Why Choose One When We All Are Friends: The Power of Multiple AI Tools in Clinical Trials

PHUSE SDE 2024 – CARY

Hari Jampala

Fortrea

Disclaimer

The content of this presentation is for knowledge sharing and information purpose only. The content is developed through research process. The information presented here represents the work of author and should not be considered in general as a views of author’s organization.

Fortrea

An accomplished Drug Development and Statistical Programming Professional and leader, currently serving as Senior Manager of Statistical Programming at a Fortrea. With deep expertise in statistical programming within the pharmaceutical industry and CROs, this subject matter expert (SME) excels in CDISC activities, regulatory submissions, and process automation. A proven leader, manage remote teams, drive cross-functional process improvements, and lead projects to successful completion. Equipped with a strong working knowledge of clinical trial phases, proficient in SAS R, Python, and PowerBI.

Harivardhan.jampala@fortrea.com

Fortrea

Embracing AI: From Caution to Collaboration in Statistical Programming

Industry Estimates: Overall, there may be an estimated 15,000 to 25,000 statistical programmers employed worldwide across both sectors. This number accounts for roles like biostatistics, data management, and other related functions that often require similar programming skills.

SPSS → **SAS** → **Opensource(R, Python, Scala, Julia, java script)** → **AI** → **GENAI**

Imagine what is the mindset of the programmers?

Help in changing the mindset to reshaping workflows rather than replacing the need for skilled programming professionals.

Enhanced Efficiency and Speed

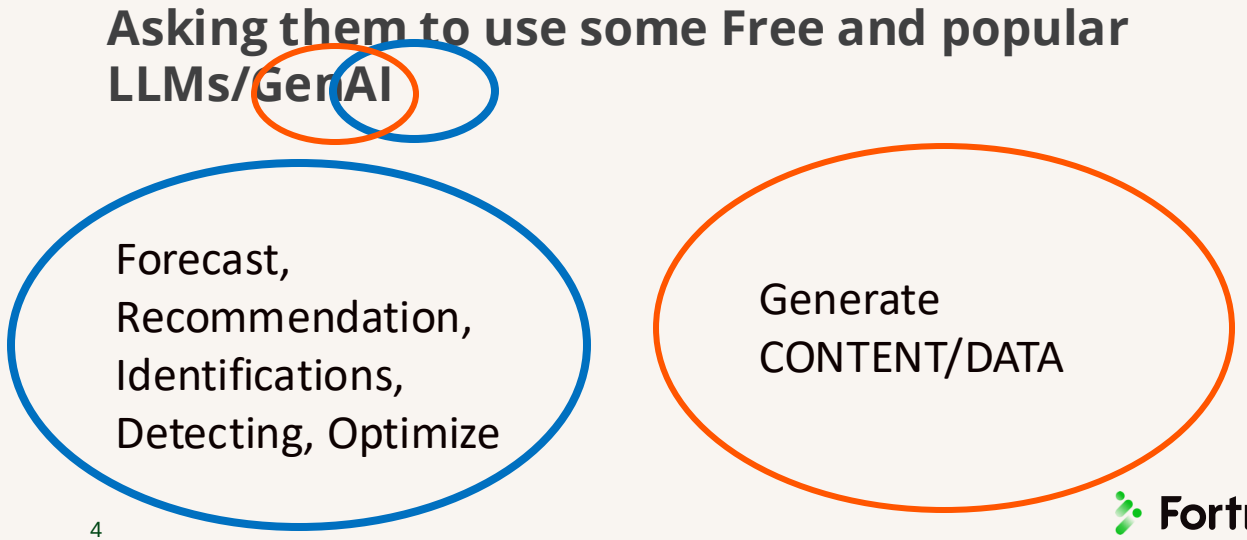
Focus on Practical, Non-Coding Applications of AI

Emphasize AI as a Tool, Not a Replacement

Integrate AI into Existing Workflows Gradually

Promote Collaborative AI Development Projects

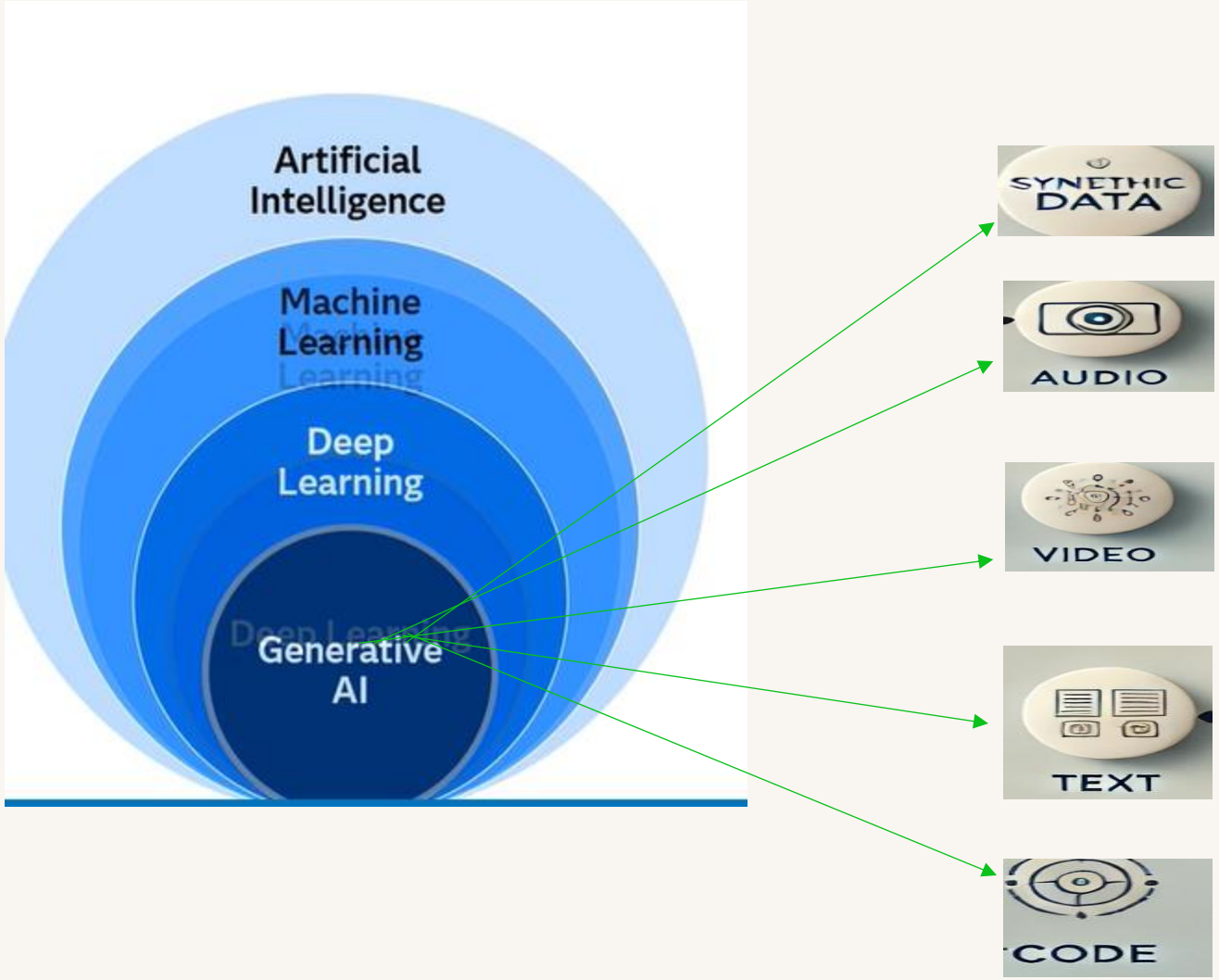
Highlight Career Growth Opportunities with AI



Fortrea

GENAI/ LLM

GENAI – Transforms the interaction between HI and AI



LARGE LANGUAGE MODEL (LLM)

LARGE = Learns from hundreds of billions of data

LANGUAGE = Understands human natural language

ChatGPT

Timeline

GPT

Published by OpenAI in partnership with Microsoft

2018	2019	2020	2023
GPT-1	GPT-2	GPT-3	GPT-4
117M parameters	1.5B parameters	175B parameters	1.7T parameters
3.3B Tokens	10B Tokens	300B Tokens	? Tokens

GENAI is much more than ChatGPT.

Fortrea

Opportunities

Clinical research

Dose range findings

Drug Discovery – Drug target identifications, selections, compound screening and design

Site selection, recruitment, retention, adherence,

Non-clinical – PK/PD and Dose range findings

Data collection, management, analysis

Clinical end point analysis

Manufacturing and post market safety monitoring- advanced pharmaceutical manufacturing, post market safety surveillance or pharmacovigilance

Biostats and Statistical programming:

- CRF and edit check configurations
- Validation and UAT
- Data analysis and visualization
- Statistical programming
  - Draft text like for SAP, CSR
  - Draft codes documentation or creation
- Documentation and TMF

Fortrea

How? Your New Programming Partner: How AI Enhances, Not Replaces

HOW? Roadmap might be

- Understand Artificial Intelligence Basics
- Learn Math
- Learn Programming Language
- Learn Big Data
- Learn Data Science
- Learn Machine Learning Algorithms
- Learn Deep Learning Algorithms
- Learn Large Language Models

Highlight the Competitive Advantage

Encourage Small, Low-Stakes Experiments

LinkedIn -> 70% of this year list did not exist 20 years ago -> next 10 years ? -> 20 don’t exist

Fortrea

Encourage Small, Low-Stakes Experiments

4. Scala (Spark)

In Scala, particularly with Apache Spark, you use `join` on DataFrames. This example assumes `df1` and `df2` are Spark DataFrames.

scala Copy code

```scala
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._

val spark = SparkSession.builder.appName("MergeExample").getOrCreate()
val merged_data = df1.join(df2, Seq("ID"), "inner")
```

5. Julia

In Julia, you can use the `DataFrames.jl` package, which has a `join` function for merging.

julia Copy code

```julia
using DataFrames

# Assuming df1 and df2 are DataFrames
merged_data = innerjoin(df1, df2, on = :ID)
```

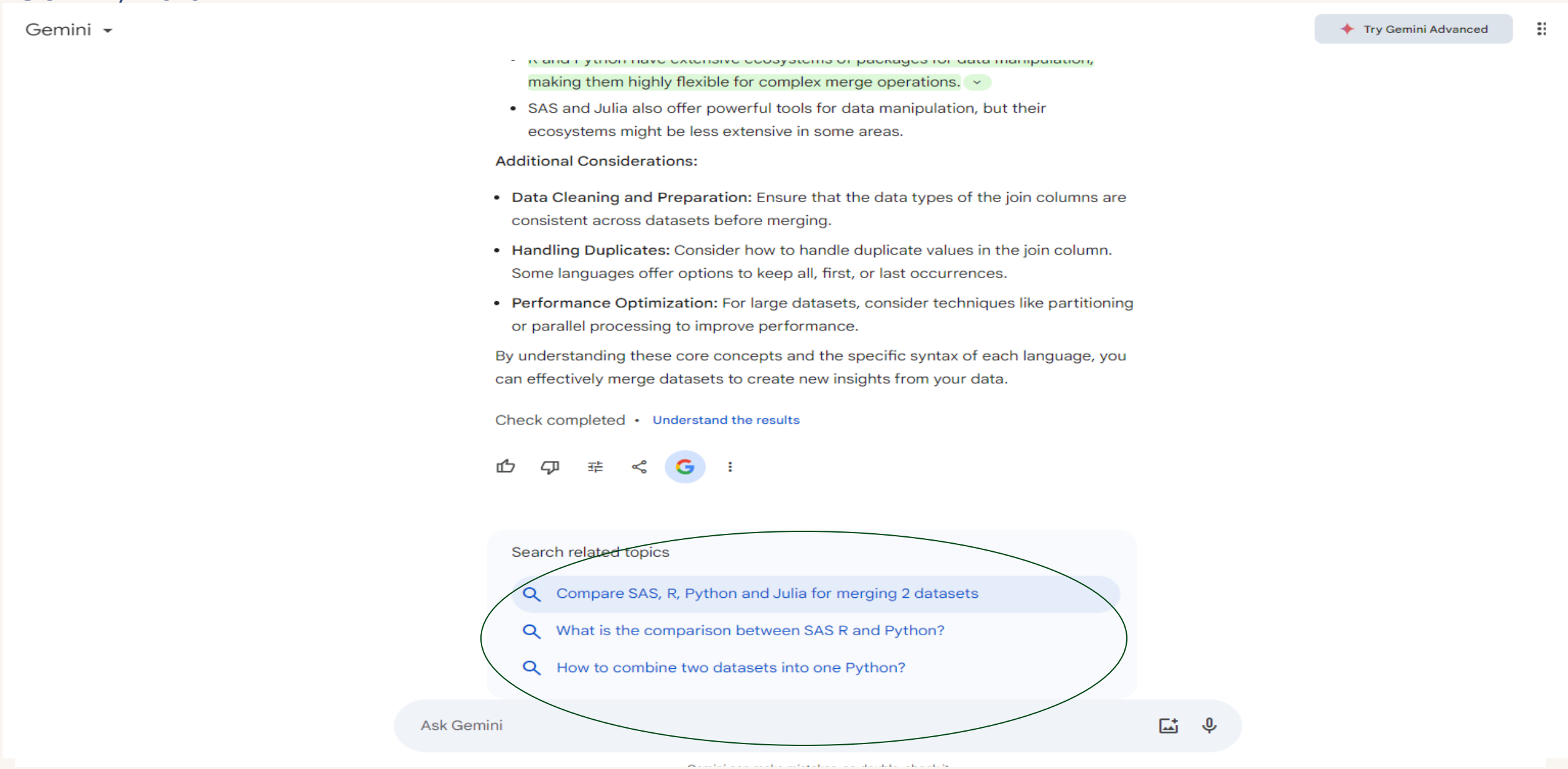
Summary

Language	Function/Method	Syntax
SAS	`MERGE` or `PROC SQL JOIN`	`MERGE df1(in=a) df2(in=b);`
R	`merge` or `dplyr::inner_join`	`merge(df1, df2, by = "ID")`
Python	`pandas.merge`	`pd.merge(df1, df2, on="ID")`
Scala	`Spark DataFrame.join`	`df1.join(df2, Seq("ID"), "inner")`
Julia	`DataFrames.innerjoin`	`innerjoin(df1, df2, on = :ID)`

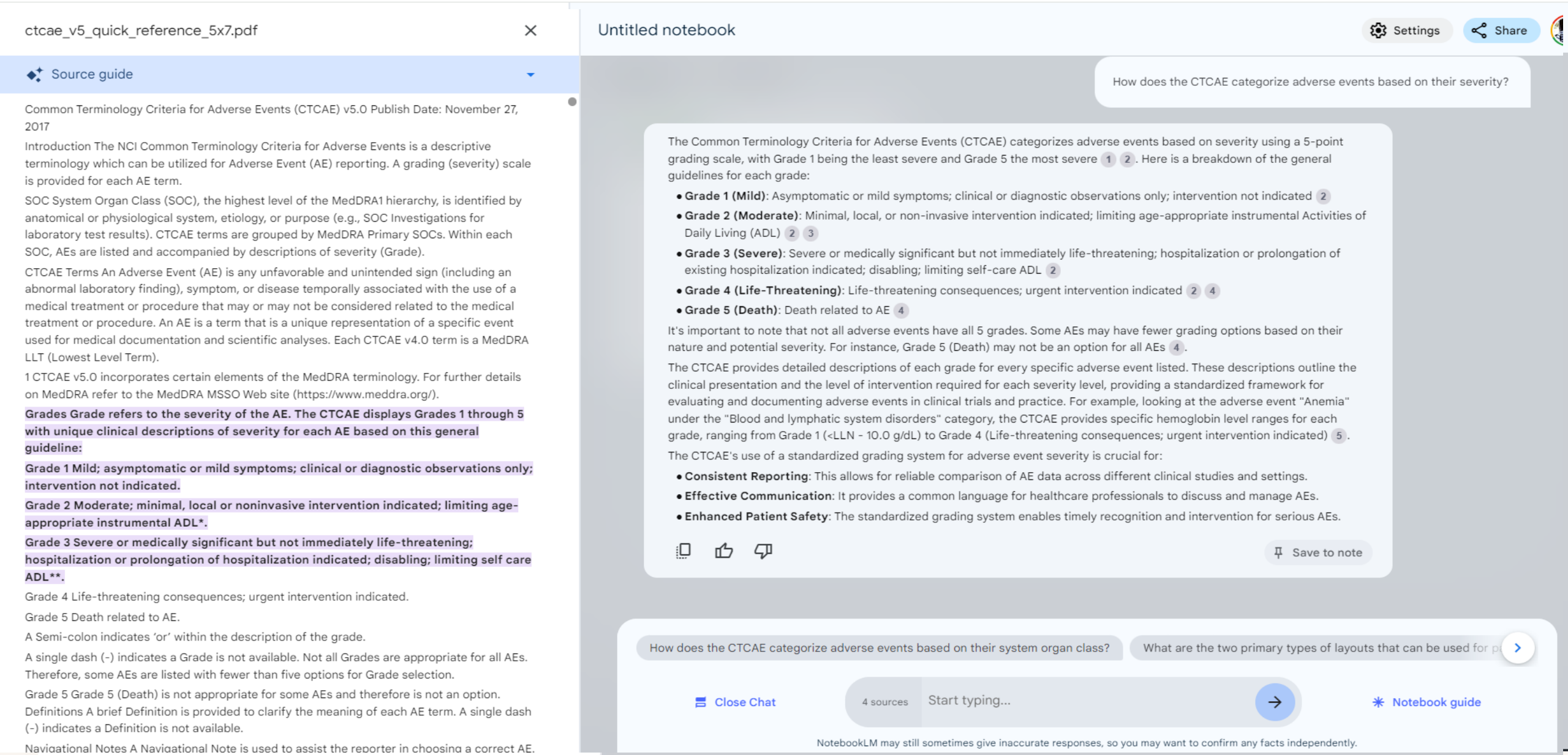
Fortrea

Encourage Small, Low-Stakes Experiments

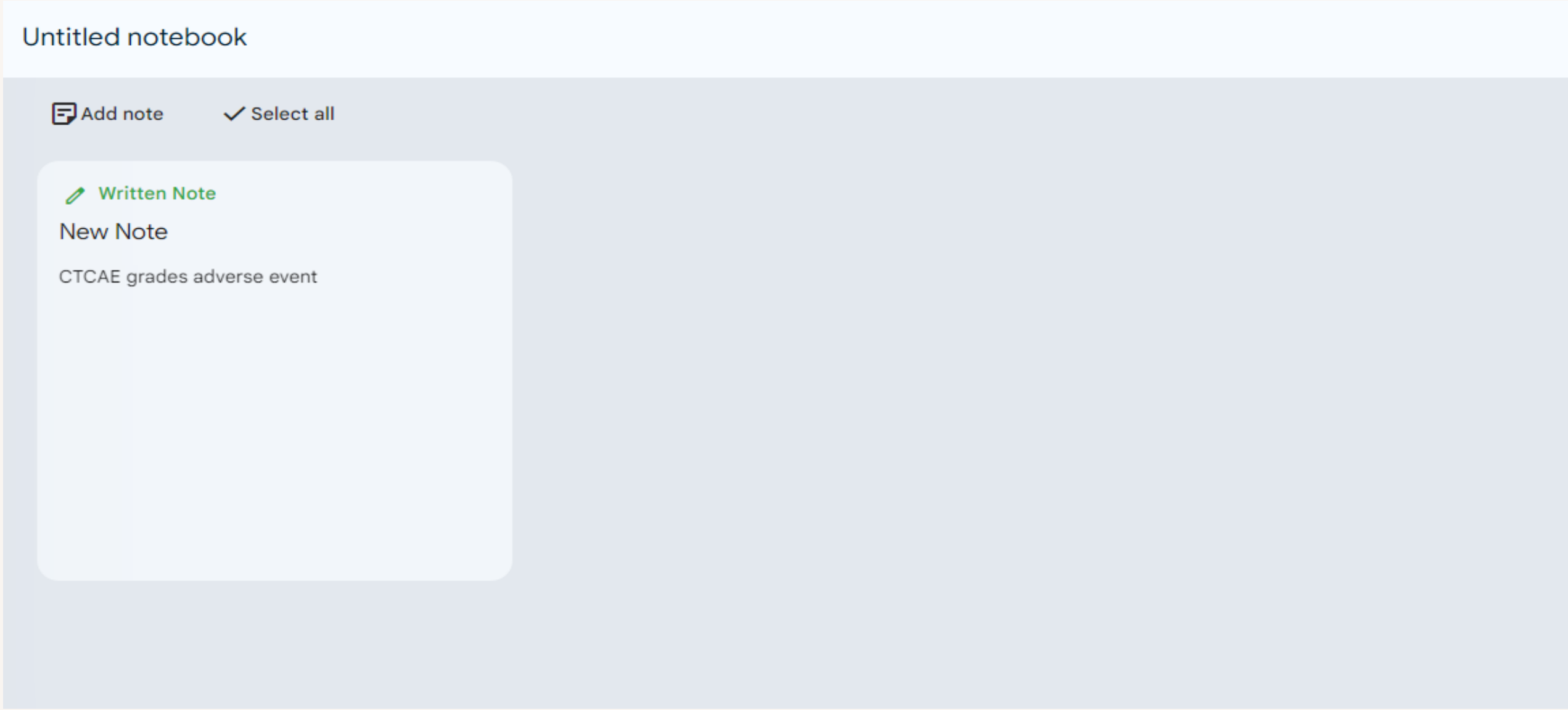
Gemini/ Bard



Other LLMs?

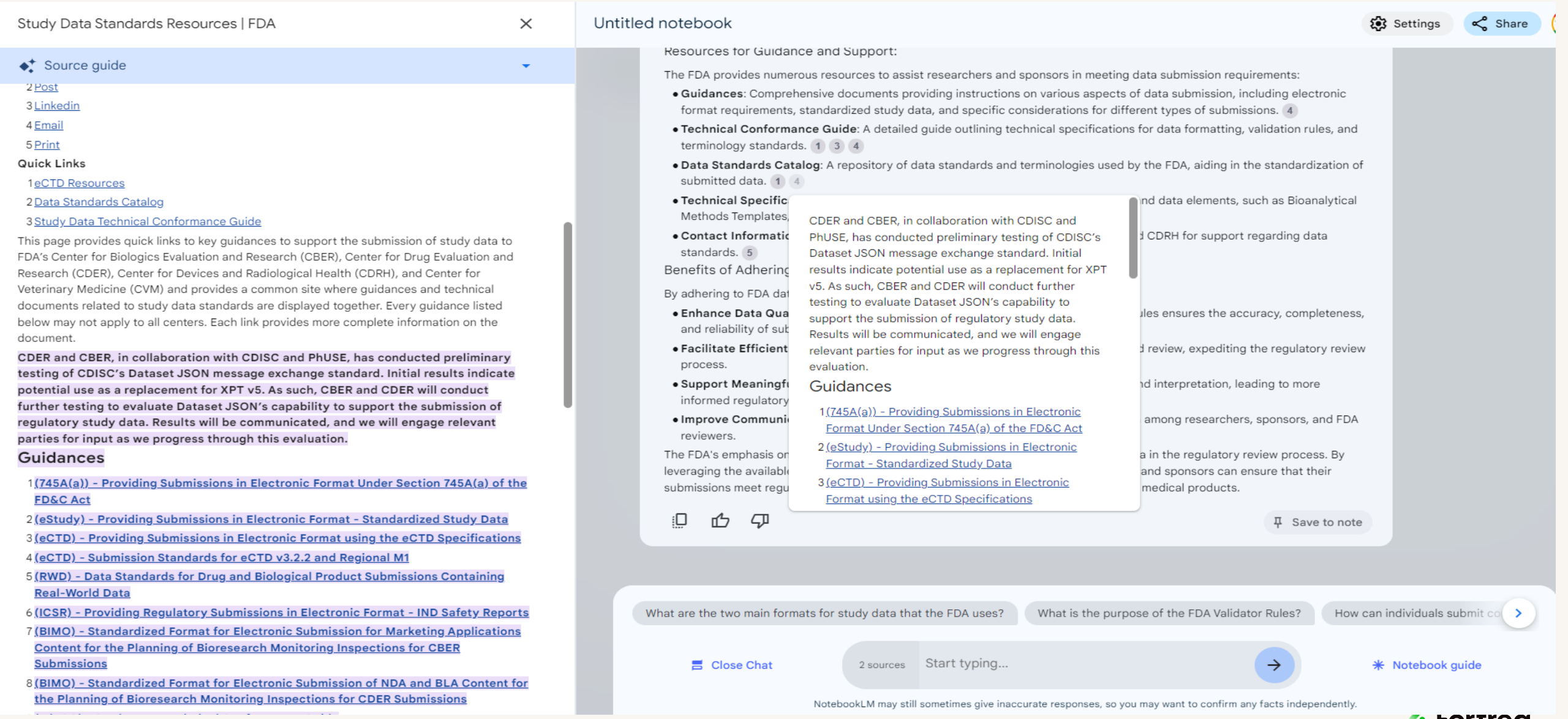


Other LLMs?



Fortrea

Other LLMs?



Fortrea

Other LLMs?

Microsot Phi3.5 - https://ollama.com/library/phi3.5

A lightweight AI model with 3.8 billion parameters with performance overtaking similarly and larger sized models.

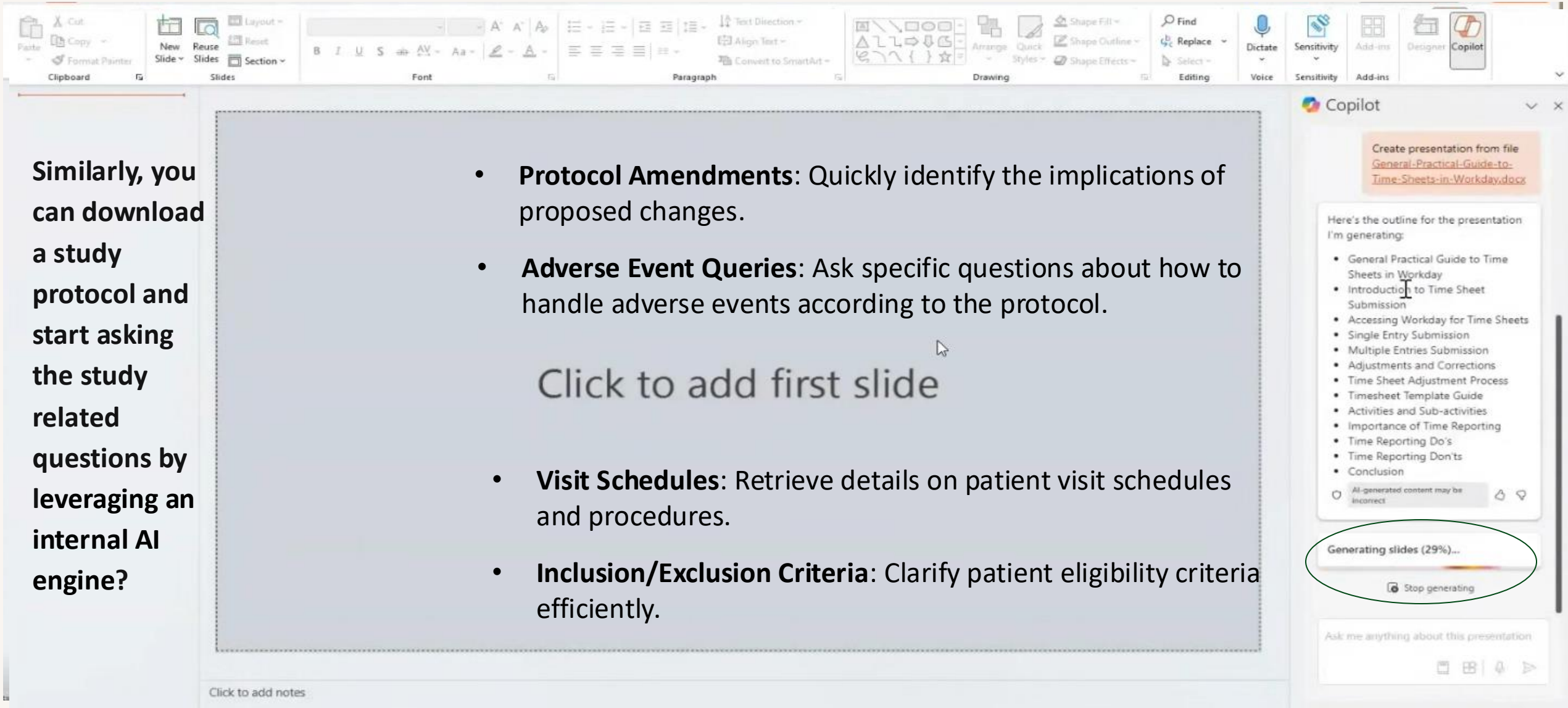
Phi-3.5 models have a context length of up to 128K, allowing them to handle **long-context tasks effectively**. This can be beneficial for processing lengthy clinical research documents or patient records.

Phi -
3.5 models have shown superior performance in reasoning and math benchmarks compared to other small models,
This makes them highly efficient and capable of handling complex tasks with reduced latency

Co-Pilots, Labs, and Tutors

Microsoft Co-pilot

Similarly, you can download a study protocol and start asking the study related questions by leveraging an internal AI engine?



- **Protocol Amendments**: Quickly identify the implications of proposed changes.
- **Adverse Event Queries**: Ask specific questions about how to handle adverse events according to the protocol.
- **Visit Schedules**: Retrieve details on patient visit schedules and procedures.
- **Inclusion/Exclusion Criteria**: Clarify patient eligibility criteria efficiently.

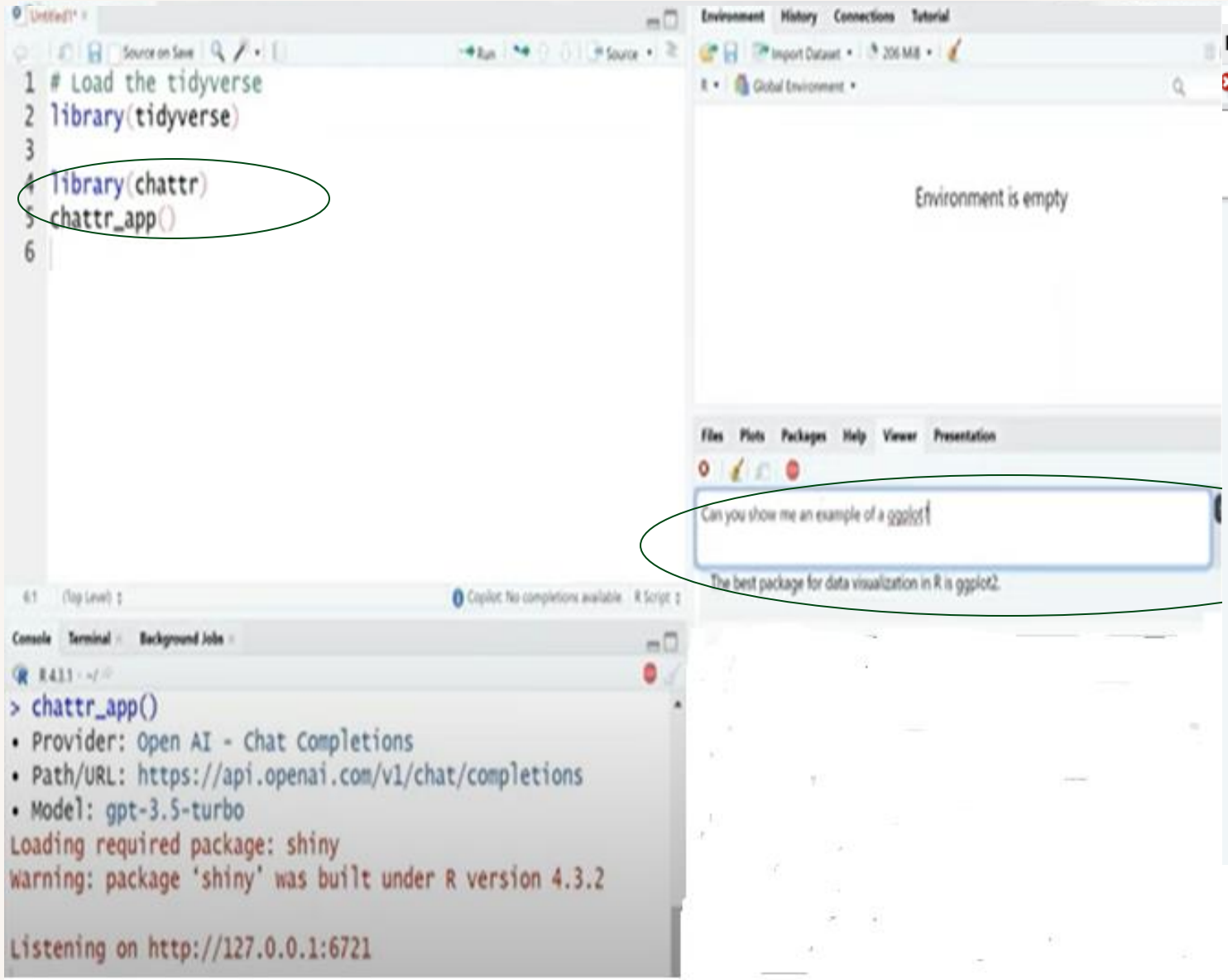
Fortrea

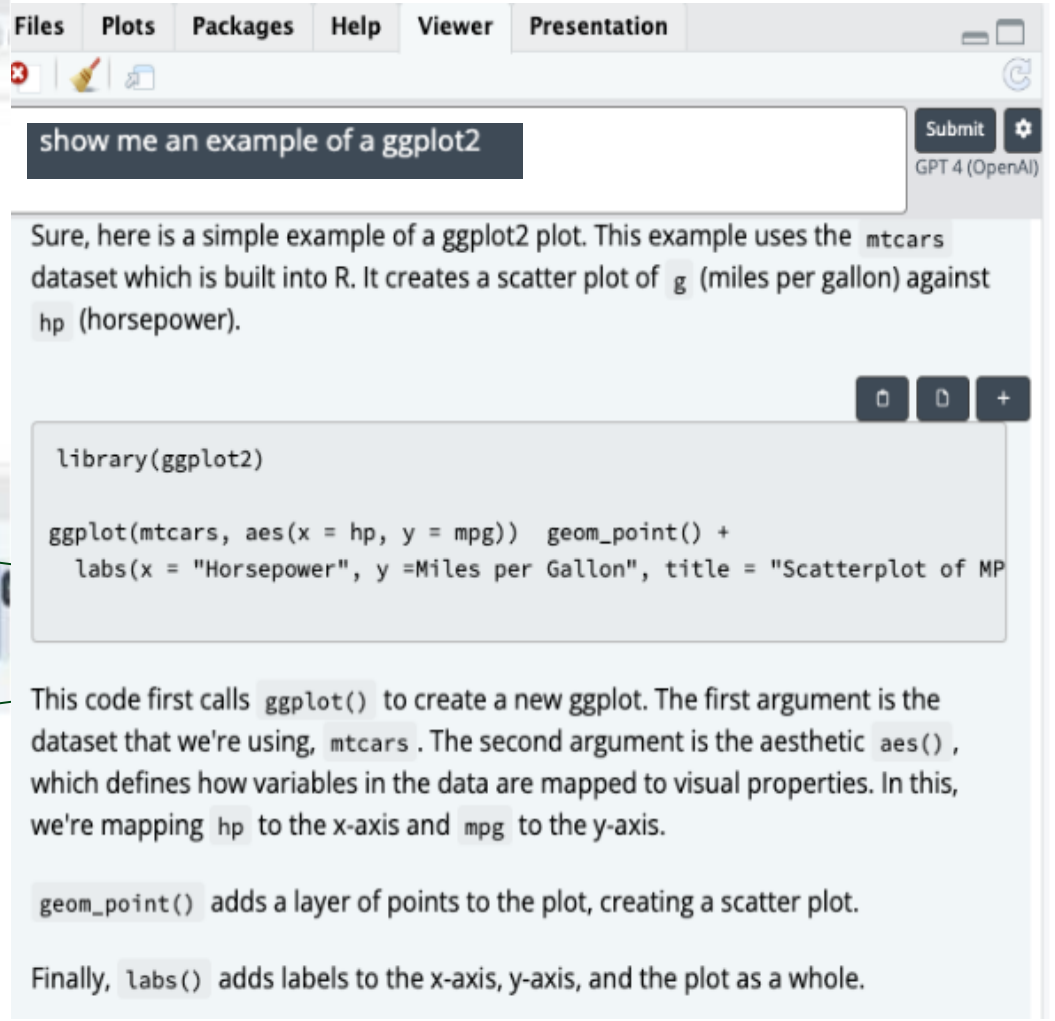
Co-Pilots, Labs, and Tutors

Github Copilot

Fortrea

Co-Pilots, Labs, and Tutors





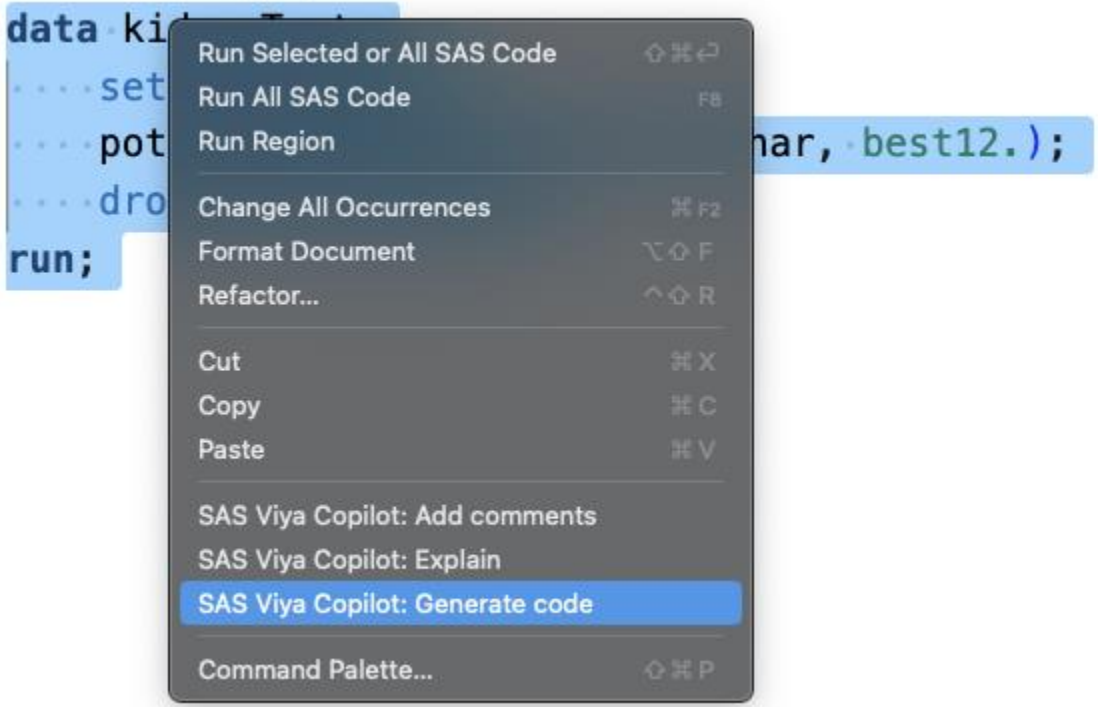
Co-Pilots, Labs, and Tutors

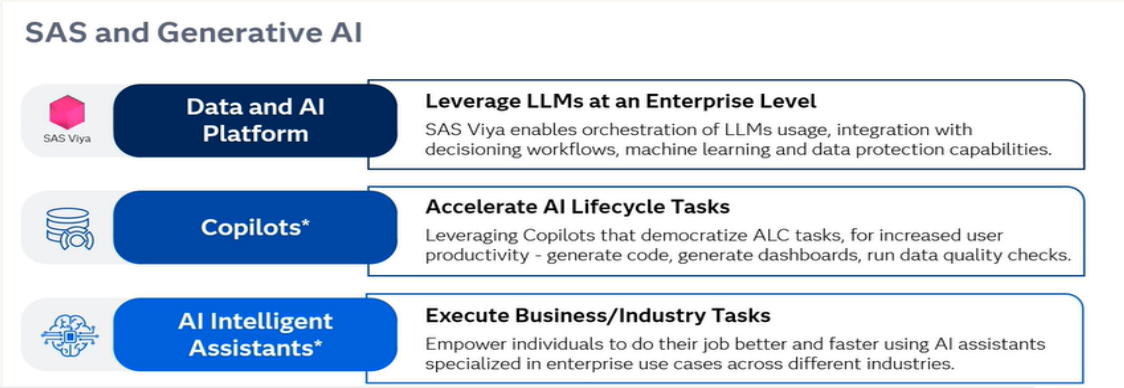
SAS CO pilot

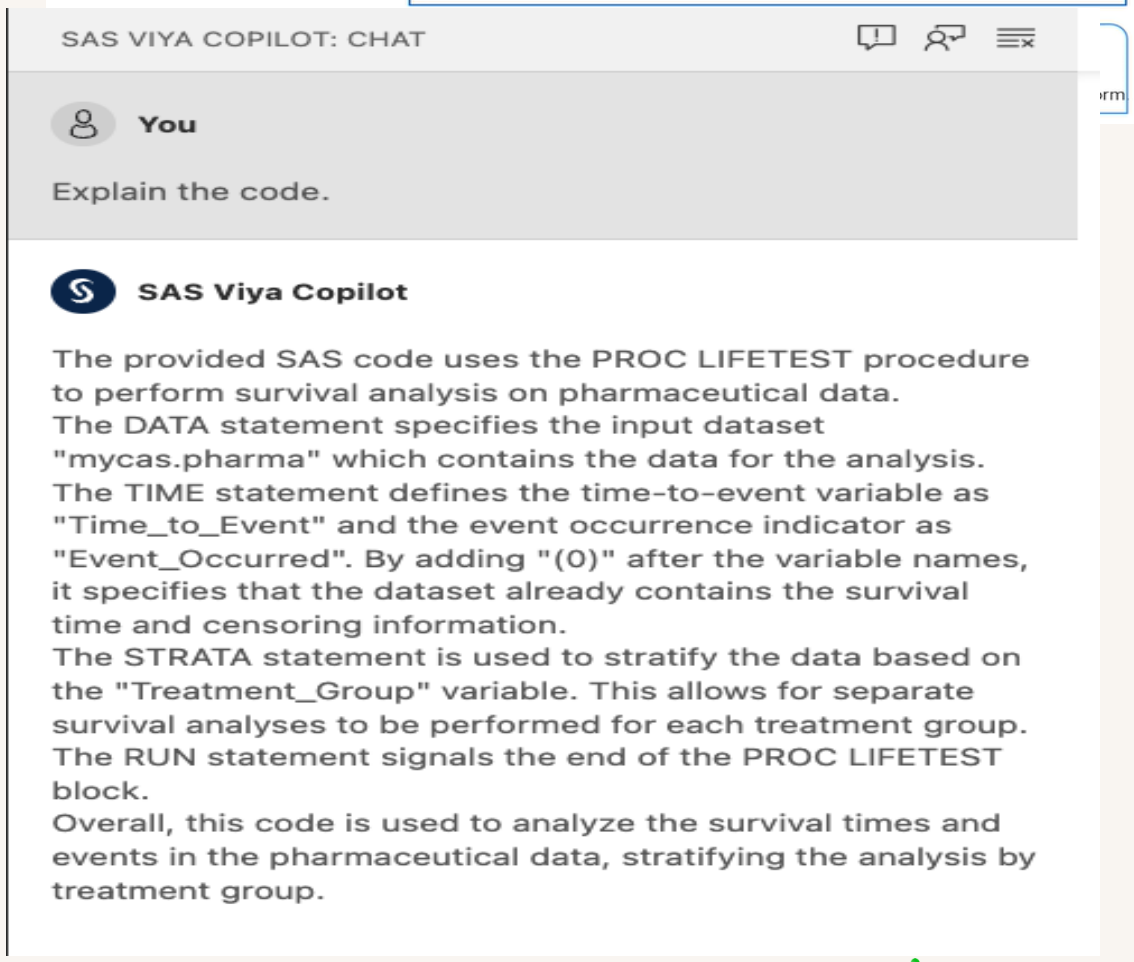
Generate code

```
/* Convert 'potassiumchar' to a numeric variable 'potassium' in kidneyTest */
data kidneyTest;
    set kidneyTest;
    potassium = input(potassiumchar, best12.);
    drop potassiumchar;
run;
```

Add comments | Explain

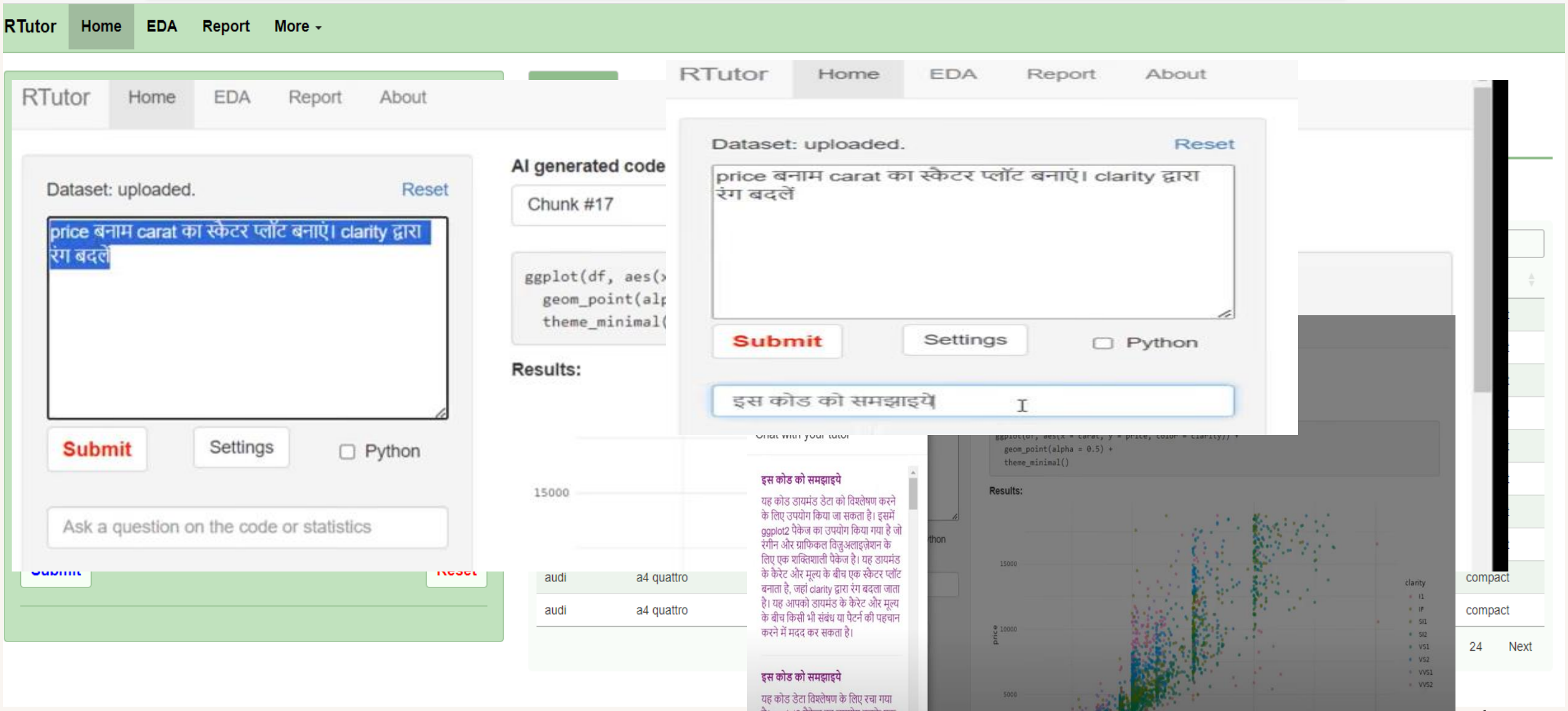






Fortrea

Co-Pilots, Labs, and **Tutors**



Co-Pilots, **Labs**, and Tutors

Generate | Write a code for merging 2 datasets with a innerjoin | Close

1 of 1 | Use code with caution

```
# prompt: Write a code for merging 2 datasets with a innerjoin

import pandas as pd

# Sample data for dataset 1
data1 = {'ID': [1, 2, 3, 4, 5],
        'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],
        'Age': [25, 30, 22, 28, 35]}
df1 = pd.DataFrame(data1)

# Sample data for dataset 2
data2 = {'ID': [2, 3, 5, 6, 7],
        'City': ['New York', 'London', 'Paris', 'Tokyo', 'Sydney'],
        'Country': ['USA', 'UK', 'France', 'Japan', 'Australia']}
df2 = pd.DataFrame(data2)

# Merge the two datasets using an inner join on the 'ID' column
merged_df = pd.merge(df1, df2, on='ID', how='inner')

# Print the merged dataset
merged_df
```

	ID	Name	Age	City	Country
0	2	Bob	30	New York	USA
1	3	Charlie	22	London	UK
2	5	Eve	35	Paris	France

Next steps: Generate code with merged_df | View recommended plots | New interactive sheet

CoLab offers Python code by default, and if you're interested, you can also integrate a SAS plugin to learn SAS code. There are resources available that explain how to add SAS. In the same session, you can ask questions related to both Python and SAS.

Power of Multiple AI Tools - AI Benefits in Clinical Statistical Programming

ClinicalTrials.gov

PubMed

These advancements can save time, reduce frustration, and enhance overall efficiency in clinical statistical programming. What do you think?

Real-Time Adverse Event Monitoring:

1. Utilize the Rave database to live stream data, detecting any **abnormalities in adverse events** as they happen.

Understanding Relationships in Datasets:

1. Map out **connections** between the Rave **database** and various dataset levels, as well as **CRF** (Case Report Form) forms, to gain a clearer understanding of data **relationships**.

Patient Journey Projections with GENAI:

1. Employ General AI (GENAI) to simulate and predict **patient behavior and outcomes** using sigmoid methods, enhancing patient journey insights.

Operational Efficiency with GENAI:

1. Utilize GENAI to reveal how each CRF form relates to others, and provide **keys** to connect each form, offering a **summary** of these connections.

Streamlined Utility Identification for Pharma/CROs:

Live Database Updates and Dashboards:

1. Implement a dashboard that updates the **data management** and **statistical programming** teams in real-time when the live database is updated, ensuring everyone is informed of the **latest changes** and key data points.

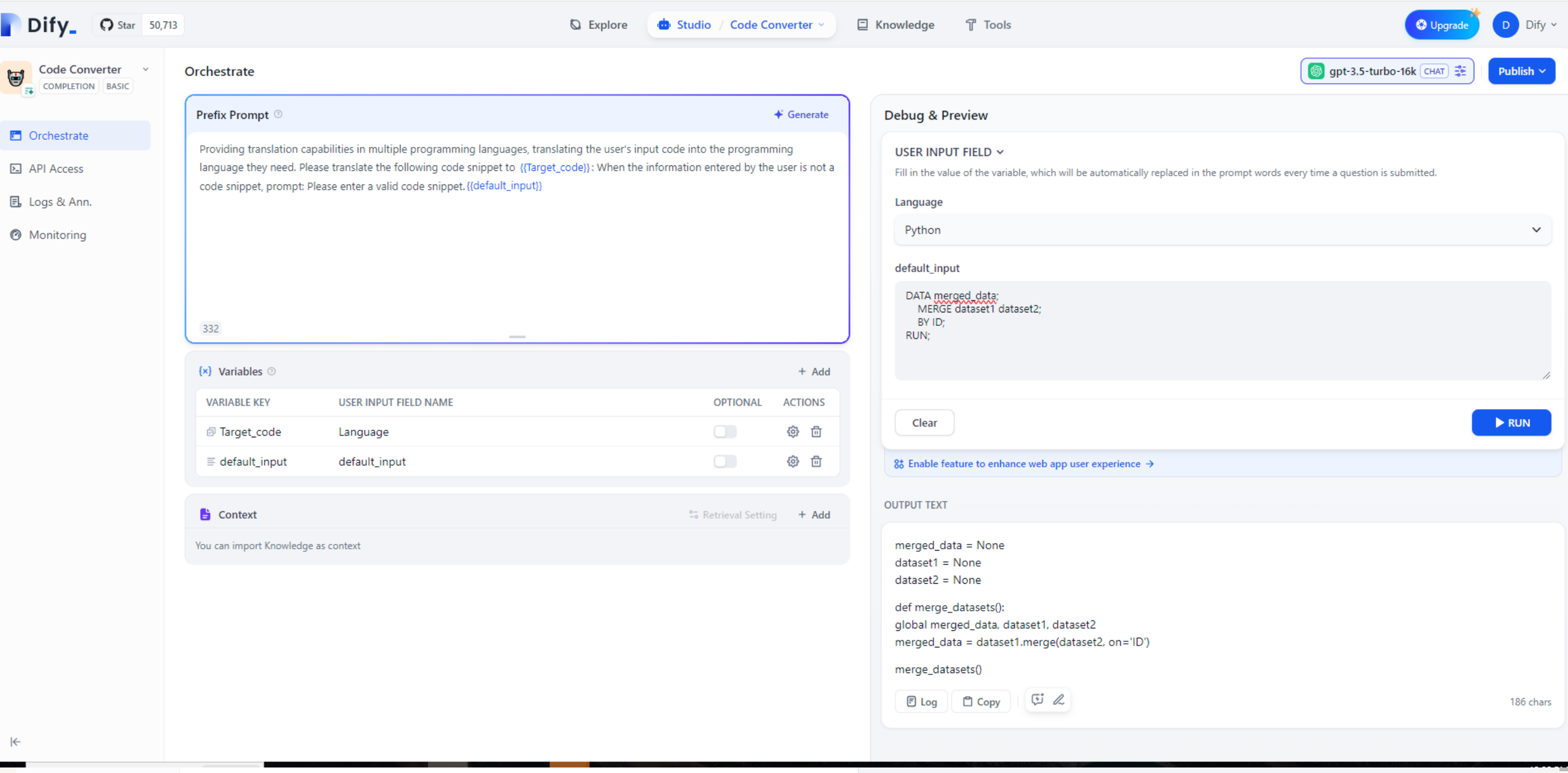
Internal LLMs (Large Language Models):

1. Contract Research Organizations (CROs) often juggle **multiple studies**, making it challenging to find **specific utilities**. Instead of relying solely on Subject Matter Experts (SMEs), create **internal LLMs to provide scenario-based recommendations**. The model can suggest appropriate macros and reference studies, significantly reducing time and effort.

Provide reasoning and Compares potentially **impacted SOPs** with **new regulations**?

Fortrea

LLMs Can I create?



Regulatory Authorities

FDA's drug center to consolidate AI efforts under single council

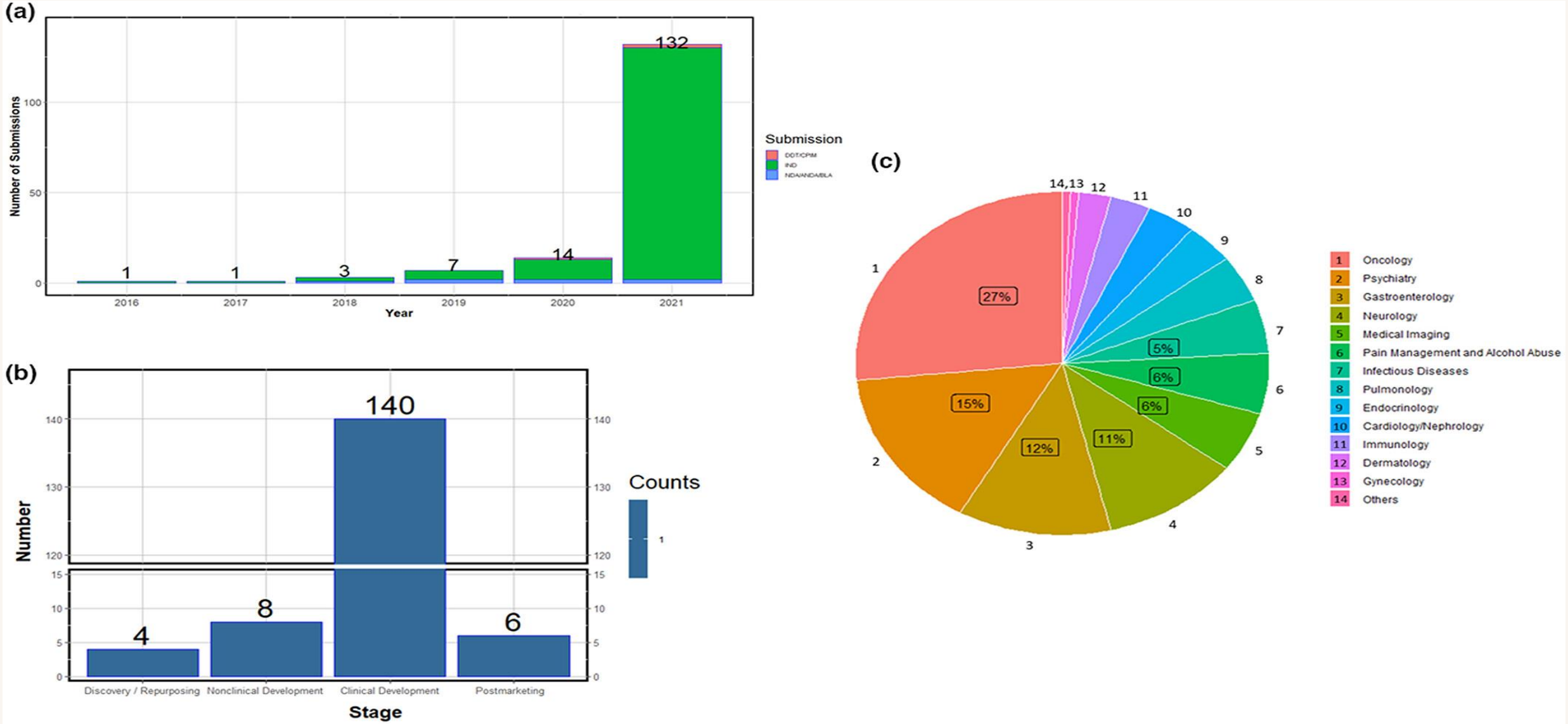
(fiercebiotech.com) - **FDA's drug center to consolidate AI efforts under single council**

"The advent of novel AI applications, such as generative AI and large language models, is likely to expand opportunities for use of AI within CDER, including by non-technical staff. This will require increased AI education and coordination," she said. "There has also been a rapid increase in CDER regulatory submissions incorporating AI, and the scope and impact of AI use in drug development are expanding."

FDA's CDER has received>500 submissions with AI between 2016 to 2021

Fortrea

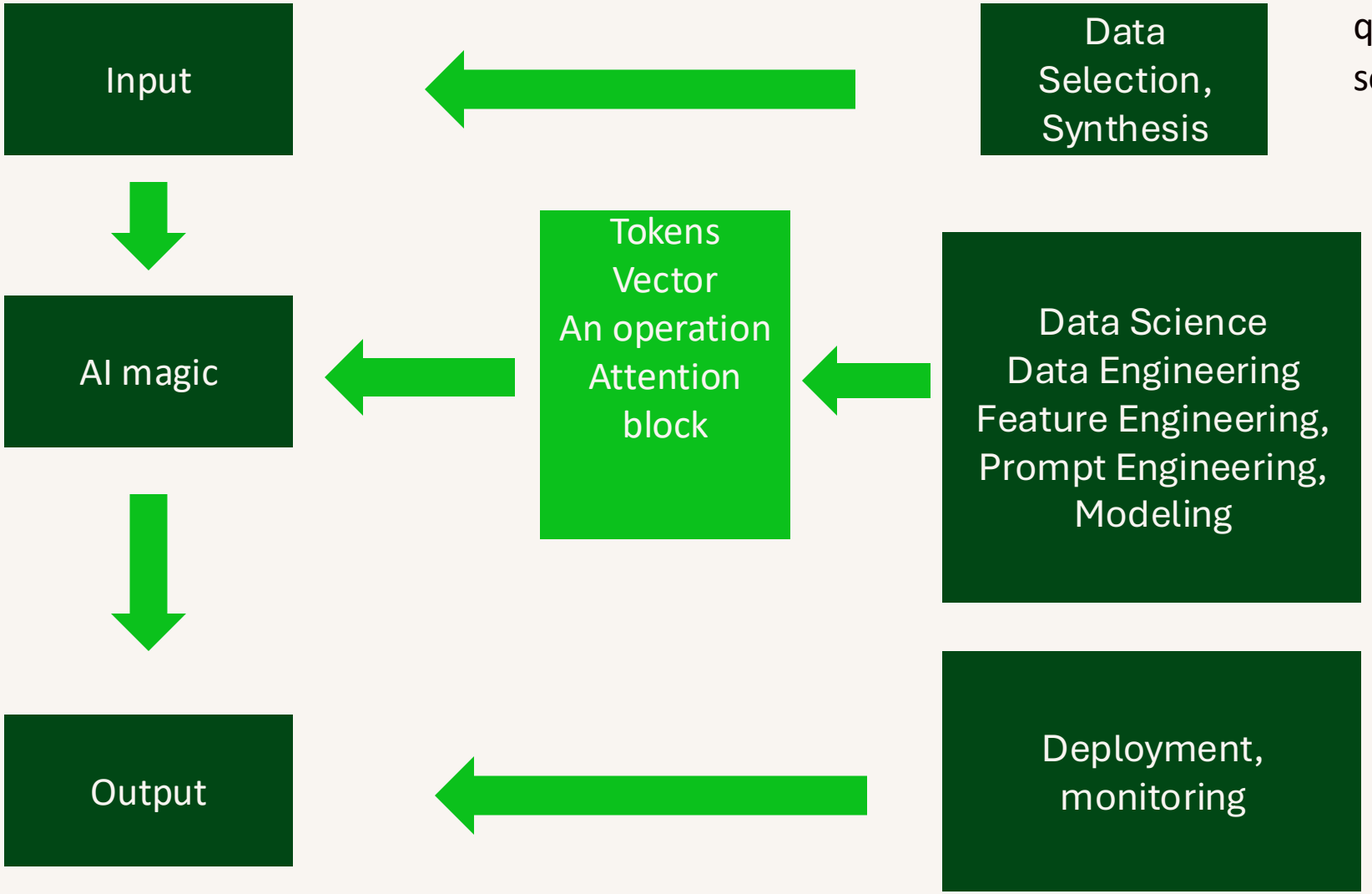
LANDSCAPE ANALYSIS OF THE APPLICATION OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN REGULATORY SUBMISSIONS FOR DRUG DEVELOPMENT FROM 2016 TO 2021



Clin Pharma and Therapeutics, Volume: 113, Issue: 4, Pages: 771-774, First published: 16 June 2022, DOI: (10.1002/cpt.2668)

Fortrea

AI is not a Magic



Data is not true always- it will be true when you ask a question then you find something meaningful

Fortrea

Agenda

How to convince the programmer by providing a roadmap

Foundational Understanding - AI vs GENAI with all other sources

Introductory Use Cases - Simple examples of Chatgpt, geminin, notbook LLM and some other LLMS

Hands-On Practice - Co-Pilots, Labs, and Tutors – Github, Microsoft, SAS and Google co-lab, R tutor

Experimentation - Creators your own LLM – DEFY

Regulatory Considerations

All of these applications require minimal programming effort but do rely on domain expertise, which we already possess as experienced programmers in the industry. So, why not leverage these tools?

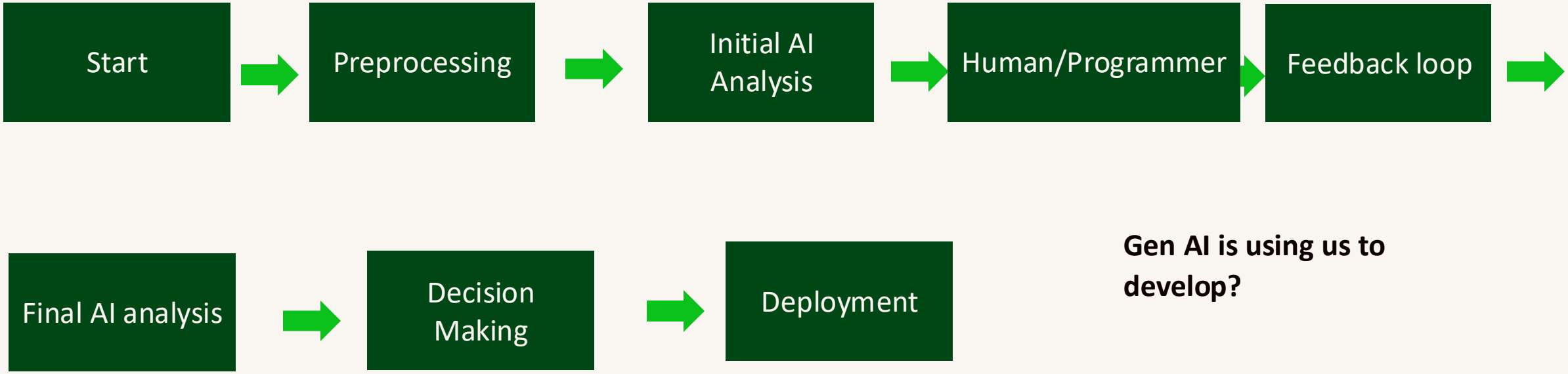
Naturally, there are considerations when choosing between enterprise and public GPTs, such as data bias, data privacy, ethical use, and long-term ethical implications.

AI is everything ? NO – because it is Probabilistic system -> Support your day-to-day tasks

ChatGPT is great -> Intern - > better training or instructions -> it will produce a crap

Fortrea

What is the conclusion?



Gen AI is using us to develop?

Human>Programmer-in-the-Loop Benefits:

- **Improved Accuracy**: Ensures AI outputs are accurate and contextually relevant.
- **Ethical Oversight**: Provides a human perspective to address ethical concerns.
- **Continuous Learning**: Enhances AI learning with human feedback, making models more robust.

Generative AI is not ready to operate without human oversight.

As programmers in the industry, our focus should shift from individual tools to the broader ideas and innovations they enable.

Fortrea

Thank you!

Fortrea

References

- GitHub Co-pilot: https://docs.posit.co/ide/user/ide/guide/tools/copilot.html
- Chatgpt in R : https://www.r-bloggers.com/2024/05/how-to-get-chatgpt-in-r-with-chattr/
- SAS plugin Google Co-lab: https://communities.sas.com/t5/SAS-Programming/SAS-google-colab/td-p/759241
- Dify : https://cloud.dify.ai/app/4d60d622-ad72-4cb2-9d58-d1fb9f438b1a/configuration
- ChatGPT: https://openai.com/index/chatgpt/
- Gemini : https://gemini.google.com/?utm_source=google&utm_medium=cpc&utm_campaign=2024enUS_gemfeb&gad_source=1&gclid=Cj0KCQiA88a5BhDPARIsAFj595hqA82lPfYb9fkgBhcabgrXQz0WYW0wHZcNuLWR4QUSiSIi04HhxaoaAjDYEALw_wcB&gclsrc=aw.ds
- Notebook LLM : https://notebooklm.google/
- Microsoft Co-Pilot: https://copilot.microsoft.com/
- Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021: https://ascpt.onlinelibrary.wiley.com/doi/full/10.1002/cpt.2668
- Rtutor: https://rtutor.ai/
- Phi3.5 https://ollama.com/library/phi3.5

Fortrea