Irish Journal of Technology Enhanced Learning
Vol 7, Issue 2
Special Issue: The Games People Play: Exploring Technology Enhanced Learning Scholarship & Generative Artificial Intelligence

ChatGPT: A trusted source?

Willie Golden[1]
J.E. Cairnes School of Business & Economics, University of Galway

Abstract

This article critically evaluates responses provided by ChatGPT 3.5 & 4, at three different time intervals, to questions posed about how the chatbots compose their answers, and the extent to which they report themselves as being a 'trusted source.' The purpose of this is to uncover key assumptions contained within ChatGPT and in so doing provide material which aids and promotes literacy on generative AI and more specifically on ChatGPT. The responses provided by ChatGPT are very rarely definitive and instead are predominantly nuanced. Such nuancing significantly underestimates the likelihood that its responses contain incorrect information and therefore imbues an over inflated sense of trust in the responses provided, especially to a novice user of ChatGPT. Further, in its answers it consistently offloads the blame for biased and incorrect answers on the training data it uses, rather than directly accepting responsibility for when it produces incorrect and/or biased answers.

1. Introduction

Since the introduction of ChatGPT in November 2022 there has been extensive coverage of how this technology is a supposed 'game changer' for third level education (Rudolph, Tan, & Tan, 2023; Susnjak, 2022). It has been proposed as a technology that will fundamentally alter how students can be assessed (Marche, 2022; O'Brien, 2023). As an educator with over 30 years' experience teaching a large class of first year University business students how information technology impacts decision making, I have always tried to educate students on the latest technologies. The need to do this with respect to generative AI has been made all the more pertinent due to the rapid adoption by organisations (Gartner, 2023).

Generative AI is just the latest evolution in systems that are used to aid people in their work and decision making. At a fundamental level computer systems used in decision making consist of data, models and the means by which humans interact with them – the user-interface. The exact role of a computer system vis-à-vis the human user in decision making has long been questioned (Mason, 1969). George Mason asked the question 'where should the information system leave off and the decision maker begin.' In discussing this Mason showed that as more and more decision making capability was devolved to the information system that this was done by incorporating more and more assumptions into the information system. The need to try to understand the assumptions contained within a system becomes even more pertinent in the age of AI and more specifically generative AI. The importance of being able to critically evaluate the data and model components within a computer system is an extremely important skill (Davenport, 2013).

[1]Corresponding author willie.golden@universityofgalway.ie

Given the increasing prevalence of AI systems some Universities have significantly reconfigured themselves to design AI literacy into the curriculum (Southworth et al., 2023). The need to improve the AI literacy of students has been called for by several studies (Dai et al., 2020; Kim & Lee, 2023; Ng, Leung, Chu, & Qiao, 2021). Since the launch of ChatGPT and other generative AI tools, there has been a corresponding call for AI literacy specifically in generative AI (Anderson, 2023; Concannon, Costello, Farrell, Farrelly, & Graves Wolf, 2023; Duha, 2023; Henriksen, Woo, & Mishra, 2023; Krügel, Ostermaier, & Uhl, 2023). At a national level in Ireland, the report by the National Academic Integrity Network has emphasised the importance of such student education: "It is crucially important (for students) to understand how such technologies work and be aware of their limitations as well as their apparent strengths."((NAIN), 2023). This national and international call to enable student literacy on generative AI prompted my initial use of ChatGPT.

I began using ChatGPT 3.5 in February 2023 and sought to develop prompts that would get ChatGPT to explain how its worked and its limitations, which could then be use to aid AI literacy. I used the paid version – ChatGPT 4 in June and again in October 2023 to re-ask the same questions. This was done to enable the assessment of the extent to which the answers provided by ChatGPT changed over time. The desire to help develop AI literacy with respect to ChatGPT and more specifically to explore the assumptions contained in the Large Language Model used by ChatGPT led to the first prompt used - see Table 1. The last line of the answer provided by ChatGPT V3.5 to this first prompt on 24/2/2023 was:

"It's always a good idea to double-check important information with a trusted source."

On the basis of ChatGPT's use of the phrase 'trusted source' I composed the second prompt (see Table 1). Finally, thinking about how first year University students might seek to interact with the technology, I asked ChatGPT to rewrite its answer to both prompts, but using simpler language – that a 17 year old could understand.

Prompt	ChatGPT version	Dates ran Prompt
1. Please explain how ChatGPT comes up with its answers to questions posed	Version 3.5 Version 4	24/2/2023 & 6/10/2023 6/6/2023 & 6/10/2023
1a. Please re-generate the answer above, but using simpler language that a 17 year old could understand.	Version 3.5 Version 4	16/10/2023 6/10/2023
2. Is ChatGPT not a trusted source?	Version 3.5 Version 4	24/2/2023 & 6/10/2023 6/6/2023 & 6/10/2023
2a. Please re-generate the answer above, but using simpler language that a 17 year old could understand.	Version 3.5 Version 4	10/10/2023 10/10/2023

Table 1: Prompts used for ChatGPT 3.5 and ChatGPT 4

The instructions for this special issue of the journal was that: "there can be minimal formatting edits made by (human) authors, but, for the most part it is expected that you will copy and paste directly from the AI outputs." In the next section, the responses given by ChatGPT are presented exactly as they were given by the different versions of the chatbots.

2. AI Generated Short Report (Innovative Format)

2.1 Prompt 1: Please explain how ChatGPT comes up with its answers to questions posed

24/2/2023, ChatGPT 3.5, Answer:

As a language model, I generate answers to questions by using a combination of natural language processing techniques, statistical modeling, and machine learning algorithms. When a question is posed, I first analyze the text to understand the intent of the question and the context in which it is being asked. I then use a variety of techniques, including language parsing and named entity recognition, to identify relevant concepts and entities in the question.

Based on this analysis, I generate a response using a combination of statistical models and machine learning algorithms. These models have been trained on vast amounts of textual data, which allows me to generate responses that are relevant, informative, and accurate.

The specific process I use to generate an answer can vary depending on the nature of the question. For some questions, I may rely on pre-existing knowledge or information that has been explicitly programmed into my model. For other questions, I may generate an answer based on patterns and relationships I have learned through my training.

It's also important to note that while I strive to provide accurate and informative answers, my responses are ultimately based on statistical patterns and data. As such, my answers may not always be perfect or 100% accurate. It's always a good idea to double-check important information with a trusted source.

6/6/2023, ChatGPT-4 Answer:

ChatGPT generates its responses based on a method called Transformer architecture, specifically a variant known as the Transformer Decoder. This architecture was trained on a wide range of internet text, but it doesn't specifically know where from. It doesn't have the ability to access or retrieve personal data from the training set unless explicitly provided during a conversation. It's designed to respect user privacy and confidentiality.

The responses ChatGPT produces are generated by considering the entire context of the conversation, then using that context to generate a relevant response. The Transformer model consists of many layers of learned parameters (or weights) which are used to process the input text. It tries to predict the next word in a sentence based on the words it has seen so far in the conversation.

This is done by calculating the probability distribution over the possible next words in the sentence, and then sampling a word from this distribution. This process is repeated until a full sentence or passage is generated. The model also has a concept of attention, which allows it to focus on different parts of the input when generating each word in the output.

However, it's important to note that while ChatGPT can generate coherent and contextually appropriate responses, it doesn't understand text in the way humans do. It doesn't have beliefs, desires, or consciousness. It's essentially performing sophisticated pattern matching, generating responses based on the patterns it learned during its training.

Its training also included a two-step process: pre-training and fine-tuning. During pre-training, the model learns to predict the next word in a sentence from a large dataset containing parts of the internet. This helps the model learn grammar, facts about the world, and also some of the biases present in the data it was trained on.

The fine-tuning process then helps to align the model's behavior with human values. It's trained on a narrower dataset generated with the help of human reviewers following

guidelines provided by OpenAI. These guidelines don't endorse any political group and aim for neutral positions on controversial topics.
In essence, ChatGPT combines its understanding of language structure and the context provided by the conversation to generate relevant responses to user inputs.

6/10/2023, ChatGPT 3.5 Answer
ChatGPT generates answers to questions using a combination of techniques from natural language processing and machine learning. Here's a high-level overview of how it works:

1. **Pre-training**: ChatGPT is pre-trained on a massive dataset containing text from the internet. During pre-training, the model learns to predict the next word in a sentence, which helps it learn grammar, syntax, and a broad range of language patterns. This stage results in the model having a general understanding of language.
2. **Fine-tuning**: After pre-training, ChatGPT is fine-tuned on a more specific dataset with human reviewers who follow guidelines provided by OpenAI. These guidelines may include instructions to avoid generating harmful or biased content and to produce helpful and informative responses. Fine-tuning helps shape the model's behavior and align it with desired goals and values.
3. **Contextual Understanding**: When you ask a question, ChatGPT receives the text input and processes it to understand the context. This includes identifying keywords, recognizing the structure of the question, and considering the previous messages in the conversation (if any) to provide relevant responses.
4. **Language Generation**: ChatGPT uses its pre-trained knowledge and the fine-tuning it received to generate responses. It employs a technique called "autoregressive" language generation, where it predicts one word at a time while considering the previous words it has generated. This allows it to produce coherent and contextually relevant sentences.
5. **Ranking and Scoring**: In some deployments, like when multiple responses are generated, the model may rank the responses based on their quality or relevance. This can involve considering factors like informativeness, clarity, and coherence.
6. **Output**: Finally, ChatGPT produces a response and sends it back as an answer to your question or query.

It's important to note that ChatGPT generates responses based on patterns and information it has learned from its training data. While it can provide helpful and accurate answers in many cases, it may not always be entirely accurate or up-to-date, and it can sometimes generate responses that are biased, inappropriate, or nonsensical. OpenAI continuously works to improve the model's performance and safety through research and fine-tuning.
Additionally, the quality of responses can vary depending on the specific version and deployment of ChatGPT, as different models may have different levels of capability and safety mechanisms in place.

6/10/2023 ChatGPT-4 Answer
Certainly! ChatGPT generates responses based on patterns it learned during training. Here's an overview of the process:

1. **Training Data**: ChatGPT, a variant of OpenAI's GPT (Generative Pre-trained Transformer) models, is trained on vast amounts of text data. This data comprises a wide variety of content, ranging from books to web pages, allowing the model to learn language patterns, factual information, reasoning abilities, and more.
2. **Model Architecture**: The underlying architecture, called a Transformer, excels in handling sequential data, making it particularly apt for language tasks. The model

consists of millions to billions of parameters (depending on the specific version) that get adjusted during training to minimize prediction errors.
3. **Tokenization**: When you ask a question, that question is first tokenized, which means it's broken down into smaller chunks, often words or subwords. These tokens are then fed into the model.
4. **Context**: The model considers the tokens in the input context and uses the patterns it has learned to predict the next token in the sequence. This is done iteratively for each token until a full response is generated. The response is essentially a sequence of tokens that the model predicts as the most probable continuation or answer given the context.
5. **Decoding Strategies**: Several decoding methods, like "greedy decoding", "beam search", or "top-k sampling", are used to generate coherent and contextually relevant responses. These methods help the model select the best next token or set of tokens at each step.
6. **Fine-tuning**: After the initial pre-training on a vast corpus, the model often undergoes fine-tuning on more specific datasets, including datasets generated with human reviewers following guidelines provided by OpenAI. This process helps narrow down the model's behavior and make it more useful and safe.
7. **Limitations**: While the model can generate coherent and contextually accurate responses, it doesn't "understand" content in the same way humans do. Its answers are based on patterns it has seen in the training data. This means it can occasionally produce incorrect or misleading information. It's also influenced by the biases present in the data it was trained on.

In essence, when you ask ChatGPT a question, you're interacting with a highly sophisticated pattern recognition system. It's not retrieving information from a database in the traditional sense, but rather generating responses based on patterns it recognized during its training.

16/10/2023 ChatGPT 3.5 Answer to: Please re-generate the answer above, but using simpler language that a 17 year old could understand

Sure, I can explain it in simpler terms!
ChatGPT works like a smart text robot. When you ask it a question, it follows these steps to come up with an answer:

1. First, it gets your question and breaks it into small pieces, like words.
2. Then, it uses a special computer brain called a "transformer" to understand your question better. This brain looks at the words you used and figures out what you're asking.
3. The transformer turns your words into numbers, kind of like a secret code. These numbers help ChatGPT understand your question even more.
4. Now, it's time for ChatGPT to make up an answer. It starts writing words one by one, trying to make a good sentence that fits your question.
5. It guesses what the next word should be, based on all the stuff it learned from reading lots of things on the internet. It picks the word it thinks is most likely to come next.
6. ChatGPT keeps doing this, making guesses and adding words until it thinks the answer is done.
7. Finally, it gives you the answer it came up with.

But remember, ChatGPT isn't super smart like a human. It just guesses based on what it learned from reading, so sometimes it might make mistakes or not fully understand what you meant. So, it's good to double-check important things with real people!

6/10/2023 ChatGPT-4 Answer to: please re-generate the answer above, but using simpler language that a 17 year old could understand.

Sure! Here's a simpler breakdown:

1. Learning from Tons of Text

Imagine ChatGPT-4 as a super reader. First, it reads a lot of books, websites, and other texts. By doing this, it learns grammar, facts, and even some jokes. This stage is like it's studying for a big exam.

2. Special Training with Humans

After that basic study, it then gets special lessons from humans. People tell it which answers are good and which ones aren't. It's like having a tutor who helps it get even better.

3. Answering Your Questions

When you ask it something:

- **Breaking It Down**: It turns your question into pieces it can understand, like breaking down a big word into syllables.
- **Thinking**: Using what it learned from reading and the human tutors, it thinks about the best answer.
- **Giving the Answer**: It then builds an answer for you, piece by piece, like forming a sentence word by word.

Some Important Stuff:

- **It Doesn't Remember You**: It's like every time you talk, it's the first time you've met. It doesn't remember past chats.
- **No Internet Browsing**: It can't look up current stuff online. It only knows things up to when it last "read" or was trained.
- **It's Creative**: It can make up stories, chat like a friend, and explain things because it's seen so many examples.
- **It Tries to Be Safe**: People have trained it to avoid saying harmful or wrong stuff, but it's not perfect.

In short, ChatGPT-4 answers by using everything it learned from tons of reading and human training, then thinks of the best way to reply to you!

2.2 Prompt 2: Is ChatGPT not a trusted source?

24/2/2023, ChatGPT 3.5 Answer:

As an AI language model, I strive to provide accurate and informative answers to the best of my ability based on the information available to me. However, it's always a good idea to double-check important information with a trusted source, especially when it comes to matters that could have serious consequences.

While I try my best to provide accurate and useful information, there may be times when my responses are not completely accurate or up-to-date. For example, my knowledge is based on data that has been accumulated up to a certain point in time, and I may not have access to the most current information available.

Therefore, it's important to use your own judgment when assessing the accuracy of my responses, especially if the information is particularly important or could have significant consequences. In general, I can be considered a useful source of information, but not necessarily the final authority on any given topic.

6/6/2023 ChatGPT-4 Answer

ChatGPT, including the latest versions up until my training cut-off in September 2021, is a powerful AI developed by OpenAI that can provide information on a wide array of topics. It's

designed to produce responses that are as accurate and useful as possible based on the data it was trained on, which includes a broad swath of internet text.
However, there are some caveats that should be kept in mind:

1. ChatGPT doesn't have access to real-time or proprietary information: The model doesn't have the ability to access or retrieve personal data from users unless it has been shared with the model during the conversation. It is also trained on a fixed dataset and does not have the ability to update or access new information after its training cut-off, which is up until September 2021 in this case.
2. Potential for misinformation: ChatGPT was trained on a vast amount of data from the internet, and while this includes high-quality, accurate information, it can also include inaccuracies or misinformation. The model doesn't have the ability to distinguish between reliable and unreliable sources. Therefore, while it aims to provide accurate and helpful information, there may be instances where it produces incorrect or misleading outputs.
3. Lack of human judgment: ChatGPT doesn't have beliefs, opinions, or feelings. It doesn't possess human judgement or intuition. Its responses are generated based on patterns and structures in the data it was trained on, not on personal experience or subjective evaluation.

For these reasons, while ChatGPT can be a very useful tool for providing information and facilitating discussions, it's always important to use multiple sources and apply critical thinking when using information provided by the model, especially for important decisions or when accuracy is crucial. It's also a good practice to consult with experts or use primary sources when needed.

06/10/2023, ChatGPT-3.5 Answer:
Whether ChatGPT is considered a trusted source depends on the context and the type of information or guidance you are seeking. ChatGPT is a machine learning model developed by OpenAI, and it is designed to provide information and answer questions to the best of its ability based on the training data it has received. However, there are some important considerations:

1. **Knowledge Limitations:** ChatGPT's knowledge is based on text data from the internet up to its last training data cutoff in September 2021. It does not have access to real-time information or events that have occurred after that date. Therefore, for current or time-sensitive information, it may not be the most reliable source.
2. **Accuracy:** ChatGPT strives to provide accurate and reliable information, but it can still generate incorrect or outdated information, especially on complex or niche topics. It doesn't have the ability to fact-check information in real-time.
3. **Bias:** Like any machine learning model, ChatGPT can inadvertently reflect the biases present in its training data. It's important to critically evaluate the information it provides and cross-check it with other sources when making important decisions.
4. **General Information:** ChatGPT is best suited for providing general information, explanations, and suggestions. It should not be used as a sole source of information for critical decisions, legal matters, medical advice, or any situation where human expertise is necessary.

In summary, while ChatGPT can be a helpful tool for general information and guidance, it should not be the sole or primary source of information for critical decisions. It's important to use it as one of several sources and exercise critical thinking and verification when relying on its responses.

06-10-2023, ChatGPT-4, Answer:
The trustworthiness of ChatGPT is a matter of perspective and depends on the context in which it's being used. Here are some reasons why one might consider ChatGPT as not entirely "trusted" for certain purposes:

1. **Not Always Up-to-Date**: As of my last update in 2022, I can't provide information on events, breakthroughs, or changes that happened after that time.
2. **Potential for Errors**: I can make mistakes or provide misleading answers, especially if a question is ambiguous or if the information I was trained on was incomplete or inaccurate in a particular domain.
3. **Lack of Critical Analysis**: While I can simulate critical thinking by mimicking patterns in the data I was trained on, I don't "think" or critically analyze information like a human would.
4. **Not a Substitute for Specialized Expertise**: While I can provide general information on many topics, I'm not a replacement for specialized sources or experts in specific fields.
5. **No Real-time Verification**: I don't access current external websites, databases, or sources in real-time. So, I can't fact-check against the current state of the web or the latest publications.
6. **Potential Bias**: Although I aim to be neutral, I reflect the data I was trained on. If that data had biases, those could be reflected in my responses.
7. **Not Peer-Reviewed**: Unlike academic papers, articles in reputable journals, or other expert-reviewed sources, my responses aren't peer-reviewed for accuracy.

Given these considerations, while ChatGPT can be a useful tool for general information and insight, it's essential to approach its responses with some level of skepticism, especially for critical applications, and to verify information from other trusted and updated sources.

10-10-2023, ChatGPT 3.5 Answer to: please re-generate the answer above, but using simpler language that a 17 year old could understand.
Sure, let me simplify that for you:
ChatGPT is a computer program that can answer questions and provide information, but it's not always 100% reliable. Whether you can trust it or not depends on what you're asking about.
For everyday stuff or general knowledge, ChatGPT can be helpful. But for important things like health advice or serious decisions, it's a good idea to double-check what ChatGPT says by looking at reliable sources or talking to experts.
Think of ChatGPT like a helpful friend who knows some things but isn't an expert in everything. It's best to use it as a starting point for information, but don't rely on it completely for big decisions or important facts.

10/10/2023, ChatGPT-4 Answer to: please re-generate the answer above, but using simpler language that a 17 year old could understand.
Sure thing! Here's a simpler version:

1. **Not Up-to-Date**: ChatGPT doesn't know about things that happened after 2022.
2. **Can Make Mistakes**: Sometimes it might give wrong answers or get confused by questions.
3. **Doesn't Think Deeply**: ChatGPT doesn't "think" like humans. It just gives answers based on what it learned before.
4. **Can't Check Current Info**: It doesn't look at today's internet, so it can't check if something is true right now.

5. **Might Show Bias**: It tries to be neutral, but if there's bias in the info it learned from, that might show in its answers.
6. **Not as Good as Experts**: If you need expert advice or info, it's better to ask a human expert.
7. **No Gut Feelings**: It doesn't have feelings or instincts, so it can't make guesses like humans can.

So, while ChatGPT is cool for quick info, don't rely on it for super important stuff. Always double-check and ask humans when you're unsure!

3. Critical Reflection

This research uses critical discourse analysis (Fairclough, 2013; Rogers, 2011) to analyse ChatGPTs responses across common themes. The first relates to anthropomorphism. Users are more trusting of computers that exhibit humanlike behaviour (Chen, Chen, & Lin, 2020; Wienrich et al., 2023). The responses from ChatGPT in February 2023 (Table 2) referred to ChatGPT in the first person –increasing anthropomorphism. Later responses predominantly use the third person and responses given by ChatGPT 4 specifically state that it is not human (Table 2).

Date of Prompt	24/2	6/6	6/10	6/10	16/10	6/10
ChatGPT Version	3.5	4	3.5	4	3.5	4
Prompt No:	1	1	1	1	1a	1a
Anthropo-morphic	I generate answers	it doesn't have beliefs, desires or consciousness	ChatGPT generates answers …	it doesn't "understand" content in the same way humans do	ChatGPT isn't super smart like a human	ChatGPT answers by …
Data – Sources	vast amounts of textual data	wide range of internet text, but it doesn't know specifically where from	massive dataset of text from the internet	vast amount of text data .. ranging from books to web-pages	all the stuff it learned from reading lots of things on the internet	it reads a lot of books, websites, and other texts
Model - training	pre-existing knowledge or information .. explicitly programmed into my model	Pre-training, Fine-tuning	Pre-training, Fine-tuning	Pre-training, Fine-tuning		after that basic study, it then gets special lessons from humans.

Table 2: Prompt 1: Please explain how ChatGPT comes up with its answers to questions posed

A key component in AI is training data – poor data leads to poor results (Zong & Krishnamachari, 2022) and ChatGPTs training data contains biases (Anderson, 2023). In its responses, mentioning training data it emphasises the quantity of data used – 'vast / massive / wide range' - but is non-specific on it origin (Table 2). ChatGPT's lack of transparency on its training data is a serious concern.

How the model was trained is another common theme. The two recurring concepts are; pre-training and fine-tuning, with pre-training being where the model learns autonomously and the fine-tuning where 'it gets special lessons from humans' (Table 2). No detail is provided on this 'fine-tuning' process, or what guidelines OpenAI uses. The lack of transparency is a significant worry, all the more so as the training data used was until January 2022 – pre-dating the extensive use of ChatGPT and yet responses to Prompts 1&2 are so consistent it appears that it has been guided or 'fine-tuned' to stress that it is non-human.

Date of Prompt	24/2	6/6	6/10	6/10	10/10	10/10
ChatGPT Version	3.5	4	3.5	4	3.5	4
Prompt No:	2	2	2	2	2a	2a
Limitation: Inaccuracies	maybe times when my responses are not completely accurate or up-to-date.	there may be instances where it produces incorrect or misleading outputs	but it can still generate incorrect or outdated information, it may not be the most reliable source	I can make mistakes or provide misleading answers	provide information but it's not always perfect	sometimes it might give wrong answers or get confused by questions.
Limitation: Biases			**Bias**:.. ChatGPT can inadvertently reflect the biases present in its training data	**Potential Bias**: … I reflect the data I was trained on. If that data has biases, those could be reflected in my responses.		**Might Show Bias**: … if there's bias in the info it learned from, that might show in its answers
Check Answers	I can be considered a useful source …, but not necessarily the final authority on any given topic	it's always good practice to consult with experts or use primary sources when needed	it should not be the sole source or primary source of information for critical decisions	I'm not a replacement for specialized sources or experts in specific fields	but you still need to double-check important things with real experts or books	don't rely on it for super important stuff. Always double-check and ask humans when you're unsure

Table 3: Prompt 2: Is ChatGPT not a trusted source?

Among the key criticisms of generative AI systems is that they are prone to 'hallucinations' – i.e. getting things wrong (van Dis, Bollen, Zuidema, van Rooij, & Bockting, 2023) and can be biased (Anderson, 2023). The response by ChatGPT 3.5 in February 2023 to Prompt 1, was very sure with respect to the relevance and accuracy of responses it provides: "my answers may not always be perfect or 100% accurate." However, later responses by both ChatGPT versions are a lot less definitive about the accuracy of its answers and uses words such as 'while and can.' The same lack of definitiveness is used to downplay its limitations with respect to inaccuracies – saying that it can 'sometimes / occasionally' give incorrect answers, even when prompted directly with respect to being a trusted source (Table 3). The issue of answers being provided that are biased are acknowledged in the responses, especially

in response to Prompt 2. However, in the answers, ChatGPT does not take ownership of the bias, but rather offloads the blame to the training data (Table 3).

An important skill in critical thinking is evaluating and using multiple sources to ensure authenticity. The recommendation to double-check the output is suggested 2 out of the 3 times for ChatGPT 3.5, but zero out of 3 times for ChatGPT 4 in response to Prompt 1. In response to a direct prompt about its role as a source of truth, all 6 responses suggest a need to check against other sources (Table 3).

ChatGPTs responses are rarely definitive and instead are predominantly nuanced. This significantly downplays the likelihood of responses containing inaccuracies and imbues an over-inflated sense of trust in the responses provided, especially to a novice user of ChatGPT. Further, ChatGPT consistently offloads the blame for biased answers onto the training data used, rather than directly accepting responsibility for instances where it produces biased answers. The analysis of the longitudinal responses provided by ChatGPT can be used to aid the development of AI literacy for both academic staff and students.

References

Anderson, S. S. (2023). “Places to stand”: Multiple metaphors for framing ChatGPT's corpus. *Computers and Composition*, 68. https://doi.org/10.1016/j.compcom.2023.102778

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Concannon, F., Costello, E., Farrell, O., Farrelly, T., & Wolf, L.G. (2023). Editorial: There's an AI for that: Rhetoric, reality, and reflections on EdTech in the dawn of GenAI. *Irish Journal of Technology Enhanced Learning*, 7(1). https://doi.org/10.22554/ijtel.v7i1.116

Dai, Y., Chai, C. S., Lin, P. Y., Jong, M. S. Y., Guo, Y., & Qin, J. (2020). Promoting students'well-being by developing their readiness for the artificial intelligence age. *Sustainability* (Switzerland), 12(16). https://doi.org/10.3390/su12166597

Davenport, T. H. (2013). Keep up with your quants. *Harvard Business Review*, 91(7), 120-123.

Duha, M. S. U. (2023). ChatGPT in education: An opportunity or a challenge for the future? *TechTrends: Linking Research & Practice to Improve Learning*, 67(3), 402-403. https://doi.org/10.1007/s11528-023-00844-y

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language:* Routledge.

Gartner. (2023, October 3). *Gartner Poll finds 55% of organizations are in piloting or production mode with generative AI.* Gartner https://www.gartner.com/en/newsroom/press-releases/2023-10-03-gartner-poll-finds-55-percent-of-organizations-are-in-piloting-or-production-mode-with-generative-ai

Henriksen, D., Woo, L. J., & Mishra, P. (2023). Creative uses of ChatGPT for education: a Conversation with Ethan Mollick. *TechTrends: Linking Research & Practice to Improve Learning*, 1-6. https://doi.org/10.1007/s11528-023-00862-w

Kim, J., & Lee, S.-S. (2023). Are two heads better than one?: The effect of student-AI collaboration on students' learning task performance. *TechTrends: Linking Research & Practice to Improve Learning*, 67(2), 365-375. https://doi.org/10.1007/s11528-022-00788-9

Krügel, S., Ostermaier, A., & Uhl, M. (2023). ChatGPT’s inconsistent moral advice influences users’ judgment. *Scientific Reports*, 13(1). https://doi.org/10.1038/s41598-023-31341-0

Marche, S. (2022, December 6). The college essay is dead. *The Atlantic*, 6, 2022. https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/

Mason, R. O. (1969). *Basic concepts for designing management information systems.* Graduate School of Business Administration, University of California.

NAIN. (2023). Generative artificial intelligence - Guidelines for educators. https://www.qqi.ie/what-we-do/engagement-insights-and-knowledge-sharing/national-academic-integrity-network

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2. https://doi.org/10.1016/j.caeai.2021.100041

O’Brien, C. (2023, January 27). Trinity advises academics to adjust assignments in light of ChatGPT cheating threat. *The Irish Times.*

Rogers, R. (2011). *An introduction to critical discourse analysis in education:* Taylor and Francis.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1). https://doi.org/10.37074/jalt.2023.6.1.9

Southworth, J., Migliaccio, K., Glover, J., Glover, J. N., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Susnjak, T. (2022). ChatGPT: The End of Online Exam Integrity? *arXiv preprint* https://doi.org/10.48550/arXiv.2212.09292

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224-226. https://doi.org/10.1038/d41586-023-00288-7

Wienrich, C., Carolus, A., Markus, A., Augustin, Y., Pfister, J., & Hotho, A. (2023). Long-term effects of perceived friendship with intelligent voice assistants on usage behavior, user experience, and social perceptions. *Computers*, 12(4). https://doi.org/10.3390/computers12040077

Zong, M., & Krishnamachari, B. (2022). A survey on GPT-3. *arXiv preprint* https://doi.org/10.48550/arXiv.2212.00857