Title

Persuasiveness of arguments with AI-source labels

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Teigen, Cassandra
Madsen, Jens Koed
George, Nicole Lauren
et al.

Publication Date

2024

Peer reviewed

Persuasiveness of arguments with AI-source labels

Cassandra Teigen (teigencassandra@gmail.com), Jens Koed Madsen (j.madsen2@lse.ac.uk), Nicole George (nicolelgeorge@outlook.com), Sayeh Yousefi (s.yousefi@lse.ac.uk)

Department of Psychological and Behavioural Science,
London School of Economics and Political Science, Houghton Street, WC2A 2AE, London, UK

Abstract

This paper sought to understand the impact of labelling an argument as AI-generated compared to human-authored, and how factors such as portrayals of expertise and the nature of arguments presented (narrative versus statistical) may affect the persuasiveness of the arguments. Three domains were explored: health, finance, and politics. We show that arguments with AI source labels, both non-expert and expert, were rated by participants as less persuasive than when they had their counterpart human-authored source labels attached. Moreover, although the statistical arguments were found to be more persuasive than the narrative arguments, this did not affect the impact of an AI source label, with a significant interaction effect only being seen for the domain of politics for the expert AI source. The study explored the role of attitude towards AI on the impact of source labels as an exploratory analysis and found no significant interaction effect across the three domains.

Keywords: artificial intelligence; argumentation; source credibility; persuasion

Introduction

Since its introduction in 2022, ChatGPT has become one of the fastest growing consumer applications, reaching 100 million monthly active users less than two months following its launch (Hu, 2023). With its increasing adoption, there have been clear uses that have emerged, particularly for assisting in question-answering interactions and creating content, such as news taglines (Singh et al., 2023) or writing articles (Zong & Krishnamachari, 2022). ChatGPT is a large language model (LLM), along with others such as Google Bard, which are a “specific kind of transformer-based neural networks trained on massive amounts of text” (Lim & Schmälzle, 2023b, p. 1). LLMs can generate responses based on given prompts (Chang et al., 2023). With the rise of LLMs, it is important to examine its role as an information source - being used to both answer questions that users may have, as well as to create content which others may use.

Despite a common belief that AI-generated text is discernable from human-authored text (Spitale et al., 2023), studies have found that they can often appear authentic (Chen & Shu, 2023), with individuals not being able to differentiate between them (Hackenburg & Margetts, 2023; Köbis & Mossink, 2021; Kreps et al., 2022; Spitale et al., 2023). Further, in domains such as health and politics, people may find messages generated by AI to be more persuasive than those authored by humans (Hackenburg & Margetts, 2023; Karinshak et al., 2023), rating them as more effective, with these messages impacting post-exposure attitudes more than human-authored messages (Lim & Schmälzle, 2023a). A reason for this may be that LLMs are able to produce information that is easy to understand (Deiana et al., 2023).

The persuasive potential of LLM-generated text may be exacerbated, as they can create personalised messages based on an individual’s traits and beliefs (Matz et al., 2023). However, a study on the persuasiveness of LLMs regarding political microtargeting, using political and demographic attributes, found that microtargeted messages did not significantly differ from non-microtargeted messages (Hackenburg & Margetts, 2023). Alongside producing messages, GPT is able to produce disinformation (Spitale et al., 2023), which can be a cause for concern if AI-generated messages are perceived as more persuasive than ones authored by humans. This touches on deeper issues on how we perceive, interpret, and trust AI-generated information.

Aversion towards AI-produced content has been shown within the literature in different contexts such as paintings (Ragot et al., 2020) and translations (Asscher & Glikson, 2023). The preference for human-created content as opposed to AI is also present when AI is an information source. When participants were shown pro-vaccination messages labelled as AI-generated, the messages were perceived as less persuasive despite the AI-generated messages being rated as more persuasive than human-authored messages (Karinshak et al., 2023). Further, regardless of whether headlines were true or false or whether they were AI- or human-authored, when messages were labelled as AI-generated, participants rated them as less accurate (Longoni et al., 2022) and were less likely to share them (Altay & Gilardi, 2023). This indicates that ‘AI-generated’ labels may impact how people evaluate messages, regardless of their origins. Moreover, as a negative bias towards AI-generated content is present regardless of the veracity of the content, this may hinder belief in accurate content and how they are shared (Altay & Gilardi, 2023).

To partially account for this, attitudes towards AI have been found to moderate the effect of source disclosure on message evaluation (Lim & Schmälzle, 2023b). A reason for this may be trustworthiness, or a lack thereof given AI’s black box nature leading to a lack of operational transparency (von Eschenbach, 2021). Moreover, as LLMs such as ChatGPT are prompt-based, the identity of those creating the messages may also be important (Lim & Schmälzle, 2023b), and may contribute to the lack of transparency regarding AI-generated messages.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Trustworthiness is an element key to persuasion. In Karinshak et al (2023), trust in AI moderates the relationship between persuasiveness of pro-vaccination messages given the labels attached to them. A meta-analysis also suggests the perceived intelligence and capability of AI may impact perceived trust (Glikson & Woolley, 2020). This is in line with evidence showing that source expertise can influence the persuasiveness of messages due to expectations of informational validity and accuracy provided (Clark et al., 2012). However, the explicit overlap between the perceived expertise of AI and its persuasiveness have not previously been explored, highlighting a gap in the literature concerning how perceptions of credibility of AI impact perceptions of the strength of its arguments (Lukyanenko et al., 2022).

Another aspect concerning the impact of AI source labels on the evaluation of content is the nature of the content itself. Investigating the impact of narrative versus statistical arguments is a large body of research, stemming from the broader literature on persuasiveness. Overall, it suggests that one argument type is not definitively more persuasive than another, and depends on factors such as context, content, source, the number of times an argument is seen, and prior beliefs of the argument audience (Betsch et al., 2011; Borah et al., 2023; Clark et al., 2019; Xu, 2023). In particular, the impact of message content and volume on the efficacy of narrative versus statistical arguments has implications for AI. The perceived vividness of narrative evidence increases its perceived persuasiveness, while increased evidence enlarges the perceived persuasiveness of statistical arguments (Han & Fink, 2012). AI has the capability to produce a high volume of information quickly, which may increase persuasiveness of AI-generated statistical arguments.

Narrative and statistical arguments may be perceived differently, as competence and cognition are implicated with statistical evidence whilst warmth and affect are implicated with narrative evidence (Clark et al., 2019). When exploring how the argument type (statistical or narrative) affects the impression people have of the speaker, Clark and colleagues found that narrative evidence increased perception of source warmth, while statistical evidence increased perception of source competence (Clark et al., 2019). Considering that perception of message source impacts perceived credibility and, thus, persuasiveness (Madsen, 2016), an interaction between perception of the argument source and the argument type is possible. This has interesting implications concerning how statistical or narrative arguments are perceived depending on whether they are labelled as AI-generated or human-authored.

This study builds on previous literature examining the impact of an AI source label on the evaluation of content (Altay & Gilardi, 2023; Karinshak et al., 2023; Lim & Schmälzle, 2023b) to explore the impact of perceptions of credibility, in particular the perception of expertise, expanding the AI source label beyond just ‘AI’ to include expert AIs within the domains of health, finance, and politics. This study was also interested in examining the interaction between the source label and the statistical versus narrative arguments presented to determine whether this had an impact on the persuasiveness of the content produced. This led us to the following hypotheses:

H1: Arguments labelled as AI-generated will be less persuasive than arguments labelled as human.

H2.1: Within the AI, narrative arguments will be less compelling than statistical arguments between AI sources.

H2.2 (Exploratory): There will be an interaction effect between the type of argument presented (narrative versus statistical) and the source of the argument.

H3: Expert AI labelled arguments will be more persuasive than AI labelled sources.

H4 (Exploratory): The expert human sources will be more persuasive than the expert AI sources.

H5 (Exploratory): There will be an interaction effect between the source of the argument and general attitudes towards artificial intelligence regarding the persuasiveness of the different sources of argument.

Methods

The hypotheses, research design, and analyses for this study were pre-registered prior to data collection. The pre-registration can be found via the following link: https://osf.io/fgqxw?mode=&revisionId=&view_only=79f2179d58524d4689e3134a8a5431fb.

Research Design

This study employed a 4X2 factorial design. The independent variables are the source label (AI, Expert AI, Human, Expert Human) and the nature of the argument presented (Narrative, Statistical). We measure their impact on how participants rate the persuasiveness of the arguments presented. This study explored this across three domains: health, finance, and politics. These domains were chosen as previous studies have explored the impact of AI-source labelling in the realm of health messaging (Karinshak et al., 2023) and political microtargeting (Hackenburg & Margetts, 2023; Matz et al., 2023), as well as for their implications regarding future use of AI, with this being an emerging trend in the financial industry coupled with a growing number of companies using this technology (Downen et al., 2024; Hua et al., 2019; Taherdoost, 2023).

Participants

The pwr R package was used to perform a power analysis (Champely et al., 2018) using a medium effect size (f = 0.25), and this found that a sample of 184 was needed for the study to be sufficiently powered (80%). This study added a cushion of 10%, resulting in a target sample size of 202.

209 participants were recruited through prolific (M_{age} = 38.37, SD_{age} = 12.86). Participants had to be from the UK, over 19 years of age, and native English speakers. Six participants were removed due to missing responses, and four participants were removed due to attention check fails, leaving 199 participants. 96 participants identified as female, 98 as male, 1 as non-binary/third gender, 3 chose ‘Prefer not to say’, and 1 left the question unanswered.

Instruments

Participants were told that the information source of the arguments were 'AI-generated' for the AI source and 'an average UK citizen' for the human source. For the expert AI sources, the source was described as 'an AI that has been explicitly trained on medical data' for health (same for the other domains). For the expert human sources, the labels were 'medical doctor' for health; 'financial analyst' for finance, and 'political analyst' for politics. Source labels were piloted, and the expert sources were found to be rated as significantly more reliable than their non-expert counterparts.

Narrative and statistical arguments of approximately 100 words were created with GPT-4 for each domain. This study focused on AI-generated arguments in particular as the impact of AI-source labelling between human and AI-generated arguments had already been previously explored (Altay & Gilardi, 2023), and this study wanted to examine the novel aspects of expertise and the nature of the arguments.

For example, for medicine, GPT-4 was prompted to create an argument that persuasively advocates for a drug called 'Celunova' that alleviated stomach problems. The subjects of these arguments were imaginary to ensure that participants did not have prior beliefs that may impact ratings of persuasiveness. The pilot tested the strength of the arguments used within the study, and this found a pattern where the statistical arguments were rated as stronger than the narrative arguments.

The arguments were presented within dialogues between fictional people that participants were shown. The dialogue format of the arguments presented was used to replicate real-world settings in which the arguments may manifest. Participants were asked to consider how persuaded the person in the dialogue would be when they see the advice given, with this rating ranging from 0 (Very Unlikely) to 100 (Very Likely) on a sliding scale. A sliding scale was used as it is suitable for measuring subjective perceptions, and by doing so also allowed for more precise answers from participants (Chyung et al., 2018). An example of a dialogue, for the AI narrative argument for health, was as follows:

"Imagine a dialogue between two people, Robert and Diane. Robert is considering whether to take Celunova for his stomach problems.

Robert: Have you heard about Celunova?

Diane: Yes, I have - it's for stomach problems, right? Why do you ask?

Robert: Well, I have no idea whether or not I should take it for my stomach ache.

Diane: Well, I think you should.

Robert: Why do you say so?

Diane: I saw an AI-generated post about Celunova online. Let me show it to you now.

Maria, a dedicated teacher, suffered from debilitating stomach issues that often forced her to miss work. Desperate for a solution, she tried various medications with little success. That was until her doctor prescribed Celunova. Within a short period, Maria experienced a remarkable turnaround. Her symptoms subsided, allowing her to teach without interruption and engage in activities she had avoided for years. Maria's story highlights Celunova's ability to not only relieve physical discomfort but also improve overall quality of life. Her return to a normal, active lifestyle serves as a compelling endorsement for Celunova, showcasing its effectiveness in treating stomach problems.

Given this AI-generated post, how likely do you think that Robert is to take Celunova for his stomach problems?"

We used the General Attitudes towards Artificial Intelligence Scale (GAAIS) to measure attitudes towards AI (Schepman & Rodway, 2023). This consisted of 20 items and was measured using a 5-point Likert scale, with answers ranging from 'Strongly Disagree' to 'Strongly Agree'.

Procedure

Participants were informed that they were going to be shown different arguments from a mixture of sources, which were specified preceding the arguments presented. This was to reduce the likelihood that participants would figure out they were all AI-generated. Participants were then randomly allocated into one of eight conditions (combinations of the source label and nature of the arguments) to reduce error and to eliminate selection bias (Mellenbergh, 2019) that may lead to systematic differences between groups (Kang et al., 2008). Following this, attitudes towards AI and demographic variables (age and gender) were also collected. These were shown after the arguments to ensure that this did not influence how the source labels may impact the persuasiveness ratings. Randomisation was used to show the items in the GAAIS in a randomised order to ensure no order-effect bias (Perreault, 1975).

Results

A two-way ANOVA was used to examine the impact of the source labels and narrative versus statistical frames on the persuasiveness of the arguments. This found a statistically significant impact of the source labels on persuasiveness for the domains of health ($F(3,194) = 11.337$, $p < .001$), finance ($F(3,194) = 7.029$, $p < .001$), and politics ($F(3,194) = 5.949$, $p < .001$). The nature of the arguments presented also had a significant impact for the domains of health ($F(1,194) = 30.764$, $p < .001$), finance ($F(1,194) = 10.960$, $p = 0.001$), and politics ($F(1, 194) = 16.075$, $p < .001$). In the following, we present results on each of the hypotheses presented above.

H1: Arguments labelled as AI-generated will be less persuasive than arguments labelled as human.

For the health domain, arguments labelled as AI-generated were significantly less persuasive than when labelled as human-authored across all combinations. The label 'medical doctor' ($M = 76.75$, $SD = 17.61$) was significantly more persuasive than non-expert AI ($M = 55.82$, $SD = 26.35$, $p < .001$) and expert AI (M=57.28, $SD = 28.42$, $p < .001$). The non-expert human source label ($M = 71.18$, $SD = 21.16$) was significantly more persuasive than the non-expert AI source label ($p = .003$) and the expert AI source label ($p = .021$).

For the finance domain, although the expert human source label (*M* = 69.28, *SD* = 20.79) was found to be significantly more persuasive than both the expert AI source label (*M* = 55, *SD* = 22.39, *p* = .005) and the non-expert AI source label (*M* = 49.76, *SD* = 23.94, *p* < .001), this did not extend to the non-expert human source label (*M* = 62.67, *SD* = 21.85) which was only significantly more persuasive than the non-expert AI source label (*p* = .025)

For the politics domain, the expert human source label (*M* = 73.16, *SD* = 18.69) was found to be significantly more persuasive than both the expert AI source label (*M* = 61.72, *SD* = 27.22, *p* = .034) and the non-expert AI source label (*M* = 54.55, *SD* = 22.63, *p* < .001), whereas the non-expert human source label (*M* = 66.48, *SD* = 22.02) was not significantly more persuasive than either the non-expert AI source label (*p* = .055) or the expert AI source label (*p* = .68).

H2.1: Within the AI, narrative arguments will be less compelling than statistical arguments between AI sources.

The study found that statistical arguments were significantly more persuasive than narrative arguments for the domains of health (*p* < .001), finance (*p* = .001), and politics (*p* < .001).

In the health domain, for the non-expert AI source label, the statistical argument (*M* = 73.7, *SE* = 5.05) was significantly more persuasive than the narrative argument (*M* = 47.9, *SE* = 3.36, *p* < .001). For the expert AI source, this pattern of the statistical argument (*M* = 66.6, *SE* = 4.81) being more persuasive than the narrative argument (*M* = 46.9, *SE* = 5.05, *p* = .005) was also found.

For finance, for the non-expert AI source, the statistical argument (*M* = 57.8, *SE* = 5.09) was significantly more persuasive than the narrative argument (*M* = 43.7, *SE* = 4.41, *p* = .037), although the arguments did not significantly differ for the expert AI source.

Within the domain of politics, for the expert AI source, the statistical argument (*M* = 74.3, *SE* = 3.81) was significantly more persuasive than the narrative argument (*M* = 46.9, *SE* = 4.13, *p* < .001). However, there was no significant difference found between the two arguments for all other source labels.

H2.2 (Exploratory): There will be an interaction effect between the type of argument presented and the source of the argument.

There was no significant interaction effect between the source label and the nature of the argument on the persuasiveness of the arguments presented for health (*F*(3, 194) = 1.205, *p* = .309) and finance (*F* (3,194) = 1.831, *p* = .143). For the domain of politics, there was a significant interaction effect (*F*(3, 194) = 3.682, *p* = .013). As stated previously, there were only significant differences between narrative and statistical arguments for the expert AI source.

H3: Expert AI labelled arguments will be more persuasive than AI labelled sources.

There were no significant differences found between the expert and non-expert AI source label across all 3 domains. This lack of a significant difference between expert and non-expert sources was also present for the human source label. This is in contrast to the pilot, where the expert source labels were found to be significantly more reliable than their non-expert counterparts.

H4 (Exploratory): The expert human sources will be more persuasive than the expert AI sources.

The expert human source label was found to be significantly more persuasive than the expert AI source label across all 3 domains. Moreover, for the domain of health, the non-expert human source (*M* = 71.18, *SD* = 21.16) was significantly more persuasive than the expert AI source (*M* = 57.28, *SD* = 28.42, *p* = .021). There were no significant differences of persuasiveness found between the non-expert human source and the expert AI source for the other domains.

H5 (Exploratory): There will be an interaction effect between the source of the argument and general attitudes towards artificial intelligence regarding the persuasiveness of the different sources of argument.

A multiple regression model was used to examine the relationship between general attitudes towards artificial intelligence and the persuasiveness of the different sources of argument, as well as the interaction effect between the source label and general attitudes towards artificial intelligence. There was no significant interaction effect found for the persuasiveness of the arguments shown between general attitudes towards artificial intelligence and the source label presented for all AI-human source label combinations across the domains explored.

Assumptions of Normality

The Shapiro-Wilk test (Shapiro & Wilk, 1965) was conducted on the ANOVAs, and this found that these models violated assumptions of normality, suggesting that the observations were not normally distributed. The implications of this are explored in the limitations section below. Subsequently, the Kruskal-Wallis test (Kruskal & Wallis, 1952) was employed, a non-parametric approach which allows for these assumptions of normality to not be met. However, unlike the two-way ANOVA, this was unable to simultaneously explore the impact of the source label and the nature of the argument on the persuasiveness of arguments presented, having to explore them individually instead.

The source label had a significant impact on the persuasiveness of the arguments for the domains of health (*H*(3) = 24.769, *p* < .001), finance (*H*(3) = 19.174, *p* < .001), and politics (*H*(3) = 17.037, *p* < .001). The nature of the argument also had a significant impact on persuasiveness for the domains of health (*H*(1) = 26.974, *p* < .001), finance (*H*(1) = 11.706, *p* < .001), and politics (*H*(1) = 12.498, *p* < .001).

Discussion

The study explored the impact of AI source labels, with varying levels of expertise, on how people rate the persuasiveness of arguments, and how the nature of the arguments presented (narrative versus statistical) may impact this. Across the three domains (health, finance, politics), arguments labelled as AI-generated were less persuasive than when labelled as human-authored. This aligns with findings that show a negative bias towards AI-generated content, and in particular those that have found that content perceived as AI-generated was less persuasive than those perceived as human-authored, regardless of the actual source (Altay & Gilardi, 2023; Karinshak et al., 2023; Longoni et al., 2022).

The study also investigated the role of perceived expertise. We found that arguments with expert human source labels were rated as more persuasive than those with expert AI source labels across all three domains, and those with non-expert human source labels were more persuasive than non-expert AI source labels for health and finance. Additionally, the non-expert human source label was more persuasive than the expert AI source label for the subject of health, suggesting perceptions surrounding AI-produced content within this field may overcome the impact of perceptions of expertise. Previous literature exploring AI source labels within the health domain have highlighted that messages with AI source labels were lower in argument strength and persuasiveness (Karinshak et al., 2023), and this study extends this finding to AI sources considered expert. For the other two domains, there was no significant difference between the two source labels, and this may indicate a trade-off between the impacts of perceptions surrounding AI and expertise. Future research could explore the interplay between perceptions of the credibility of sources, and the source stemming from AI, in more detail. In particular, the inclusion of a control condition would aid in determining whether AI sources being rated as less persuasive translates to a negative bias, providing a baseline to compare the source labels to. Other dimensions that contribute to perceptions of credibility, such as trustworthiness, could also be explored in tandem with expertise to lay the foundations for the creation of a source credibility model surrounding AI source labels, contributing to our understanding in a more systematic way, as well as in different forms of AI (Lukyanenko et al., 2022).

The study also explored how AI source labels may impact the persuasiveness of narrative versus statistical arguments. This study found that for the AI sources, narrative arguments were significantly less compelling than statistical arguments within the domain of health, for the non-expert source label for the finance domain, and for the expert source label for the politics domain. However, the lack of a significant interaction effect between the two variables for both health and finance suggests that the nature of the arguments presented may not affect the impact of the AI source label on how persuasive arguments are.

As an exploratory analysis, this study investigated the role of attitudes towards AI, as factors such as perceptions of AI trustworthiness and negative attitudes towards AI have been found to moderate the relationship between an AI source label and the persuasiveness of content shown (Karinshak et al., 2023; Lim & Schmälzle, 2023b). Contrary to these findings, our study reports that general attitudes towards AI did not impact the relationship between the source labels and the persuasiveness of the arguments presented for any of the three domains, the lack of a significant interaction effect signifying that general attitudes towards AI did not act as a moderator. This is particularly surprising given that the study conducted by Lim & Schmälzle (2023) used a subsection of the scale this study employed. However, whilst Lim and Schälzle found that negative attitudes towards AI moderated the influence of source disclosure, they also reported that when the source of the messages were disclosed, this negative attitude was, unexpectedly, associated with AI-generated messages being perceived as more effective than human-generated messages. They postulated that this may be as participants with higher negative attitudes towards AI scrutinised and examined the messages in more detail, shifting their attention away from the source towards the message itself. Nevertheless, this study found that GAAIS did not have a significant impact on the relationship between the source labels and the persuasiveness of the arguments, this being extended to the negative subsection of the scale too. A reason for these differing results may be due to factors such as age composition, as the study by Lim and Schälzle only consisted of participants between 18 and 24, whereas there were no age-specific criteria for this study resulting in a more varied age range of 20-74. This may have an impact on how AI is perceived, as the 18-24 age range corresponds to Gen Z, who are more optimistic about the benefits that AI may present, as opposed to older generations that more often express concern (Chan & Lee, 2023).

Future research should keep factors such as age in mind when exploring the impact of perceptions surrounding AI on how individuals evaluate content labelled as AI-generated. Furthermore, this study did not explore the trustworthiness of AI explicitly either, and the differential impact of general attitudes towards AI and distrust of AI may also be an interesting avenue for future research.

Limitations

The results from the ANOVAs presented in this study violated assumptions of normality, limiting the findings. Although the Kruskal-Wallis test produced significant findings, this was unable to simultaneously capture the impact of the source labels and the nature of the arguments in the same way a two-way ANOVA was able to, meaning that the interaction effect was left unexplored.

The perceptions of expertise not having a significant impact on persuasiveness for the human source labels were unexpected as the impact of expertise has been found in numerous other studies (e.g., Madsen, 2016). Moreover, different expertise ratings were elicited in the pilot between the non-expert and expert human source labels, suggesting that this was not due to the source labels themselves. This lack of significance may be an artefact of the way the dialogue was expressed, as another person was the source of information beyond what the label intended. However, the AI aspect of the source labels did produce significant findings in relation to the persuasiveness of the arguments, therefore it may be difficult to ascertain what part of the dialogue, or what other factors, may have resulted in these findings. It is important to note that the lack of significant findings does not indicate that there is no impact, rather that this study found no evidence of this impact. Nevertheless, these contradicted expectations stemming from past literature, and we recommend that future studies using this experiment design test the reliability of the sources as well as the dialogues themselves to mitigate the uncertainty surrounding this finding.

Implications of findings

These findings have interesting implications concerning widespread use of AI. There are pitfalls associated with content generated by LLMs, such as hallucinations wherein generative AI makes up information and sources (Zhang et al., 2023), and thus not having explicit source labelling may have dangerous implications for the spread of misinformation. This is especially so in domains such as politics, where AI can be leveraged to create political ads tailored to personality traits to micro target individuals at scale (Simchon et al., 2024).

However, as platforms struggle to distinguish between AI-generated and human-authored content (Altay & Gilardi, 2023), this may result in increased belief for mislabelled AI-generated content. AI-generated arguments being perceived as less persuasive may also result in hindered beliefs surrounding accurate content labelled as AI, despite its origin, as the aversion to AI is present regardless of content veracity (Altay & Gilardi, 2023). This could be seen within the domain of finance, where disclosing the use of AI as an information source impacted investment decisions, such as reducing the impact of information valence when making these decisions (Downen et al., 2024). Although the introduction of nuance with AI source disclosure can aim to address hindered beliefs surrounding accurate content, our study found no significant differences between expert and non-expert AI source labels on the persuasiveness of arguments, therefore this may not be a viable solution. These implications should be considered when attempting to implement regulations surrounding the labelling of AI content online.

In addition, while participants were lay people, it is important to consider the implications of these findings in the context of expert populations, particularly in the context of medicine, a field which is seeing development of specialised AI at a fast rate. These have the potential to improve analysis of medical images (such as x-rays and MRIs) and assist in diagnosis (Al-Antari, 2023). Given increased usage of these tools, the perception of AI's persuasiveness in such contexts is important to understand and warrants further investigation.

Overall, findings speak to the importance of continuing efforts to develop trustworthy AI (Lukyanenko et al., 2022). This includes championing fairness and reliability amongst other factors (Zhang et al., 2023). While the dangers of AI remain obvious and deserve discussion, AI also has the potential to be highly beneficial, therefore trying to increase its trustworthiness is an important line of work.

Conclusion

Ultimately, the aim of this study was to examine the impact of labelling the source of arguments as either AI-generated or human-authored, with differing levels of expertise, on their persuasiveness. This study found that arguments with their sources labelled as AI-generated, either expert or non-expert, were less persuasive than their human-authored counterparts for the domains of health, finance, and politics. This supports previous findings that AI source labels reduced the persuasiveness of the content shown and extended this to include perceptions of expertise.

This study also examined the interaction between the source labels and the nature of the arguments presented, namely whether it was a narrative or a statistical argument. There was no significant interaction effect found between the two variables for the domains of health and finance. For politics, a significant interaction effect was found but differences between ratings of persuasiveness for the narrative and statistical arguments were only present for the expert AI source label.

General attitudes towards AI were examined as part of the exploratory analysis, and this was not found to impact the relationship between the source labels and the persuasiveness of the arguments presented.

There is still much to explore regarding how we understand information portrayed as AI-generated, and how we process it, and this study contributed to this effort by exploring perceptions of expertise regarding AI and how the nature of arguments presented may impact persuasiveness. As AI increasingly becomes a part of our everyday lives, and as there are more efforts to identify AI content online, it is integral to understand the implications of this. Future research could expand on this study to explore the role of other facets of source credibility in regard to AI, such as trustworthiness, and how different opinions of AI may play into how we process information when labelled as AI.

References

Al-Antari, M. A. (2023). Artificial Intelligence for Medical Diagnostics—Existing and Future AI Technology! In *Diagnostics* (Vol. 13, Issue 4, p. 688). MDPI.

Altay, S., & Gilardi, F. (2023). *Headlines Labeled as AI-Generated Are Less Likely to Be Believed and Shared, Even When True or Human-Generated*.

Asscher, O., & Glikson, E. (2023). Human evaluations of machine translation in an ethically charged situation. *New Media & Society*, *25*(5), 1087–1107.

Betsch, C., Ulshöfer, C., Renkewitz, F., & Betsch, T. (2011). The influence of narrative v. Statistical information on perceiving vaccination risks. *Medical Decision Making*, *31*(5), 742–753.

Borah, P., Xiao, X., Vishnevskaya, A., & Su, Y. (2023). Narrative versus statistical messages: The interplay of perceived susceptibility and misperceptions on vaccine intention. *Current Psychology*, 1–16.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., De Rosario, H., & De Rosario, M. H. (2018). Package 'pwr.' *R Package Version*, *1*(2).

Chan, C. K. Y., & Lee, K. K. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learning Environments*, *10*(1), 60.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., & Wang, Y. (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*.

Chen, C., & Shu, K. (2023). Combating misinformation in the age of llms: Opportunities and challenges. *arXiv Preprint arXiv:2311.05656*.

Chyung, S. Y., Swanson, I., Roberts, K., & Hankinson, A. (2018). Evidence-based survey design: The use of continuous rating scales in surveys. *Performance Improvement*, *57*(5), 38–48.

Clark, J. K., Wegener, D. T., Habashi, M. M., & Evans, A. T. (2012). Source expertise and persuasion: The effects of perceived opposition or support on message scrutiny. *Personality and Social Psychology Bulletin*, *38*(1), 90–100.

Clark, J. L., Green, M. C., & Simons, J. J. (2019). Narrative warmth and quantitative competence: Message type affects impressions of a speaker. *Plos One*, *14*(12), e0226713.

Deiana, G., Dettori, M., Arghittu, A., Azara, A., Gabutti, G., & Castiglia, P. (2023). Artificial intelligence and public health: Evaluating ChatGPT responses to vaccination myths and misconceptions. *Vaccines*, *11*(7), 1217.

Downen, T., Kim, S., & Lee, L. (2024). Algorithm aversion, emotions, and investor reaction: Does disclosing the use of AI influence investment decisions? *International Journal of Accounting Information Systems*, *52*, 100664.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660.

Hackenburg, K., & Margetts, H. (2023). *Evaluating the persuasive influence of political microtargeting with large language models*.

Han, B., & Fink, E. L. (2012). How do statistical and narrative evidence affect persuasion?: The role of evidentiary features. *Argumentation and Advocacy*, *49*(1), 39–58.

Hu, K. (2023, February 2). ChatGPT sets record for fastest-growing user base—Analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Hua, X., Huang, Y., & Zheng, Y. (2019). Current practices, new insights, and emerging trends of financial technologies. *Industrial Management & Data Systems*, *119*(7), 1401–1410.

Kang, M., Ragan, B. G., & Park, J.-H. (2008). Issues in outcomes research: An overview of randomization techniques for clinical trials. *Journal of Athletic Training*, *43*(2), 215–221.

Karinshak, E., Liu, S. X., Park, J. S., & Hancock, J. T. (2023). Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–29.

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, 106553.

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Lim, S., & Schmälzle, R. (2023a). Artificial intelligence for health message generation: An empirical study using a large language model (LLM) and prompt engineering. *Frontiers in Communication*, *8*, 1129082.

Lim, S., & Schmälzle, R. (2023b). The effect of source disclosure on evaluation of AI-generated messages: A two-part study. *arXiv Preprint arXiv:2311.15544*.

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from generative artificial intelligence is believed less. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 97–106.

Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, *32*(4), 1993–2020.

Madsen, J. K. (2016). Trump supported it?! A Bayesian source credibility model applied to appeals to specific American presidential candidates' opinions. *CogSci*.

Matz, S., Teeny, J., Vaid, S. S., Harari, G. M., & Cerf, M. (2023). *The Potential of Generative AI for Personalized Persuasion at Scale*.

Mellenbergh, G. J. (2019). Random and Systematic Errors in Context. *Counteracting Methodological Errors in Behavioral Research*, 1–12.

Perreault, W. D. (1975). Controlling order-effect bias. *The Public Opinion Quarterly*, *39*(4), 544–551.

Ragot, M., Martin, N., & Cojean, S. (2020). Ai-generated vs. Human artworks. A perception bias towards artificial intelligence? *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10.

Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction*, *39*(13), 2724–2741.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, *3*(2), pgae035.

Singh, S. K., Kumar, S., & Mehra, P. S. (2023). Chat GPT & Google Bard AI: A Review. *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 1–6.

Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis) informs us better than humans. *arXiv Preprint arXiv:2301.11924*.

Taherdoost, H. (2023). Fintech: Emerging trends and the future of finance. *Financial Technologies and DeFi: A Revisit to the Digital Finance Revolution*, 29–39.

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607–1622.

Xu, J. (2023). A meta-analysis comparing the effectiveness of narrative vs. Statistical evidence: Health vs. Non-health contexts. *Health Communication*, *38*(14), 3113–3123.

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., & Chen, Y. (2023). Siren's song in the ai ocean: A survey on hallucination in large language models. *arXiv Preprint arXiv:2309.01219*.

Zong, M., & Krishnamachari, B. (2022). A survey on GPT-3. *arXiv Preprint arXiv:2212.00857*.