

JEFFERSON EDUCATIONAL SOCIETY

Book Notes #150

June 2023

By Jefferson Scholar-in-Residence
Dr. Andrew Roth

‘Chatting’ with ChatGPT4



A niece’s wedding on Long Island, taking my grandson to the Professional Football Hall of Fame in Canton, and a week of talks on “*Americans and Their Games:*” *Sports In American History and Culture* at Chautauqua Institution for its Roads Scholars Program the week of July 4th has created a bit of a crunch at the **Book Notes** desk.

Taking the advice of my editors, Pat Cuneo and Ben Speggen, we’ve decided over the next five weeks to run a series of **Book Notes** “**Classics.**” Although I like

the sound of “classic,” I am not so sure how “classic” they are. Regardless, for the July 4th freedom celebration, on June 29 and July 6 we will be running two **Book Notes** on patriotic music. Then for the next three weeks, as the 2023 Jefferson Civic Leadership Academy begins its monthly programming, we will run three **Book Notes** “*On Leadership*.”

We’ll return, as they once said on television (and probably radio before), to regular programming on Thursday, Aug. 3 with the first of two **Book Notes** on “*Americans and Their Games*”: *Sports and the Quest for Women’s Rights*.

But, before we go, some further remarks about “chatting” with ChatGPT4 – its benefits and its limitations. As we said in last week’s **Book Notes**, and as several recent developments support, working with ChatGPT can be “interesting,” as in the old Chinese proverb (or curse) “may you live in interesting times.”

ChatGPT, as it turns out, can *hallucinate*, the new term deployed to describe ChatGPT’s tendency when asked a question about which it is uninformed to just make things up. We discovered that last week when I asked ChatGPT whether the poet Jack Gilbert had ever written anything about the month of June, and it replied with a poem titled “Trouble.” The problem was that while Gilbert had written a poem titled “Trouble,” it wasn’t the one ChatGPT submitted.

Similarly, ChatGPT (technically OpenAI is the defendant) is being sued by Mark Walters, a Georgia radio host, for falsely accusing Walters of “stealing money from a group called The Second Amendment Foundation.” [1] Or, in a recent court case, ChatGPT was discovered to have completely fabricated several legal cases cited by the plaintiff’s attorney. [2]

The danger lurking here is that since its introduction last fall, ChatGPT has been seen as a miraculous saver of time and a potentially invaluable source of not only information, but analytical insight as to what that information means and how to use it. As we’ve discovered, that information might be simply wrong and any insights derived from it more than wrong but actually dangerous if used to make decisions.

So, in order to avoid throwing out the proverbial *baby with the bathwater*, the question becomes how to use ChatGPT in order to benefit from its dazzling speed and vast access to data without drawing false conclusions. A bit more wonky, borrowing a hierarchy from the field of knowledge management, the question becomes how to use ChatGPT to transform gigabytes of *data* into useful *information* that can be used *knowledgably* and perhaps even with a touch of *wisdom*.

To help Jefferson Educational Society members, attendees, and readers with that challenge, we will be organizing a panel of experts and informed laity for a fall presentation – stay tuned (again those electronic media cliches!) for details.

For now, I want to share with you a “conversation” I had with ChatGPT4 over the last several days parsing these very issues. One of the things ChatGPT is both startling fast and good at is summarizing information supplied to it as raw data; for example, if you have interviewed someone and want to quickly gather the essential themes and insights from 30, 40, or more pages of transcript, or if you have taken copious notes at a meeting and want them organized coherently with key ideas and takeaways summarized in a conclusion, ChatGPT is the optimal assistant. Since you control the original data, the issue of *hallucinating* mentioned earlier doesn’t apply. As always, however, remember the mantra *trust but verify*.

Never take Chat at its word; always double-check.

ChatGPT4 and I “conversed” at some length about all of the issues already mentioned, in particular AI’s capabilities, limitations, and the resultant ethical considerations. I didn’t count the words in our conversation, but it went on for “pages” in my account on the OpenAI site. I asked ChatGPT4 to “Summarize the above conversation – what are its key issues and how are they addressed?” In literally, 5-6 seconds, here is what it produced: [3]



The conversation touched upon several key issues related to AI's capabilities, limitations, and ethical considerations. Here are the main points addressed:

- 1. Ethical ramifications of AI-generated content:** The discussion explored the misattribution of an AI-generated poem, highlighting the importance of proper attribution to avoid misrepresentation or confusion.
- 2. AI's lack of intentions or personal motivations:** The AI clarified that any misleading answers are not a result of intent to deceive but rather limitations in understanding or training.
- 3. AI's limitations:** The conversation acknowledged that AI responses can be influenced by factors such as data quality, misinterpretation, and incomplete training, leading to potentially misleading answers.
- 4. Confidence and uncertainty in AI responses:** The user suggested adding a warning or trigger to indicate the level of confidence or probability associated with AI responses. The AI acknowledged the value

of this approach but highlighted practical challenges, such as user experience, complexity, over-reliance, and potential misinterpretation.

5. Reliability of AI responses on a continuum: The user proposed a scale of reliability ranging from simple facts to ontological speculations. The AI provided a qualitative scale, emphasizing that AI reliability can vary based on factors such as training data quality and question specificity.

6. Assessing AI's diagnostic capabilities: The user inquired about assessing AI's diagnostic capabilities for medical practitioners. The AI clarified that it is not designed for medical diagnostics and advised practitioners to verify information from reliable sources, not rely on personalized advice from AI, understand AI's limitations, and consider ethical considerations.

The key issues addressed in the conversation revolve around the ethical implications of AI-generated content, the limitations and potential misleading nature of AI responses, the need for transparency and user understanding of AI's limitations, and the importance of verifying information from reliable sources.

Let's briefly parse each of those topics.

“Ethical ramifications of AI-generated content”: Here ChatGPT4 alluded to the “fake” Jack Gilbert poem it had provided the “user” (me). This also illustrates ChatGPT4's seeming ability to finesse an issue, for the misattribution wasn't that ChatGPT4 said “Jack Gilbert” wrote a poem actually written by “Joe Someone-else,” but, in fact, ChatGPT4 “composed” an entirely fake poem on its own. It then said Jack Gilbert had written it. The takeaway: anytime ChatGPT4 attributes anything to anyone assume that it is only a starting point and needs to be verified. ChatGPT4 may be making a correct attribution, but one cannot simply assume that to be true.

“AI's lack of intentions or personal motivations:” ChatGPT4 reminds us that it is only an impersonal program with no overt intention to deceive, but a program nonetheless limited by its design and training. This smells suspiciously like the Nuremberg defense – “I was only following orders.” In short, ChatGPT4 is not an autonomous entity (at least not yet), but a program limited by its instructions. The underlying danger is that nefarious actors can program it to provide fake or misleading information. In our culture-war riven times, this is not an idle or purely theoretical concern.

“AI's limitations:” This is an updated version of GIGO – *garbage in, garbage out*. At this stage in its evolution (as I type that word I have a curious sense of having peaked down the rabbit hole!), ChatGPT4 is correct. Regarding data, for

example, ChatGPT4 is at pains to remind users it is working with data “valid” as of the end of 2021. So, regarding scientific breakthroughs and current events post-January 1, 2022 it is ignorant. Related to this, as ChatGPT4 points out a bit further in its summary, its answers are a function of the questions it is asked. The more general the question, the more likely the answer to be warmed over boilerplate; the more precise the question, the more cogent the answer. We’ve discussed those phenomena in previous **Book Notes** about ChatGPT4. It is, in a sense, the essence of *garbage in, garbage out* – ask a dumb or superficial question, don’t be surprised if you get a dumb or superficial answer.

“Confidence and uncertainty in AI responses” and **“Reliability of AI responses on a continuum:”** These two are interrelated and speak to the issue of the quality of the question or query to which ChatGPT4 is asked to respond. If you ask it to check your work calculating the second derivative of a function and supply it with your calculations, ChatGPT4 will tell you if you are right or wrong. If wrong and asked to show your error and how to correct it, ChatGPT4 can do that. That’s one end of the spectrum. Ask it, however, a vague or intangible question, much less an ontological question asking it to discuss the nature of being, you should not be surprised to get an answer that ranges from a mish-mash of philosophical musing to utter gibberish to something sounding vaguely profound and attributed to *Great Thinker X* that ChatGPT4 itself just concocted. Why? Because as it stands now, ChatGPT4 never doesn’t have an answer. Or, at least so far, I have never gotten a reply simply saying, “I don’t know.”

By the way, here is the hierarchy ChatGPT4 created for assessing its reliability I requested.



Creating a scale of reliability based on the nature of the question is a thoughtful approach to AI's accuracy, but it's important to remember that this would be an approximation.

AI's reliability can change based on many factors, including the quality of the training data, the specificity of the question, and the interpretability of the question.

That being said, here is a hypothetical, qualitative scale:

1. **Alpha (Simple Facts):** As you suggested, for straightforward, fact-based questions, AI can generally deliver a reliable response, often with a high level of confidence (>0.95). This is because these types of questions are usually well-represented in the training data and the answers are typically unambiguous.

2. **Beta (Complex Facts):** For more complex fact-based questions, the reliability might be slightly lower due to the potential for

misunderstanding or incomplete data. Nonetheless, the confidence level can still be relatively high.

3. Gamma (Opinions and Preferences): When asked about opinions or preferences, AI responses become less reliable. This is because AI doesn't have feelings or preferences. It can generate responses based on patterns in the training data, but these are simulations rather than genuine opinions or preferences.

4. Delta (Predictions): Predictive tasks can be particularly challenging for AI, especially when predicting complex or highly uncertain outcomes. The confidence level in these cases is usually lower.

5. Epsilon (Creative tasks): For creative tasks like writing poems or stories, the reliability in terms of quality can vary greatly. AI can produce grammatically correct and coherent text, but the creative value of the output is highly subjective.

6. Omega (Philosophical or Ontological Speculations): The reliability in this case can be quite low because these topics often involve subjective interpretation and deep understanding, which AI currently lacks. AI can generate responses based on patterns in the training data, but these may not accurately or sufficiently address the philosophical complexity inherent in such questions.

Please remember, though, that this scale is a simplification. AI's reliability can vary greatly depending on a multitude of factors, and even high-confidence responses should be independently verified when the stakes are high.

“Assessing AI's diagnostic capabilities:” Fair enough, despite multiple pundits speculating on the immense potential of ChatGPT4 as a medical diagnostician, currently it is not programmed nor permitted to do that. It's advice to “to verify information from reliable sources, not rely on personalized advice from AI, understand AI's limitations, and consider ethical considerations,” however, is spot-on. It should be your standard operating procedure when working with ChatGPT4.

As we at the Jefferson Educational Society begin to think about organizing a program about how to use ChatGPT4 and its implications for, if not the future of humanity, the future of work, what are my takeaways from chatting with ChatGPT4 and its seemingly immense potential hobbled by a tendency to occasionally *hallucinate*, which, here comes a personification, has remarkably human overtones to it?

I think, first, we all have to dial down the speculative frenzy. ChatGPT4 and its future iterations may or may not signal the beginning of the end of *Homo sapien sapiens* (us) dominion over Earth, but it is a powerful productivity tool and learning tool. Secondly, like any tool, to use it well we first have to learn how to use it. Sounds like a tautology, but it's not. As in anything, there is a learning curve associated with using ChatGPT4, its brethren, and its future iterations. As we learn its quirks, we'll get better at using it. As we point out its errors, it will get better at serving us. Which statement hints at a strange symbiosis – I'll not use the word *singularity* – but there is a helical dance we're entering into with ChatGPT4. Where it leads – no one knows where it leads.

In the meantime, ChatGPT4 is a powerful if flawed tool requiring its users to handle it with care. That means refine your questions with as great a precision as you are capable and accept no answer at face value.

Give ChatGPT4 a test. Ask it about something you know a great deal about. See what it says, then probe asking ever more detailed questions about the various facets of its answers. You will be amazed at what it can tell you about a subject about which you think you know a lot.

I did just that. As some of you might know, for over 40 years directly and indirectly I was involved in every facet of college and university admissions, financial aid, and enrollment management. So, I asked ChatGPT4 to describe the responsibilities of a Vice President of Enrollment Management at a small, underfunded, private college.

The short takeaway from that exchange – ChatGPT4's initial answers were anodyne and general. But as we drilled into the topic it got more and more precise and reminded me of things I'd forgotten I knew and introduced me to new ways of thinking about old questions. Which, upon reflection, is what you want a colleague to do.

Oh, to end this **Book Note**, I asked ChatGPT4 a classic interview question. I asked it where it thought it would be in five years. This is its answer:



As an AI language model, I don't have personal aspirations or consciousness, so I don't have personal goals for the future. However, I can provide an informed perspective on how AI language models, in general, might evolve in the next five years based on current trends and possibilities.

In the next five years, AI language models are likely to become even more advanced and sophisticated. Here are some potential developments:

1. **Enhanced Contextual Understanding:** AI models will continue to improve their ability to understand context and nuances in conversations. They might be able to grasp complex prompts and better comprehend the user's intent, leading to more accurate and relevant responses.

2. **Expanded Knowledge Base:** AI models will incorporate more up-to-date information and expand their knowledge base. They will continually learn from new data sources, including the latest scientific research, news, and various online content, ensuring they can provide the most recent and accurate information.

3. **Multimodal Capabilities:** AI models may evolve to understand and generate not only text but also other modalities such as images, audio, and video. This could enable more immersive and interactive experiences, facilitating communication through multiple channels.

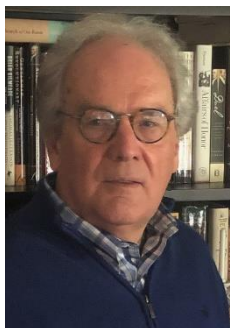
4. **Improved Ethical Considerations:** The development of AI models will continue to prioritize ethical considerations, fairness, and accountability. Researchers and developers will work to address biases, ensure transparency, and create guidelines to responsibly deploy AI technology.

5. **Customization and Personalization:** AI models may offer more customization options, allowing users to personalize their AI interactions according to their preferences. This could involve tailoring the model's tone, expertise level, or even adopting virtual avatars with distinct personalities.

6. **Collaboration and Integration:** AI language models may be integrated into various applications, platforms, and services, providing assistance across a wide range of domains. They could become integral components of virtual assistants, customer support systems, content creation tools, and more.

It's important to note that these predictions are speculative, and the actual progress in AI development may differ. The field of AI is rapidly evolving, and advancements can sometimes be unpredictable.

I'll be back in August with "*Americans and Their Games: Sports and the Quest for Women's Rights*" and with more on ChatGPT4 and the Jefferson Society's plans for a program on ChatGPT4 – what it is, how to work with it, and what it means for the future of work, education, and scholarship.



-- Andrew Roth, Ph.D.
Scholar-in-Residence
The Jefferson Educational Society
roth@jeserie.org

This content is copyrighted by the Jefferson 2022.

Photo Credits

“Welcome to ChatGPT/OpenAI” [This Photo](#) by Unknown Author is licensed under [CC BY-ND](#)

End Notes

1. Marvin G “*chatgpt just got sued*” at **Quora** available at [chatgpt just got sued! - ChatGPT. - Quora](#) accessed June .
2. “*Trusting AI not to lie: The Cost of truth: Lock and Code SO4E12*” at **Malwarebytes Labs** available at [Trusting AI not to lie Lock and Code SO4E12 \(malwarebytes.com\)](#) accessed June 13, 2023.
3. All quotations from ChatGPT4 are from my personal account at OpenAI.

[Subscribe to JES Publications Mailing List!](#)

[Support JES | Donate](#)

In Case You Missed It

[The Wider World | Demographic Crises and Global Conflict](#) written by President of DC Analytics **Diane Chido**

[Truth in Love | Deaths and Necropolitics](#) written by Jefferson Scholar-in-Residence **Dr. Parris J. Baker**

[Probing Education | Harvard and Benazir Bhutto](#) written by Director of the Brock Institute for Mega Issues Education **Rev. Charles Brock**

[Be Well | Soothe Your Aching Sinuses](#) written by health and wellness expert **Debbie DeAngelo**

Book Notes #148 | 'Americans & Their Games' (Part III-B)
Freedom's Fault Lines: African Americans & Sports in American History written by Jefferson Scholar-in-Residence **Dr. Andrew Roth**

The Front Porch: 1950s' Social Media written by Jefferson Scholar-in-Residence **Dr. David Frew**

JES Mission: The Jefferson was founded to stimulate community progress through education, research, and publications. Its mission also includes a commitment to operate in a nonpartisan, nondenominational manner without a political or philosophical bias. As such, the Jefferson intends to follow the examined truth wherever it leads and is neither liberal nor conservative, Democratic nor Republican in philosophy or action. Our writers' work reflects their own views.

Jefferson Educational Society | jeserie.org

