

AI Trials: Midpoint Reflections

Six Months of AI Experimentation

WILLIAM TREKELL

JUL 12, 2024

Share

Fertigkeits

I started a project six months ago to test AI platforms and understand their potential for everyday use. What began as a simple evaluation grew into a series of experiments. Now, at the halfway mark, I have learned more than I can share in a post. As such, this will have to serve as a summary, setting the stage for future in-depth pieces.

When I started, I was looking to expand my toolkit. Instead, I was reminded that creativity thrives when free, unchained by rigid processes and procedures. AI promised efficiency and productivity but gave me a challenge:
"Dig deeper – I'll hold your scotch."

My approach evolved through exploring various prompting techniques. While "cheatsheets" and claims of "simple prompts" were clearly basic, they served as signposts, guiding me towards more substantive experiments with roles, templates, tones, and beyond.

Additionally, the past year has seen a number of releases, notably Anthropic's Claude iterations and OpenAI's GPT-4 updates. These advancements enhanced AI's ability to understand context and solve problems. As new models emerged, I expanded testing to include various model versions, adding another dimension to my experiments.

Major AI Model Releases:

- **ChatGPT-3.5** (November 2022)
- **Bard** (March 2023)
- **ChatGPT-4** (March 2023)
- **Custom GPTs** (July 2023)
- **Claude 2.1** (November 2023)

- **ChatGPT-4 Turbo** (November 2023)
- **Gemini Pro 1.0** (December 2023)
- **Gemini Pro 1.5** (February 2024)
- **Claude 3 Sonnet** (March 2024)
- **ChatGPT-4o** (May 2024)
- **Claude 3.5 Sonnet** (June 2024)

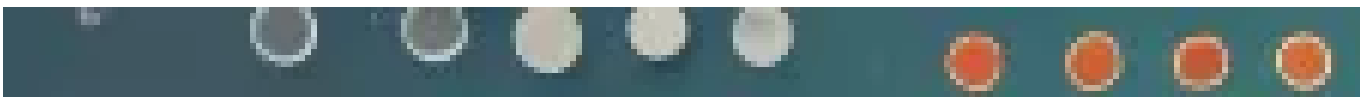
With AI evolving so quickly, I focused on maintaining baselines for comparing improvements across model iterations. Areas of focus in this project include:

1. **Data Collection:** Creating simple event guides, which have evolved into comprehensive holiday tables which revealed inconsistencies across AI models and required meticulous cross-referencing.
2. **Specialized Roles:** Assigning roles, which progressed from general to highly specialized, task-specific roles, culminating in the development of a GPT for the purpose of creating them with little effort.
3. **Template Development:** Creating standardized templates for consistent article structure, allowing me to establish scoring mechanism despite occasional AI deviations.
4. **Output Examples:** Using example articles to guide authoring and evaluation, which proved challenging when attempting to calibrate scoring spectrums independently.
5. **Evaluation Methods:** Developing a framework to assess quality, ethics, and cultural sensitivity, which continues to face challenges in achieving consistent success across multiple tests.
6. **Tone Consistency:** Exploring AI's ability to maintain consistent tones and style across diverse content, revealing varied results and the depth added by prescribed tones.

Looking Ahead

As experiments shift from exploratory to being fine-grained in nature, I've begun standardizing the prompts I use regularly and reading AI documentation in-depth. This standardization will validate progress and shape the finalization of current findings.

Next, I'll explore advanced AI prompting techniques and share insights through a series of focused articles. These articles will cover specific aspects of my findings, helping readers maximize AI benefits.



Quiet Evolution is about experimenting and sharing insights. If you find this helpful, [coffee is always appreciated](#) (no pressure). Proceeds are used strictly to cover AI costs; any excess goes to the [American Cancer Society](#).



Disclaimer: The views and opinions expressed in this article are solely those of the author and do not reflect the official policy or position of [Amazon Web Services \(AWS\)](#). The author is a UX designer at AWS and has no involvement in, nor does the work pertain to, any collaborative agreements that AWS may have with [Anthropic](#), the creators of [Claude](#). The insights and analyses presented here are entirely independent and unrelated to any projects or initiatives between AWS and Anthropic. All content in this post is based on publicly available interfaces and is not influenced by the author's employer.

Discussion about this post

Comments

Restacks



Write a comment...