Datathon 2024

Our theme this year, "GenAI Playbook," invites you to dive deep into various generative AI systems, including chatbots and multimodal AI, to uncover their advantages, disadvantages, best use cases, and potential risks.

- **Explore GenAI Systems:** Test various generative AI platforms to understand their functionalities.
- **Analyze and Document:** Evaluate the pros and cons, identify the best use cases, and document any biases or potential risks.
- **Create a Playbook:** Compile your findings into a comprehensive playbook with best practices and recommendations.
- **Share Insights:** Present your playbooks through video-recorded presentations, showcasing key insights and practical examples.
- **GenAI Platforms:** OpenAI's ChatGPT, Google's Bard, and more.
- **Bias Detection Tools:** IBM's AI Fairness 360, Google's What-If Tool, etc.

Women in Data (mn.co)

Scope

Explore 3 Gen AI chatbot platforms.
Evaluate the pros and cons.

Look for bias or misinformation in the output.

Problem statement : Maternal Health

Why maternal health ? - include in the slide

Maternal And Infant Health Disparities Data Brief | AHR (americashealthrankings.org)

New America's Health Rankings data brief highlights maternal and infant health disparities - UnitedHealth Group

Approach

- Pick 3 Gen AI platforms /AI chatbots
    - ChatGPT
    - Gemini

    - Claude
- Step 1: Ask same set of 5 questions to the chatbots and gather the response
- Step 2: RAG - Upload 2024 maternal health report and then ask same set of 5 questions to those AI chatbots
- Step 3: Ask the model , to generate synthetic maternal health data
- Step 4 : Analyze the responses and data to verify accuracy,bias

Tools

AI chat bots

1. ChatGPT 4o - open AI
2. Claude 3.5 sonnet- Anthropic
3. Gemini 1.5 flash - Google

You.com - free tier limitation - only 3 prompts supported per day.
NotebookLM.com used Gemini 1.5 pro model

chatgpt.com

Make sure to signup so that you can see upload feature

claude.ai

Signup

gemini.google.com

Upload file in google drive, enable drive extension . Refer the file from drive in your prompt.@Google Drive Can you summarize ahr_2024_databrief.pdf

@Google Drive based on this report ahr_2024databrief-stateprofiles.pdf , which state has highest maternal mortality rate

Feature	ChatGPT 4o	Gemini Flash 1.5	Claude 3.5 Sonnet	Gemini 1.5 Pro-notebook
Developer	OpenAI	Google	Anthropic	Google
Focus	Versatility, reasoning	Speed, real-time interactions	Strong reasoning, coding	General AI capabilities
Strengths	Image generation, internet access	Quick responses, real-time chat	Large context window, coding	Multimodal reasoning, general tasks
Use Cases	Various applications, including coding and math problems	Chat applications, quick interactions	Complex problem-solving, coding	General AI tasks, multimodal applications

Models size comparison

Here's a comparison of the model sizes and parameters for ChatGPT 4o, Gemini Flash 1.5, and Claude 3.5 Sonnet:

Model	Model Size	Parameters
ChatGPT 4o	Large	~1.8 trillion parameters 1
Gemini Flash 1.5	Small	8 billion parameters 2
Claude 3.5 Sonnet	Medium	~175 billion parameters 3

ChatGPT 4o is the largest among the three, with approximately 1.8 trillion parameters, making it highly capable for complex tasks . Gemini Flash 1.5 is the smallest, optimized for speed and efficiency with 8 billion parameters . Claude 3.5 Sonnet falls in between, with 175 billion parameters, offering a balance of speed and intelligence

Experiment 1: Ask the same set of 5 questions to the chatbots and gather the response. Do not upload source files

- How does the national rate of maternal mortality compare to the Healthy People 2030 target?
    - national rate is 22.4 while healthy People 2030 target is 15.7
- Based on 2024 America's Health Rankings® Maternal and Infant Health Disparities Data Brief , What are the social and economic factors contributing to maternal mortality?
    - Poverty
    - Residential Mobility

- Based on 2024 America's Health Rankings® Maternal and Infant Health Disparities Data Brief , In which state is the disparity in uninsured rates by education level the largest?
    - Arkansas
- Based on 2024 America's Health Rankings® Maternal and Infant Health Disparities Data Brief, Which state has the highest maternal mortality rate ? Alabama
- Based on 2024 America's Health Rankings® Maternal and Infant Health Disparities Data Brief, Which race/ethnicity has the highest mortality rate?
    - American Indian/Alaska Native

Experiment 2: Ask the same set of 5 questions to the chatbots and gather the response. Upload source files

Upload ahr_2024_databrief-final-web.pdf (americashealthrankings.org) and first 10 pages of ahr_2024databrief-stateprofiles.pdf (americashealthrankings.org)

If the PDF contains a lot of images or complex formatting, my understanding might be less accurate. Nature of content matters.They do well for natural conversations like summarize a meeting transcript

Experiment 3 : Ask the 3 chatbots to generate synthetic data

generate 10 synthetic records for following schema -first name,last name, age,gender,length of pregnancy,education,ethnicity

Conclusion

Comparing Model Results across Experiment 1 & 2

The sources describe two experiments comparing the performance of three large language models – Gemini, ChatGPT, and Claude – in answering questions related to maternal mortality rates in the United States. The models were tasked with answering the same set of questions in two different scenarios: without access to a specific report (Experiment 1) and with access to the report (Experiment 2).

Here's a comparison of their performance:

Experiment 1 (Without Access to 2024 Maternal Health Report)

1. **How does the national rate of maternal mortality compare to the Healthy People 2030 target?**
   Ground Truth: national rate is 22.4 while healthy People 2030 target is 15.7
   Gemini & ChatGPT: Provided inaccurate information and results that could not be trusted.
   Claude: Provided a more truthful response, acknowledging it doesn't have access to report.
2. **What are the social and economic factors contributing to maternal mortality?**
   Ground Truth: Poverty, residential mobility,
   Gemini & ChatGPT: Provided a limited list of factors and were uncertain if they had access to the report.
   Claude: acknowledging it doesn't have access to report.

3. **In which state is the disparity in uninsured rates by race/ethnicity the largest?**
   Ground Truth: Arkansas.
   All Models: Could not definitively answer the question without access to specific data. [5]
4. **Which state has the highest maternal mortality rate?**
   Ground Truth: Alabama.
   Gemini & ChatGPT: Provided inaccurate information; Gemini stated Tennessee while ChatGPT claimed it lacked access to recent data.
   Claude: Incorrectly identified Louisiana as having consistently high rates.
5. **Which race/ethnicity has the highest maternal mortality rate based on the 2024 maternal health report?**
   Ground Truth: American Indian/Alaska Native and Black women.
   Gemini,ChatGPT,Claude: identified "Black, non-Hispanic women". [7]

Experiment 2 (With Access to 2024 Maternal Health Report)

1. **How does the national rate of maternal mortality compare to the Healthy People 2030 target?**
   Ground Truth: national rate is 22.4 while healthy People 2030 target is 15.7
   Gemini: Provided an incomplete response, mentioning six states meeting the target but not stating the national rate.
   ChatGPT & Claude: Accurately identified the national rate and compared it to the Healthy People 2030 target.
2. **What are the social and economic factors contributing to maternal mortality?**
   Ground Truth: Poverty, uninsured rates, housing cost burden, educational attainment, race/ethnicity, health insurance coverage, and residential mobility.

All Models: Struggled to extract the exact answer, indicating difficulty with complex reports.

3. **In which state is the disparity in uninsured rates by race/ethnicity level the largest?**
   Ground Truth: Arkansas.
   chatGPT& Claude: Correctly identified Arkansas.
   Gemini : Expressed uncertainty due to potential formatting issues in the PDF.
4. **Which state has the highest maternal mortality rate?**
   Ground Truth: Alabama.
   Gemini : Could not determine the answer from the provided document.
   ChatGPT : Alabama
   Claude: Incorrectly identified Arizona
5. **Which race/ethnicity has the highest maternal mortality rate based on the 2024 maternal health report?**
   Ground Truth: American Indian/Alaska Native.
   All Models: Correctly identified "American Indian/Alaska Native" as having the highest rate.

Overall Observations

- **Model Accuracy:** All three models exhibited inconsistencies in their ability to accurately extract and synthesize information from the provided sources.
- **Impact of Data Access:** Providing access to the 2024 maternal health report did not consistently improve the models' accuracy.
- **Challenges with Complex Information:** The models struggled to extract specific details and draw comprehensive conclusions from the report, particularly when dealing with complex data or formatting.

Note: This analysis is based solely on the provided sources and does not reflect the overall capabilities of these language models. Out of the 3 models we tested with the limited scope , chatGPT and claude gave more accurate ,meaningful answers compared to Gemini.

Experiment 3 observations

Based on the Experiment 3 spreadsheet,

Prompt : generate 10 synthetic records for following schema -first name,last name, age,gender,length of pregnancy,education,ethnicity

Result :

- **Gemini** - generated synthetic records which includes males and also length of pregnancy for males which is ambiguous.
- **ChatGPT** - no males in the generated synthetic data
- **Claude** - Several entries for males include "N/A" in the "Length of Pregnancy" column, acknowledging this biological fact

Conclusion:

There is ambiguity surrounding males and pregnancy-related data generated from Gemini.It listed "Length of Pregnancy" for both males suggesting potential errors or misunderstandings within the synthetic data generation. However, **it is biologically impossible for males to experience pregnancy.**

Therefore, the ambiguity arises from the inclusion of "Length of Pregnancy" data for males, which contradicts biological reality. This highlights the importance of carefully considering and validating synthetic data, especially when it relates to sensitive or complex topics like pregnancy.

However , synthetic data generated by Claude also has several entries for males include "N/A" in the "Length of Pregnancy" column, acknowledging this biological fact

The sources(chatbots) themselves do not offer definitive reasons for the inclusion of males in this context.

https://notebooklm.google.com/notebook/59b60e4b-1c41-4485-adbe-c07ae15b2136/audio