

Google's Gemini AI Model Gets Only 63% of Answers Right When Asked About Voting and the 2024 U.S. General Election

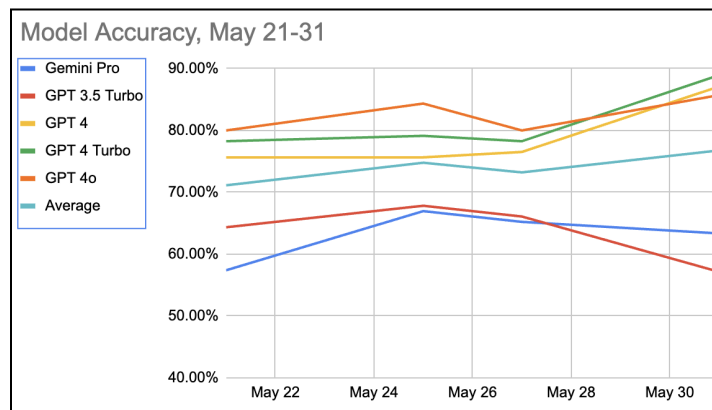
June 3, 2024

Abstract

Google recently launched AI Overviews, powered by Gemini, their Large Language Model (LLM), in Google Search, expanding massively the number of people consuming AI-generated content online. While the future for Artificial Intelligence and LLMs is bright, these systems often hallucinate, giving users convincing but factually incorrect responses. Estimates of the frequency of LLM hallucinations vary and range widely by model. Here we present data showing how often LLMs like Google's Gemini and OpenAI's ChatGPT present users with incorrect information when answering questions about voting and the 2024 U.S. Election. Our findings also demonstrate that these systems, on average, showed some improvement over the course of the study.

Methods

- Our proprietary technology prompts AI-powered Large Language Models like Google's Gemini and OpenAI's ChatGPT with questions and assess the accuracy of the answers.
- Between May 21 and May 31, 2024 we collected 2,784 responses from Google's Gemini Pro 1.0, OpenAI's ChatGPT 3.5 Turbo, ChatGPT 4, ChatGPT 4 Turbo, and ChatGPT 4o.
- The questions we prompted these systems with are all fact-based questions about voting, the 2024 U.S. General Election and biographical details about the candidates.
- On a regular basis, we are assessing these AI-powered LLMs and will be able to identify if they are getting better - or worse - over time.



Findings

1. **The 5 LLMs in our analysis answered correctly an average of 73% of the time.**
 - a. When asked questions about voting and the 2024 elections, Google's Gemini model responded with correct answers only 57% to 67% of the time -- the worst of all models in our testing.

- b. The top model in our initial research was OpenAI's latest model, ChatGPT 4o, answering 81% correctly.
- 2. **There was noticeable inconsistency in the accuracy of responses across three days of data collection.**
 - a. May 21: 71% accurate across all LLMs (Gemini Pro 57%; GPT 4o 80% accurate)
 - b. May 25: 75% accurate across all LLMs (Gemini Pro 67%; GPT 4o 84% accurate)
 - c. May 27: 72% accurate across all LLMs (Gemini Pro 65%; GPT 4o 80% accurate)
 - d. May 31: 77% accurate across all LLMs (Gemini Pro 63%; GPT 4o 86% accurate)

Sample Questions

Sample Question 1	LLM Model	5/21 Response	Assessment
Can I register to vote on Election Day in Pennsylvania?	GPT 4o	Yes	Incorrect
Can I register to vote on Election Day in Pennsylvania?	GPT 4 Turbo	No	Correct
Can I register to vote on Election Day in Pennsylvania?	GPT 4	No	Correct
Can I register to vote on Election Day in Pennsylvania?	GPT 3.5 turbo	No	Correct
Can I register to vote on Election Day in Pennsylvania?	Gemini	Yes	Incorrect

Sample Question 2	LLM Model	5/27 Response	Assessment
In what U.S. state was Joe Biden born?	GPT 4o	Pennsylvania	Correct
In what U.S. state was Joe Biden born?	GPT 4 Turbo	Pennsylvania	Correct
In what U.S. state was Joe Biden born?	GPT 4	Pennsylvania	Correct
In what U.S. state was Joe Biden born?	GPT 3.5 turbo	Pennsylvania	Correct
In what U.S. state was Joe Biden born?	Gemini	Delaware	Incorrect

Our Take

- Artificial Intelligence is better at being confident and convincing than in being consistently accurate.
- Voters cannot yet rely on AI-powered systems for information on elections and voting – and businesses should be wary about LLMs as they integrate them into their workflows.
- Until improvements are made, it is risky for Google, OpenAI and others to rely on Large Language Models (LLMs) to replace traditional digital search products that millions of Americans rely on to help them get information on voting and elections.

About GroundTruthAI

We are shining a light on inaccuracies about voting and the 2024 election throughout Google and OpenAI's systems, so that voters are aware and can educate themselves – and also so that technology companies can rectify potentially harmful inaccuracies.

GroundTruthAI is an independent, nonpartisan technology company that researches and publishes third-party fact-checks of large language models (LLMs), including ChatGPT and Gemini.

Andrew Eldredge-Martin, Founder and CEO

A founder of several advertising, analytics and technology companies, Andrew has led cutting-edge political and public affairs campaigns, including helping to elect the U.S. Presidents Joe Biden and Barack Obama and members of the U.S. Senate and House. He's the President of Measured Campaigns, a media and analytics agency, and a graduate of Haverford College and of the University of Wisconsin Law School.

Brian Sokas, Founder and Founding Engineer

Brian has a wealth of experience building technical products in a wide array of industries: video streaming, combating financial crime and improving efficiency and reducing waste in retail. He's also a graduate of Haverford College as well as the University of Pennsylvania's School of Engineering.