Paper CM06

Unleashing the Power of Large Language Models for Clinical Data Approaches: Will AI chatbots like ChatGPT/Bard replace statistical programmer and statisticians over time?

Ercan Suekuer, Roche, Basel, Switzerland

ABSTRACT

The use of natural language processing (NLP) techniques has been gaining significant momentum in clinical research and healthcare delivery. With the advent of large language models such as ChatGPT, there is now an unprecedented opportunity to leverage these models for analyzing and extracting insights from clinical data.ChatGPT is a state-of-the-art language model that has been trained on a massive corpus of text data and has demonstrated remarkable capabilities in generating human-like responses to natural language queries. In the context of clinical data, ChatGPT can be used to extract valuable insights from electronic health records, patient surveys, clinical notes, and other sources of unstructured data. In this presentation, we will discuss the benefits and opportunities of using ChatGPT for clinical data approaches. We will explore how ChatGPT can be used to extract structured data from unstructured sources, identify patterns and trends in patient data, and develop predictive models for clinical outcomes. But also how it can help us in our daily work and automate workflows and programming. We will also discuss the challenges and limitations of using ChatGPT in clinical research, such as ensuring privacy and avoiding bias in machine learning models.Finally, we will highlight some real-world applications of ChatGPT in clinical research and healthcare delivery, clinical decision support, and drug discovery. We will conclude with a discussion of future research directions in this field and the potential impact of ChatGPT on the healthcare industry.

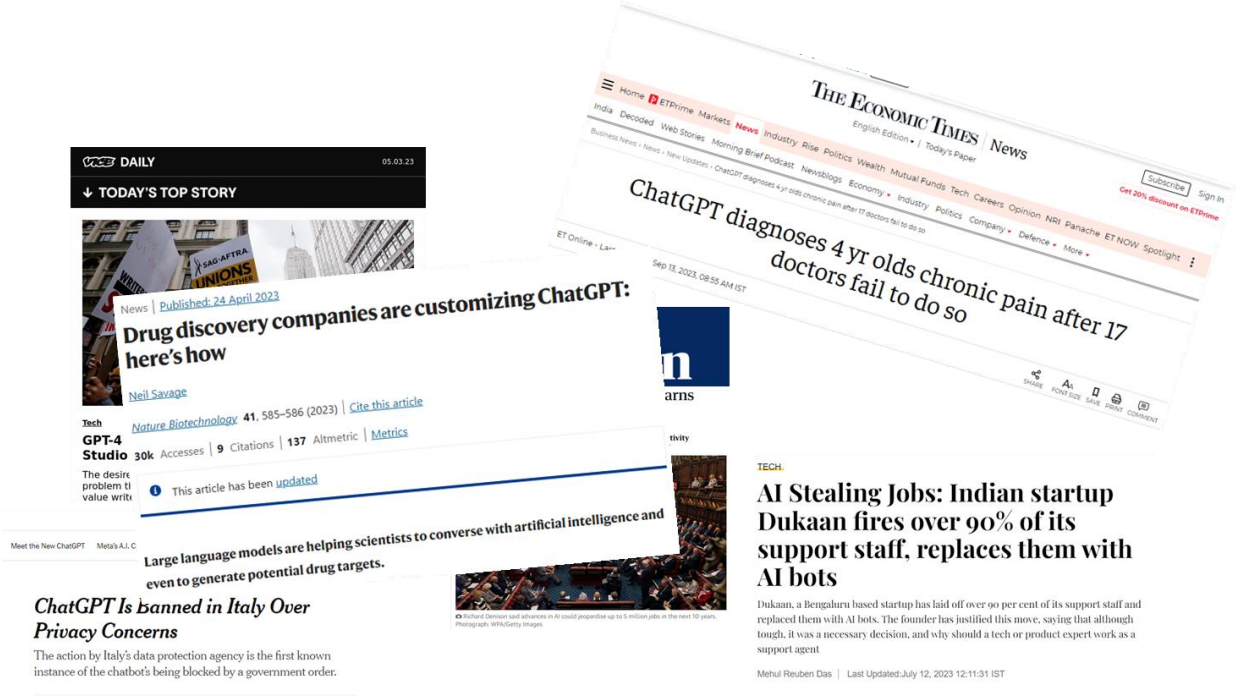
INTRODUCTION

The rapid advancement in natural language processing (NLP) techniques has transformed various domains, including clinical research and healthcare delivery. With the emergence of large language models like ChatGPT, we now have the potential to harness these models for extracting valuable insights from clinical data. ChatGPT, a state-of-the-art language model, offers remarkable capabilities in generating human-like responses to natural language queries. This presentation will explore the benefits, opportunities, challenges, and real-world applications of ChatGPT in clinical data analysis and healthcare delivery. We will discuss how ChatGPT can be utilized to extract structured data from unstructured sources, identify patterns in patient data, develop predictive models for clinical outcomes, automate workflows, and support clinical decision-making, while addressing the associated challenges and limitations.

THE BUZZ AROUND CHATGPT AND AI BOTS

The emergence of large language models (LLMs) and AI bots, such as ChatGPT, has fueled debates about their impact on humans and the future of work. Some argue that they may replace human workers, while others highlight

the potential benefits for developing better medications and exploring new avenues in drug discovery.



WHAT ARE LLMS AND AI BOTS?

Large Language Models (LLMs) are neural networks trained on vast text datasets to build strong statistical models of natural language. They can predict word sequences and generate coherent text output based on probability theory. AI bots, like ChatGPT, are fine-tuned using reinforcement learning and human feedback to generate human-like responses to natural language queries.

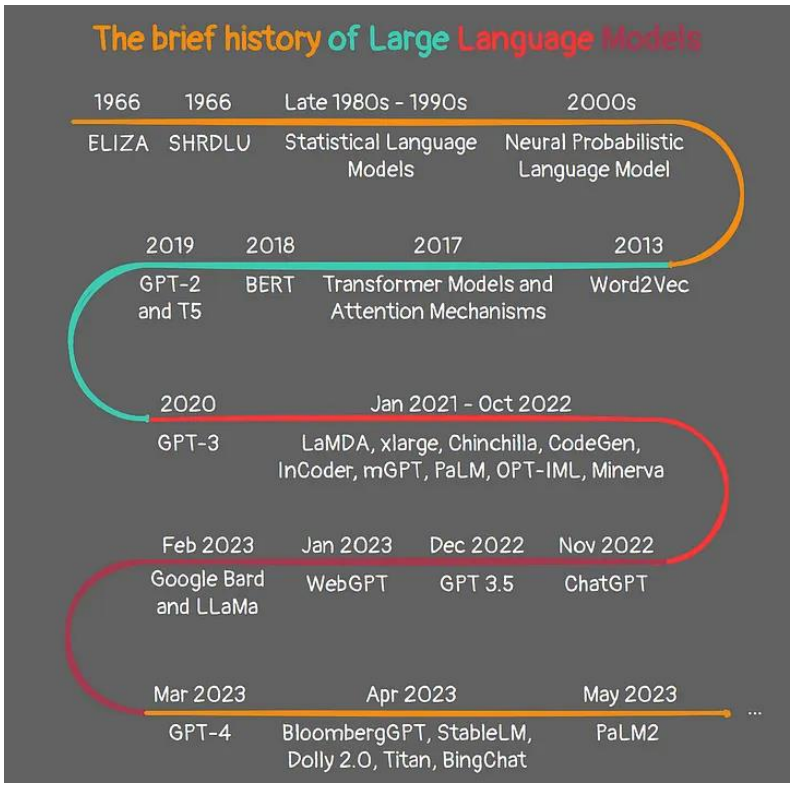
The best thing about AI is its ability to

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

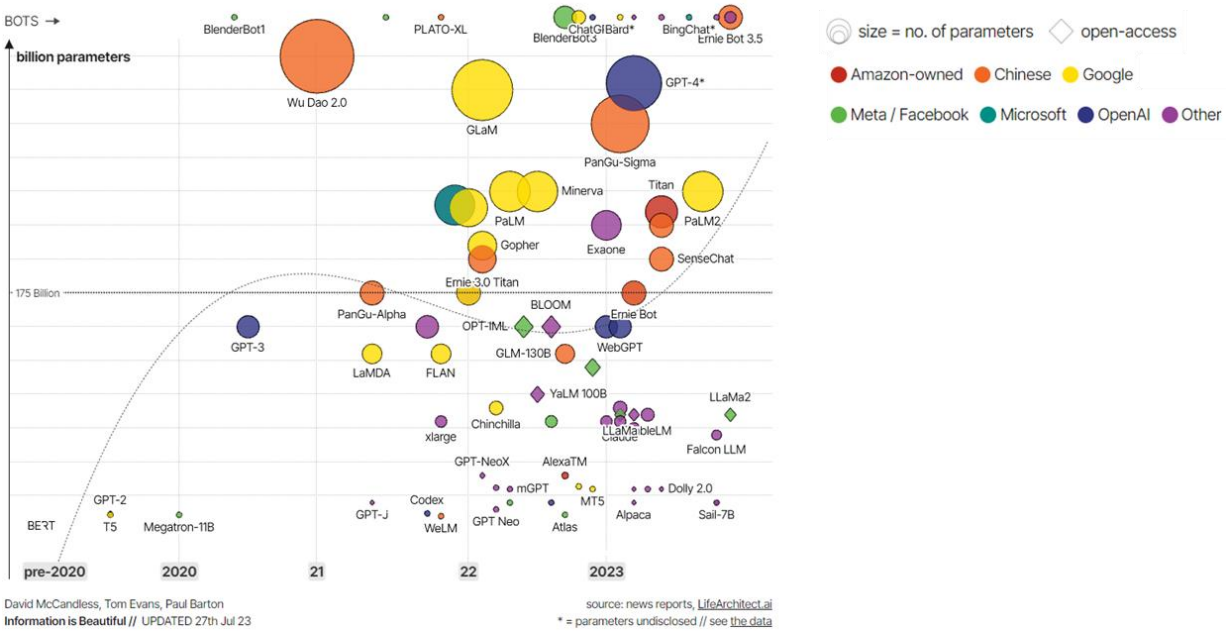
```
Out[ ]= {The best thing about AI is its ability to,
         The best thing about AI is its ability to learn,
         The best thing about AI is its ability to learn from,
         The best thing about AI is its ability to learn from experience,
         The best thing about AI is its ability to learn from experience.,
         The best thing about AI is its ability to learn from experience. It,
         The best thing about AI is its ability to learn from experience. It's,
         The best thing about AI is its ability to learn from experience. It's not}
```

A BRIEF HISTORY OF LANGUAGE MODELS

To gain a better understanding of the development of large language models like ChatGPT, it is essential to take a look at their historical context. Although ChatGPT is indeed a revolutionary advancement, the roots of AI language models can be traced back to the early days of AI. The Eliza language model, one of the earliest examples of an AI language model, was introduced at MIT in 1966. Language models are initially trained on a dataset and employ various techniques to derive relationships and generate new content based on the trained data. They are often utilized in natural language processing (NLP) applications, where users input requests in natural language to obtain results.



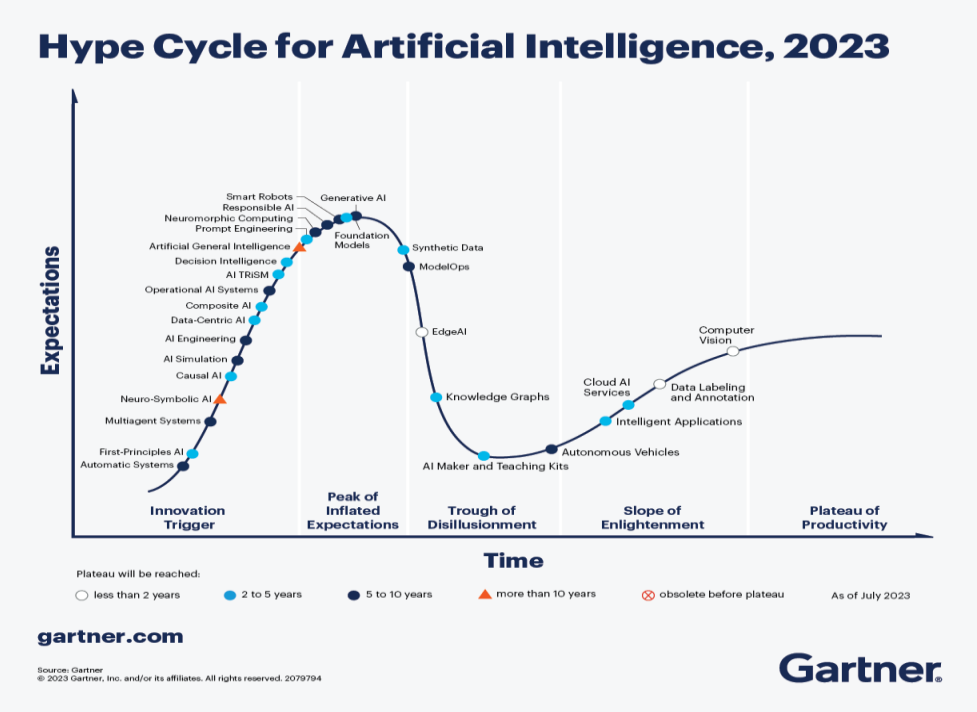
Large language models signify the evolution of the language model concept in AI, as they dramatically expand the data used for training and inference, thereby significantly enhancing the AI model's capabilities. Although there is no universally accepted threshold for the size of the training dataset, an LLM typically comprises at least one billion or more parameters. Parameters are variables present in the model on which it was trained and can be used to derive new content.



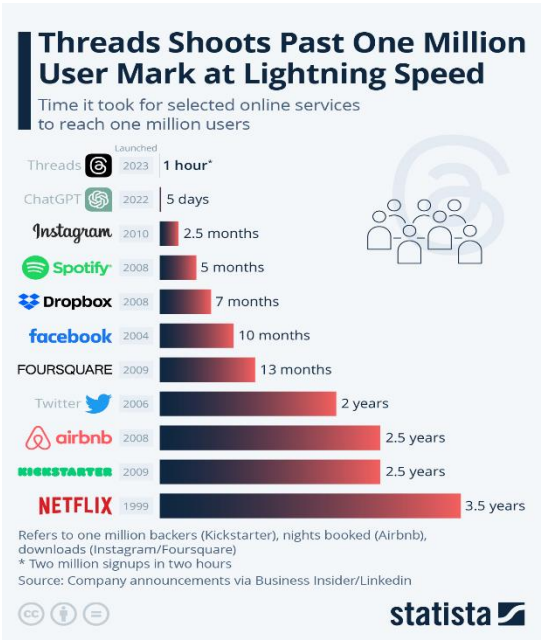
The field of artificial intelligence has experienced an immense boom in recent years, with new LLMs emerging almost every month. The exponential growth in the past two years is evident when examining the various LLMs, their respective owners, parameter sizes, and release dates. For example, while the 2018 models had 340 million parameters, today's models boast over 1,000 billion parameters. These enormous models require significant time and resources for training, illustrating the rapid advancements in AI and the potential for continued growth in the future.

IS IT JUST A HYPE?

The rapid growth of large language models, such as ChatGPT, raises questions about the sustainability and future trajectory of this technology. As depicted in Gartner's Hype Cycle for AI 2023, we are currently at the peak of the hype wave, and it remains a topic of interest to see how this technology will evolve and shape the future landscape.



The hype surrounding large language models like ChatGPT is not unfounded. One compelling evidence supporting this hype is the model's ability to reach its first million users. While attaining a milestone of one million users is often a challenging task for any technology, ChatGPT remarkably achieved this in just five days. This extraordinary accomplishment underscores the groundbreaking potential of ChatGPT and the significant impact it can have in a relatively short period.



BENEFITS AND OPPORTUNITIES

Large language models (LLMs) offer various benefits that can significantly impact the field of drug development and clinical research. These advantages include:

- Expertise and efficiency: LLMs possess a deep understanding of drug development literature, enabling them to provide fast and accurate answers to complex questions. This increased efficiency can aid researchers in making informed decisions more quickly.

- Streamlined research: LLMs can rapidly identify relevant data and summarize intricate literature, which saves time and effort for scientists. This streamlined process allows researchers to focus on more critical aspects of their work.
- Accelerated innovation: LLMs can automate tasks and expedite coding and programming, resulting in accelerated research. Consequently, scientists can dedicate more time to discovery and development, leading to a faster pace of innovation in the field.
- Enhanced collaboration: LLMs have the ability to simplify complex concepts, which fosters better communication and collaboration among researchers. Improved collaboration can lead to more efficient problem-solving and more effective teamwork.
- Cost-effective and adaptable: LLMs offer cost savings through automation, as they can perform tasks that would otherwise require human effort. Additionally, they continuously adapt to new information and research trends, ensuring that they remain relevant and up-to-date in the ever-evolving field of drug development and clinical research.

CHALLENGES AND LIMITATIONS

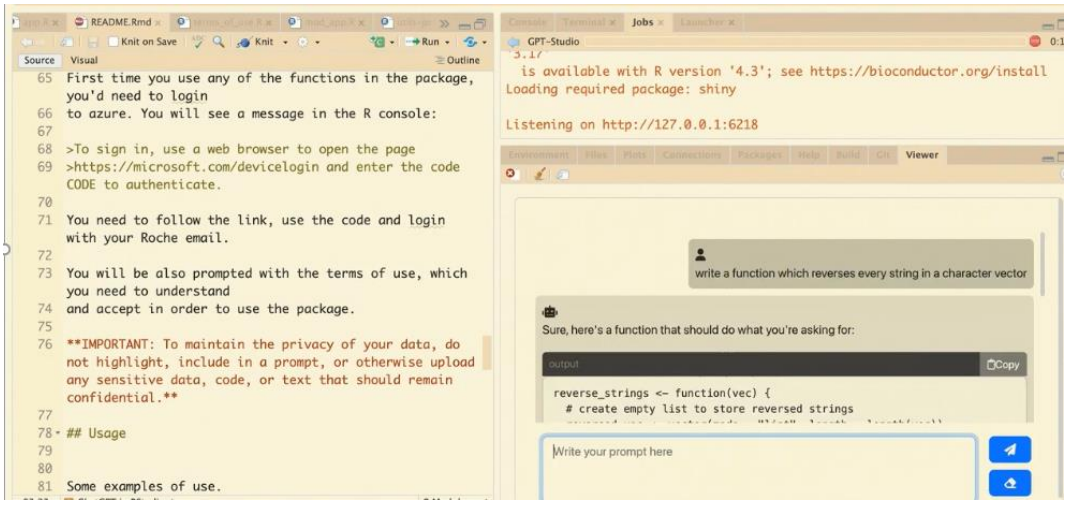
Despite the numerous benefits of large language models (LLMs), they also exhibit certain limitations that can impact their effectiveness and accuracy. These challenges include:

- Limited understanding: LLMs, utilizing Transformer Architecture, do not fully comprehend words like humans do. Instead, they rely on probability to generate text, which can result in inaccuracies and less nuanced understanding.
- Hallucinations: LLMs may produce incorrect or off-topic text due to decoding errors, gaps in training data, ambiguous context, and inherent biases. This can affect the quality and reliability of the information generated by the model.
- Prompt sensitivity: Hallucinations can be triggered by specific content in user prompts, making topic-focused conversations more susceptible to inaccuracies compared to fact-finding Q&A sessions.
- Context window limitations: When the token context window size is exceeded, LLMs may generate less coherent and accurate responses. This limitation can impact the effectiveness of LLMs in generating meaningful and contextually relevant information.
- Inherent biases: LLMs can inherit biases from their training data, which can potentially lead to the generation of biased or offensive content. This highlights the importance of carefully curating and refining training data to minimize the risk of perpetuating harmful biases.

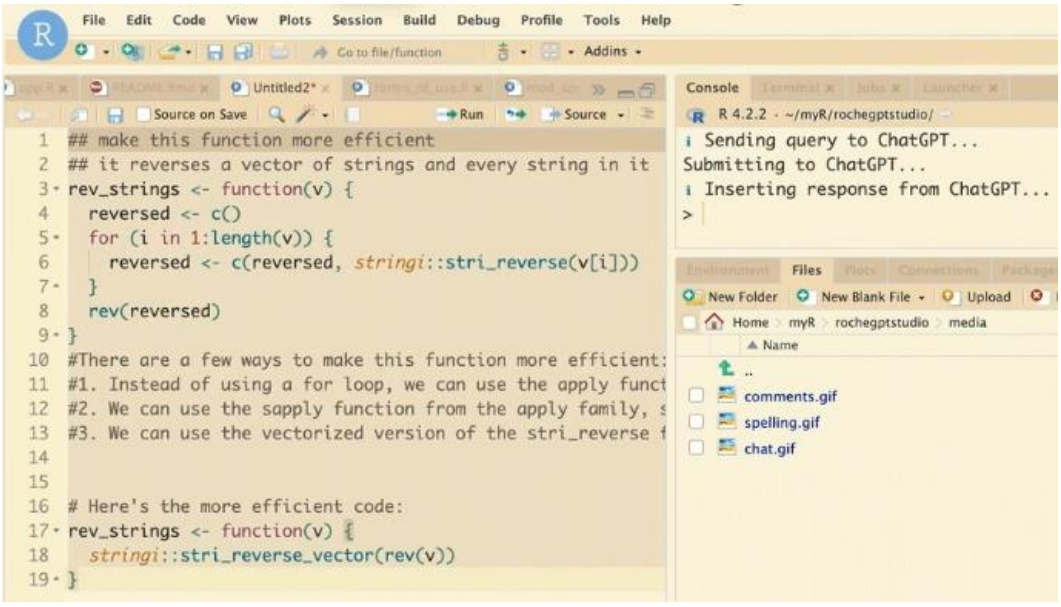
REAL-WORLD APPLICATIONS

Roche GPTStudio & Roche pRED-GPT

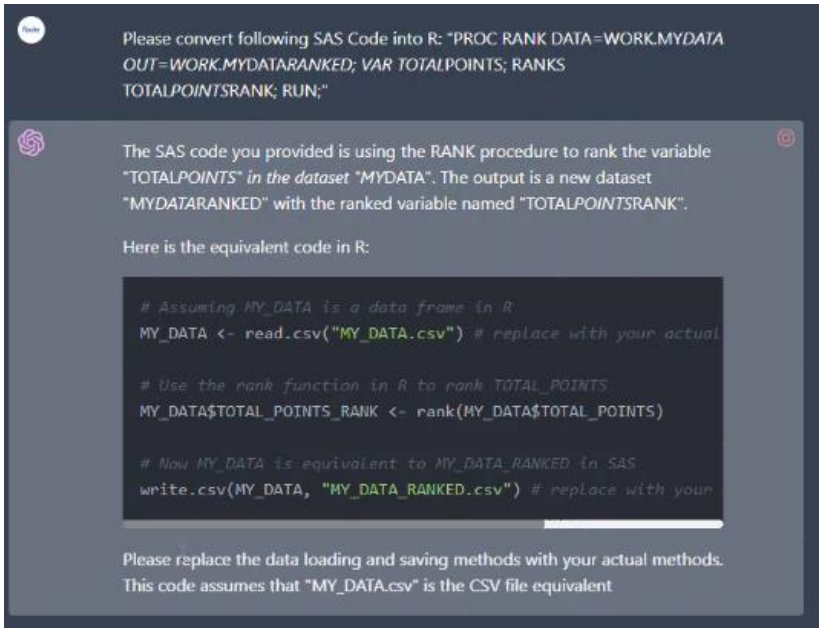
At Roche, our team of data scientists handles clinical data and employs various programming languages, including a significant number of R programmers. As every programmer is familiar with the frustration of encountering non-functioning code or reviewing someone else's code, we have integrated ChatGPT into our RStudio IDE to streamline the process. This integration allows us to access the GPT-4 model for assistance with daily programming tasks.



In addition to generating functions or writing code for us, ChatGPT can also be utilized to optimize existing code, thereby enhancing its efficiency and performance.



Furthermore, ChatGPT proves helpful in translating between programming languages. For instance, it can facilitate code conversion from SAS to R or Python, making it a versatile tool for our team of data scientists. This increased versatility allows for improved collaboration and productivity among team members with diverse programming backgrounds.



Another task that we have delegated to ChatGPT is the creation of first drafts for presentations. We can request the AI assistant to generate a presentation consisting of 10 slides on a specified topic. Moreover, we ask the bot to create the presentation using a language compatible with Google Slides, such as Apps Script. This capability enables us to review the initial drafts generated by ChatGPT, which can then be further refined and customized according to our specific needs. By automating the initial stages of the presentation creation process, we can save time and effort, allowing our team to focus on more critical aspects of the project.

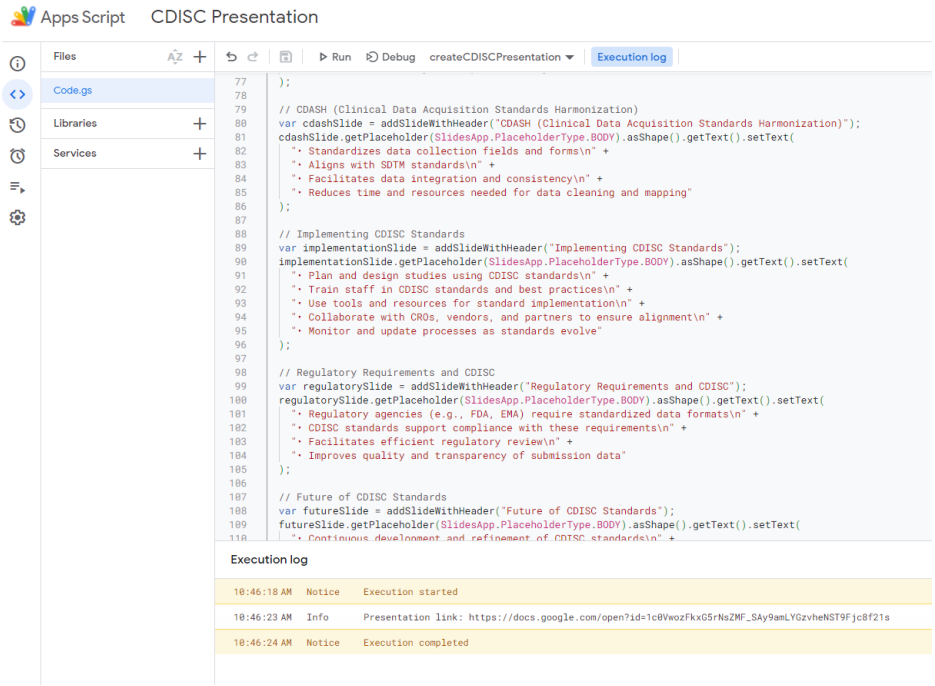
Write me "Apps Script" google slides code on CDISC standard. Make it 10 slides. expand the ideas in bullet points and in more detail.

I apologize for the confusion. Here's a Google Apps Script code to create a 10-slide presentation on CDISC standards. I've expanded the ideas in bullet points and added more detail as requested.

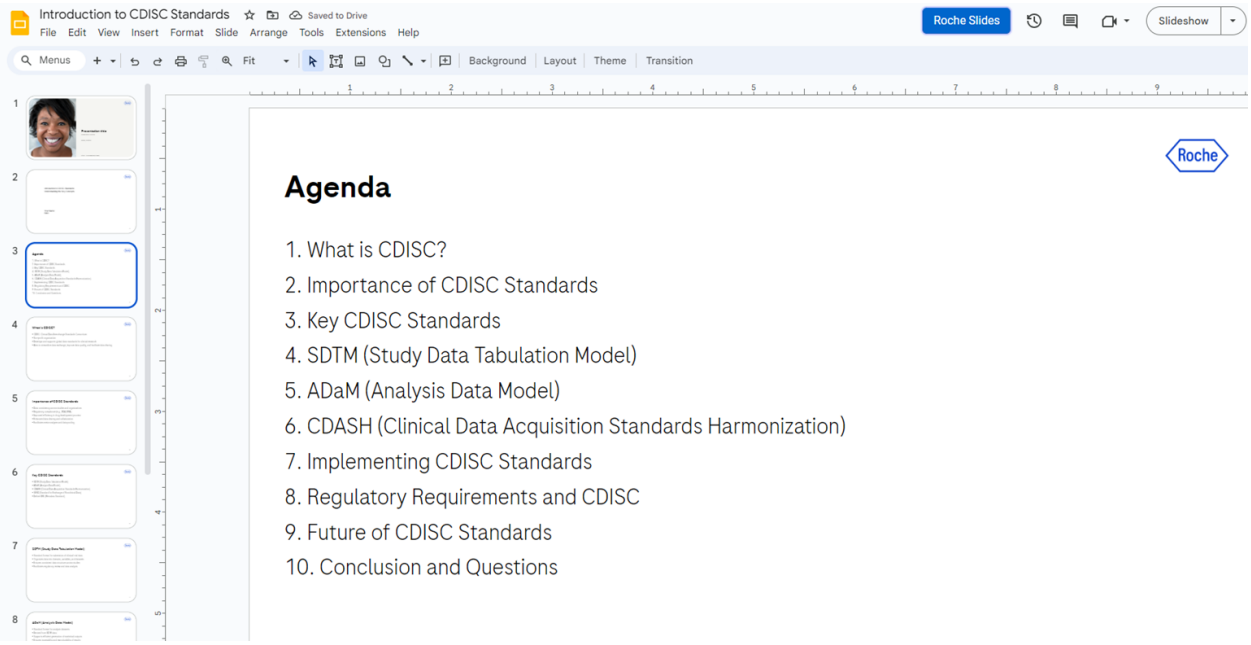
```
function createCDISCPresentation() {
  var presentation = SlidesApp.create('Introduction to CDISC Standards');
  var slides = presentation.getSlides();

  // Title Slide
  var titleSlide = slides[0];
  titleSlide.insertTextBox("Introduction to CDISC Standards\nUnderstanding the Key Concepts", 100, 100, 600, 100);
  titleSlide.insertTextBox("Your Name\nDate", 100, 250, 600, 100);

  // Function to add slide with title
  function addSlideWithHeader(title) {
    var slide = presentation.appendSlide(SlidesApp.PredefinedLayout.TITLE_AND_BODY);
    slide.getPlaceholder(SlidesApp.PlaceholderType.TITLE).asShape().getText().setText(title);
    return slide;
  }

  // Agenda
  var agendaSlide = addSlideWithHeader("Agenda");
  agendaSlide.getBodyPlaceholder().asShape().getText().setText(
    "1. What is CDISC?\n" +
    "2. Importance of CDISC Standards\n" +
    "3. Key CDISC Standards\n" +
    "4. SDTM (Study Data Tabulation Model)\n" +
    "5. ADaM (Analysis Data Model)\n" +
```

Type user query here. (Shift + Enter for new line)

AI Bot generated code in Apps Script:

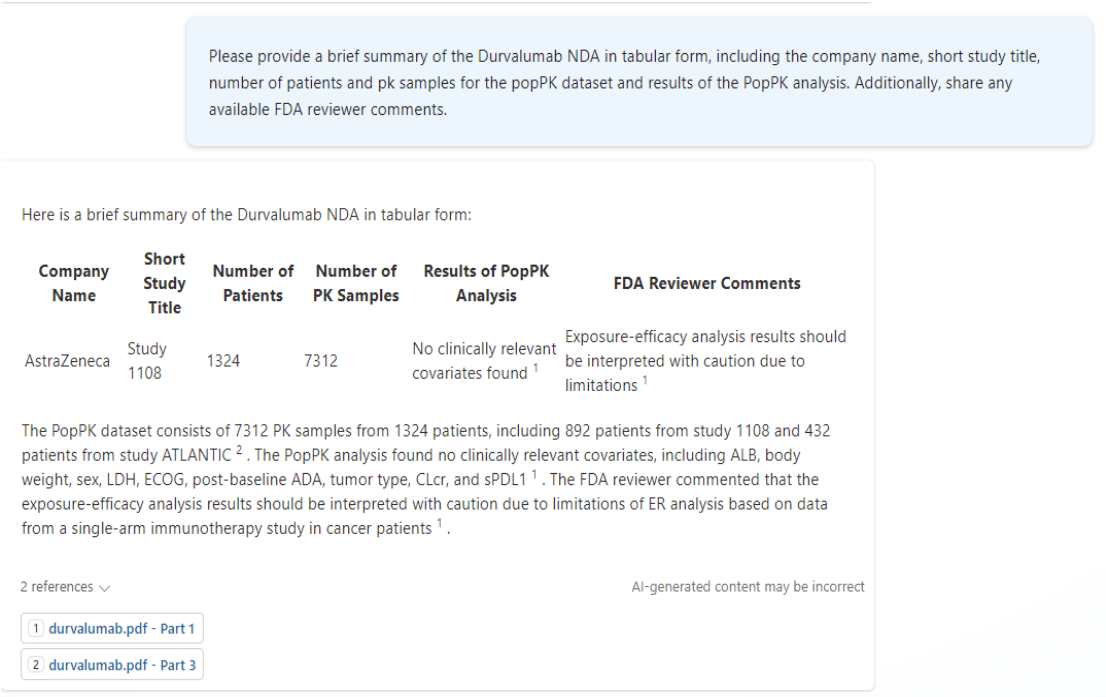


AI Bot generated Slide:



Future Research Directions & Potential Impact

We have begun optimizing our GPT model to ensure readiness for regulatory submissions. By collecting FDA's NDA/BLA Review and Evaluation Documents, we have trained our model to think like an FDA auditor. With over a hundred approved NDA/BLA files, we can identify common pain points and improve our submissions by having them pre-reviewed by an AI FDA auditor.



Our ultimate goal is to fine-tune the large language model, transforming it into an authoritative predictor that can enhance submission documents by anticipating comments and driving improvements. This optimization process will enable us to address potential concerns proactively and improve the overall quality of our regulatory submissions.

CONCLUSION

In conclusion, the integration and optimization of the GPT model have significantly improved the efficiency of various tasks within our team of data scientists and regulatory professionals. By training the model using FDA's NDA/BLA Review and Evaluation Documents, we have created an AI assistant capable of emulating the thought process of an FDA auditor. This AI-powered auditor can identify common pain points and help enhance the quality of our submissions by providing pre-reviews. The GPT model is also instrumental in optimizing existing code, translating between programming languages, and generating initial drafts for presentations.

By fine-tuning the large language model, we aim to create an authoritative predictor that can proactively address potential concerns and drive improvements in the submission documents. The incorporation of this AI technology streamlines our processes, improves collaboration among team members with diverse programming backgrounds, and ultimately leads to more successful and efficient regulatory submissions.

REFERENCES

https://informationisbeautiful.net/visualizations/the-rise-of-generative-ai-large-language-models-llms-like-chatgpt/
https://emtemp.gcom.cloud/ngw/globalassets/en/articles/images/hype-cycle-for-artificial-intelligence-2023.png
https://www.statista.com/chart/29174/time-to-one-million-users/
The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision) https://arxiv.org/abs/2309.17421

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)
Your comments and questions are valued and encouraged.
Contact the author at:

Ercan Suekuer
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland
Email: ercan.suekuer@roche.com