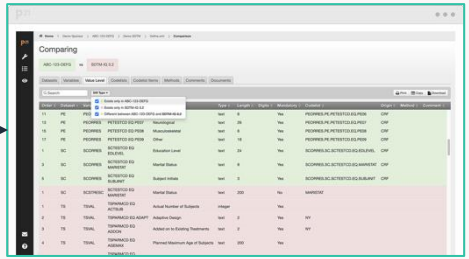
Creating a toxicology report



Scientist analyzes findings from study



AI generates report from findings and other content



Human adds reviewed report to submission

Summary

- GPT technologies are a novel and valuable addition to the life sciences toolkit
- While we may go through a hype cycle regarding GPTs, now is the time for life sciences, healthcare and governmental organizations to identify and derive value from these technologies
- RAG architectures enable specialized, secure, highly effective deployment of GPT technologies for life sciences specific tasks at a level that is similar to large GPT models and for a fraction of the cost
- There are numerous highly specialized applications of GPT technologies in the life sciences space. Identifying areas of value followed by focused application in these specialized areas is the critical next step for organizations who want to harness these technologies
- Expert knowledge of critical use cases is just as important now as it was before the advent of these technologies. Working with organizations that have the expertise in GPT technologies AND specialized use cases is critical

Questions?

Thank you!