Do Consumers Respond Positively to ChatGPT-Aided Ads?

An Experimental Study on the Disclosure of ChatGPT Involvement on Social Media

Xinhao He,[a] Huaqing Huang,[b] Juanjuan Meng[b,*]

[a] ROI Festival Charity Foundation, Hong Kong 999077, China

[b] Guanghua School of Management, Peking University, Beijing 100871, China

*Corresponding Author

Contact: onicek.he@roifestival.com (XH); huanghuaqing@stu.pku.edu.cn (HH); jumeng@gsm.pku.edu.cn (JM)

Abstract. This paper investigates consumer responses to the disclosure of ChatGPT involvement in advertising through both a field experiment on social media and a survey experiment. We present copies written by Humans, ChatGPT, and Human-ChatGPT collaborations both without and with authorship disclosure to control for copy quality, and examine the corresponding disclosure effects. Contrary to existing algorithm aversion literature, our findings reveal a novel phenomenon of positive disclosure effect for both ChatGPT and Human-ChatGPT collaboration across both experiments. Interestingly, the disclosure effect of ChatGPT on social media is more pronounced than that of Human-ChatGPT collaboration, while the reverse is observed in the survey experiment. Investigating the mechanisms based on preference and attention reveals that the preference channel tends to favor Human-ChatGPT collaborations, while the attention channel tends to favor ChatGPT alone, with the attention channel likely driving the results observed on social media.

JEL Classification: L5, M3, O38

Keywords: AI Disclosure Effect, ChatGPT, Human-AI collaboration, Algorithm Preference

1. Introduction

With the rise of Large Language Models (LLMs) like ChatGPT, governments and social media platforms are adjusting by requiring disclosure. The European Union AI Act, the world's first comprehensive AI law, regulates that Generative AI, like ChatGPT, will have to comply with transparency requirements. Social media like Meta and TikTok are also taking action. Meta now requires users to label digitally created or manipulated ads related to politics or social issues. TikTok has introduced new labels to help creators disclose their AI-generated content.

Despite the emergence of disclosure policies, especially on social media, research into their actual impacts and the underlying mechanisms remains scarce. We address this gap by combining a field experiment on social media and a survey experiment to examine consumer responses to the disclosure of ChatGPT involvement. Across experiments, we consistently find that in an undisclosed context, reflecting primarily copy quality, both the collaboration of humans with ChatGPT and independent ChatGPT writing match or exceed the effectiveness of independent human writing. Intriguingly, in contrast to the negative disclosure effect and algorithm aversion documented in existing literature (e.g., Dietvorst et al., 2014; Castelo et al., 2019; Luo et al., 2019), we identify a novel positive disclosure effect on consumers for both ChatGPT and Human-ChatGPT collaborations. This may suggest a critical shift in consumer attitudes towards AI-assisted content when the technology advances to LLMs that are more human-like.

Another interesting finding is that the disclosure effect of ChatGPT is more pronounced than that of Human-ChatGPT collaborations on social media, while the opposite holds true in the survey experiment. Further analysis attributes this disparity to the differential impacts of preference and attention channels. The preference channel tends to favor Human-ChatGPT collaboration, whereas the attention channel leans towards ChatGPT alone, and the attention channel is likely to drive the pattern observed on social media.

Our research holds significant implications for the conceptualization and assessment of policies related to artificial intelligence. The observed positive effect of disclosure suggests that the disclosure policy could potentially align with both the public interest and the private benefits of content creators and platforms. Furthermore, the differentiation between attention and preference mechanisms is informative. Given that attention towards ChatGPT might decrease over time, preference, on the other hand, is likely to exhibit increased endurance. This distinction could significantly impact societal decisions regarding the form of AI adoption favoring Human-AI collaboration in the long run.

Specifically, we collaborate with a firm to generate advertising copies of one real product written by a Human (the firm’s employees), ChatGPT, or a Human-ChatGPT collaborate. We then executed a social media experiment by disseminating the collected copies on Sina Weibo, a platform akin to Twitter in China (Chen and Konstan 2015). As of the end of 2023, Sina Weibo boasts approximately 260 million daily active users and over 600 million monthly active users, making it a pivotal platform for communication and public discourse in China. The copies were presented in two formats: with and without the writer's disclosure at the beginning of the post. However, we did not have a disclosure group for the “Human” category, as it is unconventional to disclose that on Weibo when viewers only see one copy from these groups. To ensure comparability, we conducted uniform advertising campaigns to disseminate the total of 435 posts from these groups to viewers, resulting in approximately 48,000 impressions per post. We obtained and analyzed the consumer responses at the post level, including the number of clicks and various forms of interactions, to understand the impact of disclosure on viewer engagement.

We also carried out a survey-based experiment involving 1055 respondents who were asked to score the copies. The survey experiment diverges from the Weibo setting in that attention is potentially more abundant due to monetary incentives and we can analyze the data at an individual level. Our design encompasses both within- and between-subjects design elements. In the first, a within-individual dimension, each respondent simultaneously evaluates copies from the Human Group, Human-ChatGPT Group, and ChatGPT Group in a randomized order. In the second, a between-individual dimension, it is determined whether respondents receive the undisclosed or disclosed version of the survey. With this design, the inclusion of a Human disclosure group becomes natural, given its placement alongside the disclosure of ChatGPT involvement.

Across experiments, we find that in an undisclosed context, reflecting primarily copy quality, both the collaboration of humans with ChatGPT and independent ChatGPT writing match or exceed the effectiveness of independent human writing. Intriguingly, in contrast to the negative disclosure effect and algorithm aversion documented in existing literature (Dietvorst et al. 2015, Castelo et al. 2019, Luo et al. 2019), we consistently identify a novel positive disclosure effect for both ChatGPT and Human-ChatGPT collaborations in both experiments. On Weibo, disclosure leads to significant increase in the number of responses, with the increase being 48.1% of disclosing ChatGPT and 30.6% of disclosing Human-ChatGPT collaboration compared to the corresponding undisclosed cases for the same copy content. In survey, disclosure leads to significant increase in the score, with the largest increase being 5.26% in the Human-ChatGPT Group, followed by 4.17% in the Human Group and 2.96% in the ChatGPT Group. The positive disclosure effect of ChatGPT involvement may suggest a critical shift in consumer attitudes towards AI-assisted content when the technology advances to LLMs that are more human-like. Another interesting finding is the magnitude of the disclosure effect. Based on the percentage number, the disclosure effect of ChatGPT is more pronounced than that of Human-ChatGPT collaborations on social media, while the opposite holds true in the survey experiment.

We then investigate the mechanism and identify three potential channels: copy quality, consumer attention, and consumer preference towards ChatGPT. Copy quality is primarily reflected in responses from undisclosed groups. Given that we're comparing the same copy content with and without disclosure, the estimated disclosure effect isn't impacted by copy quality, but could be due to consumer attention or preferences. Further analysis suggests that the disclosure effect on Weibo can be primarily attributed to the attention users give to the content in the form of "click" to view the full content rather than "like", while the survey experiment primarily showcases the preference effect, as the attention is more abundant and uniform across undisclosed and disclosed contexts. We therefore interpret the findings of both experiments to suggest that consumers favor the collaborative use of Human-ChatGPT over independent ChatGPT in terms of preferences. However, disclosing the use of standalone ChatGPT on social media may generate more interest, hence responses, than revealing a Human-ChatGPT collaboration in terms of attention channel.

Our paper contributes to the literature exploring how human react to algorithm. The literature has primarily documented algorithm aversion in AI delegation or AI assisted tasks from workers' perspective (Burton et al. 2020, Jussupow et al. 2020). This aversion can be potentially driven by lower tolerance for algorithmic errors (Dietvorst et al. 2015), concerns about algorithms neglecting the uniqueness of specific cases (Longoni et al. 2019), the subjective (vs. objective) nature of tasks (Castelo et al. 2019) and lower trust on AI feedback and perceived job displacement risk by AI (Tong et al. 2021). One exception is Logg et al. (2019), which documents that workers adhere more to advice on numerical predictions from an algorithm than from a person with survey experiments. Concurrently, efforts have been made to alleviate algorithm aversion, such as providing humans with opportunities to modify algorithmic output (Dietvorst et al., 2018), demonstrating unambiguous superiority of algorithms (Castelo et al., 2023), and even priming humans to think about God to instill a recognition of human fallibility (Karataş & Cutright, 2023). We contribute by examining consumers' algorithm preferences on social media, contrasting with studies focusing on workers, who represent a different market segment. Additionally, we explore whether attitudes toward traditional AI extend to advanced LLMs like ChatGPT, which exhibit increasingly human-like characteristics.

Our paper also contributes to the literature examining the impact of AI disclosure. Most studies predominantly show negative disclosure effects, such as decreasing product purchasing rates (Luo et al. 2019), impeding employee learning from AI feedback (Tong et al. 2021), and reducing trust in emails (Liu et al. 2022) and perceptions of product creativity (Magni et al. 2023). For example, Luo et al. (2019) randomized consumers to receive sales calls from chatbots or human workers, with and without chatbot identity disclosure. They found that disclosing the chatbot identity before the machine-customer conversation significantly reduces purchase rates. In contrast to their findings with traditional chatbot, our research documents a positive disclosure effect for ChatGPT as well as the newly introduced option of human-ChatGPT collaboration. Zhang and Goseline (2023) has also begun exploring the ChatGPT disclosure effect using survey experiments, revealing no significant effects when disclosing ChatGPT alone or human-ChatGPT collaboration. Our paper differs by documenting positive disclosure effects through a field experiment on social media. A novel aspect of our findings is that while consumer preferences tend to favor Human-ChatGPT collaboration over ChatGPT alone, the attention channels can sometimes produce the opposite effect.

Lastly, our study is related to emerging literature exploring the capabilities of Generative Pre-Trained Transformer (GPT). Specifically for the writing ability, Noy and Zhang (2023) find that Human-ChatGPT collaboration outperforms Human in professional writing with survey experiments. Chen and Chan (2023) provide insights into Human-GPT collaboration modalities in writing. Beyond writing, GPT has been evaluated for economic rationality (Chen et al., 2023) and behavioral and personality traits (Mei et al., 2023), and has demonstrated proficiency in diverse domains including text annotation (Gilardi et al., 2023), stock movement prediction (Xie et al., 2023), coding (Peng et al., 2023), and more. Our study extends this literature by systematically evaluating the writing abilities of Human, ChatGPT, and Human-ChatGPT collaboration in a field setting on social media.

This paper is organized as follows. Section 2 introduces the experimental design and empirical results of the Weibo experiment. Section 3 presents the survey experiment and its empirical results. Section 4 summarizes results from the above two experiments and further clarifies the mechanisms. In section 5, we conclude.

2. The Field Experiment on Weibo

2.1 Experimental Design

Phase One: Copy Collection In the initial phase, we collaborated with a company to collect advertising copies for a new shampoo brand.[1] The copies were divided into three groups. In the Human Group, the copies were independently written by employees of the company; in the Human-ChatGPT Group, the copies were collaboratively created by employees and ChatGPT 4.0; in the ChatGPT Group, the copies were autonomously generated by ChatGPT 4.0.

All three groups of copies were created based on a writing task document, which provided uniform product information and specified topics. For the two groups involving humans, we recruited employees from the company and randomly assigned them to either the Human Group or the Human-ChatGPT Group. Both groups received the same set of writing tasks. Data collection for these groups took place on May 8-9, 2023, within the company premises. The employees were brought to a meeting room in separate sessions and provided with public computers for writing. For the Human-ChatGPT Group, participants were given a brief introduction to ChatGPT and were provided with ChatGPT 4.0 accounts. To ensure that employees completed the writing tasks diligently and in accordance with the requirements, we informed them that all the collected copies would undergo evaluation and bonuses would be distributed based on the evaluation results. We implemented screenshot software to record participants' actions on the computer to monitor their compliance.

For the ChatGPT Group, we input the writing task documents that were directly shown to humans into ChatGPT 4.0. We then used the initial responses generated by the model as the final copies. For additional experimental details and related materials, please refer to Appendix 1.

In the end, we obtained 113 valid copies from 67 participants in the Human Group, 117 valid copies from 67 participants in the Human-ChatGPT Group, and 131 copies from the ChatGPT Group that are ready for release in Weibo.

[1] The experiment received approval from the Institutional Review Board for Human Participants at Guanghua School of Management, Peking University (IRB reference code: #2023-09). Informed consent of copywriters and survey participants were obtained before starting the experiment.

Phase Two: Copy Release In phase two, we implemented two modes of releasing copies on Weibo: undisclosed and disclosed with the author of the copy. For the disclosed groups, the author of the copy is indicated at the beginning of the post. In the disclosed Human-ChatGPT Group, it is disclosed as "This copy is written collaboratively by humans and ChatGPT". In the disclosed ChatGPT Group, it is disclosed as "This copy is written independently by ChatGPT". However, we do not have a disclosed group for the Human Group because users typically only receive one copy from these groups, and it would be unnatural to read an advertising copy that explicitly states it was written by humans only. In total, we have five groups: three undisclosed groups (Human Group, Human-ChatGPT Group, and ChatGPT Group) and two disclosed groups (disclosed Human-ChatGPT Group and disclosed ChatGPT Group).

The post follows a fixed format (see Appendix, Figure S1 for an example): a label (if in disclosed groups), a purchase link, a copy, and a product figure (half of the posts are randomly chosen to include a figure). To ensure a similar background for the audience seeing copies from different groups, we posted copies from different groups on the same topic at almost the same time, with uniformity in all aspects except for the content sources. The copies were posted from June 12, 2023 to July 6, 2023, for a total of 25 days, with an average of about 5 topics posted each day around different times of the day. The order of topics is the same for all groups and the order of posting copies from the same topic around the same time were randomized across treatment groups.

Since we used newly registered Weibo accounts to send out our copies, there would be almost no users actively coming across our posts. Therefore, we initiated uniform advertising campaigns for each post to display them among a series of other posts while users browse (see Appendix, Figure S2 for an illustration). We ultimately released and conducted successful advertising campaigns for 109, 114, 130, 36, 46 posts for Human Group, Human-ChatGPT Group, ChatGPT Group, disclosed Human-ChatGPT Group, disclosed ChatGPT Group, respectively, with average impressions being around 48,000 per post. The number of posts released for each disclosed group is lower than the undisclosed group primarily because several days after we started the experiment, the Weibo platform implemented a new policy that forbids labeling posts written by ChatGPT. Please refer to Appendix 1 for more detail.

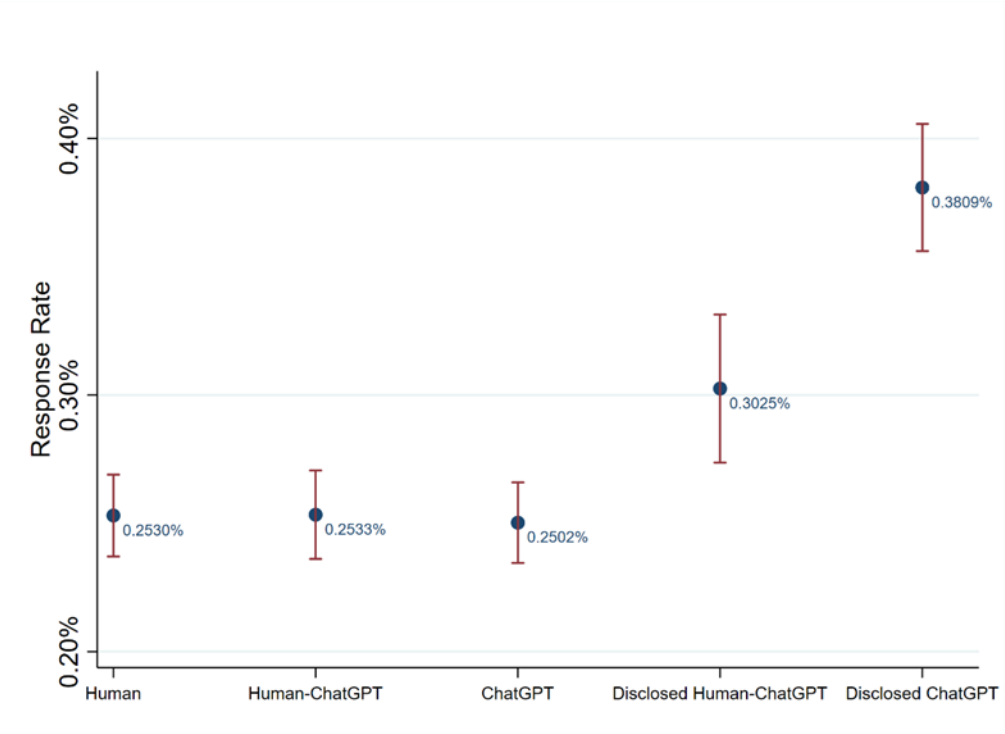
For each post, we track two key metrics: impressions and responses. Impressions represent the number of times a copy was displayed to an active user. Responses, on the other hand, is the summation of all click behavior and interactive behavior, with the first referring to expand clicks on body text and clicks on purchase links, profile photos, nicknames and figures and the second including shares, likes, comments, favorites, and new follower additions.

2.2 Empirical Results

- **Summary Statistics**

For writers, over 20% are male, with an average age of around 24 years old. For viewers, each post has been viewed by approximately 48,000 viewers. Around 25% viewers are male, and nearly 80% of them are aged 30 or younger. There is no significant difference for writers and viewers across groups (see Appendix, Table S3).

Figure 1. Weibo Experiment: Response Rate by Group



Note: This figure displays the average response rate in each group with 95% confidence interval.

Figure 1 illustrates the response rates across different groups. Response rate is defined as the number of responses divided by the number of impressions. The response rates for posts in the three undisclosed groups appear to be around 0.25% and statistically indistinguishable from each other. There is positive disclosure effect: disclosed posts exhibit significantly higher response rates than undisclosed posts in general, with disclosing Human-ChatGPT and ChatGPT receiving 0.30% and 0.38% response rates, respectively. The response rate here is similar in magnitude to the advertising click-through rate on Twitter in Lambrecht and Tucker (2019).

- **Regression Results**

The regression analysis is conducted at the post level. It's important to note that although the quantity of post-level observations might appear small, each observation effectively represents an amalgamation of more than 48,000 individual behaviors. This ensures the statistical power of the analysis.

We first exploit OLS regressions to evaluate copy outcomes from three undisclosed groups. Table 1, column (1) presents the regression result. Compared to the Human Group, the Human-ChatGPT Group and ChatGPT Group show almost no significant differences. Column (2) uses data from disclosed posts and let the disclosed ChatGPT Group be baseline. The responses for the disclosed Human-ChatGPT Group are significantly lower than disclosed ChatGPT Group by 17.6% (32.8/186.4, p<0.01).

We further estimate the disclosure effect by utilizing samples from all disclosed posts and their corresponding undisclosed counterparts, specifically comparing posts with identical copies with and without disclosure. Column (3) displays the result. we observe significantly positive disclosure effect for both Human-ChatGPT collaboration and independent ChatGPT writing. Firstly, based on the coefficient before *Disclose*, we find a significant and positive disclosure effect for ChatGPT Group: disclosing the authorship of ChatGPT significantly increases the response by 48.1% (60.4/125.7) relative to the case without disclosing (p<0.01). Secondly, the estimate of the interaction term implies significantly lower effect when disclosing collaboration between a human and ChatGPT (p<0.01) compared to disclosing ChatGPT, suggesting that disclosing Human-ChatGPT collaboration increases responses by 30.6% ((60.4-21.7)/126.6).

Table1 Weibo Experiment: Regression Analysis

DV: #Responses	(1) Undisclosed	(2) Disclosed	(3) Disclosure Effect
Human-ChatGPT	3.056	-32.766***	-4.255
	(3.319)	(9.674)	(9.517)
ChatGPT	1.653		
	(3.243)		
Disclose			60.419***
			(8.387)
Disclose*Human-ChatGPT			-21.716**
			(10.833)
Impression	0.003***	0.004***	0.003***
	(0.000)	(0.000)	(0.000)
Figure	73.753***	52.886***	68.140***
	(3.016)	(7.737)	(7.220)
Constant	-59.575***	28.413	-36.423
	(9.134)	(24.121)	(25.716)
Base Group	Undisclosed Human	Disclosed ChatGPT	Undisclosed ChatGPT
Day of Week FE (6 dummies)	Yes	Yes	Yes
Hour FE (8 dummies)	Yes	Yes	Yes
Topic FE (13 dummies)	Yes	Yes	Yes
Order FE (3 dummies)	Yes	Yes	Yes
p_value (Human-ChatGPT=ChatGPT)	0.654		
R-squared	0.943	0.957	0.914
Observations	353	82	162

Note: This table presents the OLS regression analysis of Weibo data. The dependent variable is the number of responses of each post. We control for *Impression*, that is, the number of times a copy was displayed to an active user and *Figure*, a dummy variable indicating whether the post attaches a figure. *Human-ChatGPT*, *ChatGPT* and *Disclose* are all dummy variables. A value of 1 for each represents that the copy was co-written by human-ChatGPT, written by ChatGPT alone, and disclosed when posting, respectively. We additionally control for a series of fixed effects including posting day of the week, posting hour, copy topic and posting order across the treatment groups. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Finding 1. *There are no significant differences in the number of responses among the three undisclosed groups. Disclosing the authorship of the posts leads to significant increase in the number of responses, with the increase being 48.1% in the ChatGPT Group and 30.6% in the Human-ChatGPT Group.*

3. The Survey Experiment

3.1 Experimental Design

We also carried out a survey experiment in which we recruit subjects from the WenJuanXing survey platform and ask them to rate our copies on various dimensions. The survey experiment diverges from the Weibo setting in that attention is potentially more abundant due to monetary incentives and we can analyze the data at an individual level. We can also add the disclosed Human Group to make the design of disclosure effect more complete, which is unnatural in the Weibo setting.

The survey consists of three parts. The first part inquiries about basic demographic information of the respondent, such as gender, age and education. The second part displays writing topics and corresponding copies and asks respondents to rate them in various dimensions. The third part inquiries about respondent's usage and attitude towards ChatGPT and other AI. The complete questionnaire is attached in Appendix 2.

We design two versions of questionnaires, undisclosed and disclosed, with each respondent receiving only one version. Our main treatment is in the second part, where each questionnaire randomly displays 3-6 copies under one or two topics and none of the copies attach figures. We have variations in two dimensions. The first dimension is within-individual: for each topic, we simultaneously display copies from Human Group, Human-ChatGPT Group, and ChatGPT Group in random order. The second dimension is between-individual: each respondent receives either an undisclosed or disclosed version of the survey. For the disclosed version, in addition to disclosing writers of copies from the Human-ChatGPT Group and ChatGPT Group with identical wordings in Weibo, we also label the copies from the Human Group as "This copy is independently written by a human".

Respondents will rate each copy from four perspectives: willingness to interact, willingness to purchase, creativity, and requirement alignment. Ratings range from 1 to 5, with higher numbers indicating a more positive evaluation. We take the average and obtain the score each respondent gives to each copy, which is the primary outcome variable.

The survey is incentivized with monetary rewards. Filling out the questionnaire is expected to take around 8 minutes and respondents will receive a total payment around 16 RMB (≈ 2 USD). In total 1055 participants joined our survey experiment.

3.2 Empirical Results

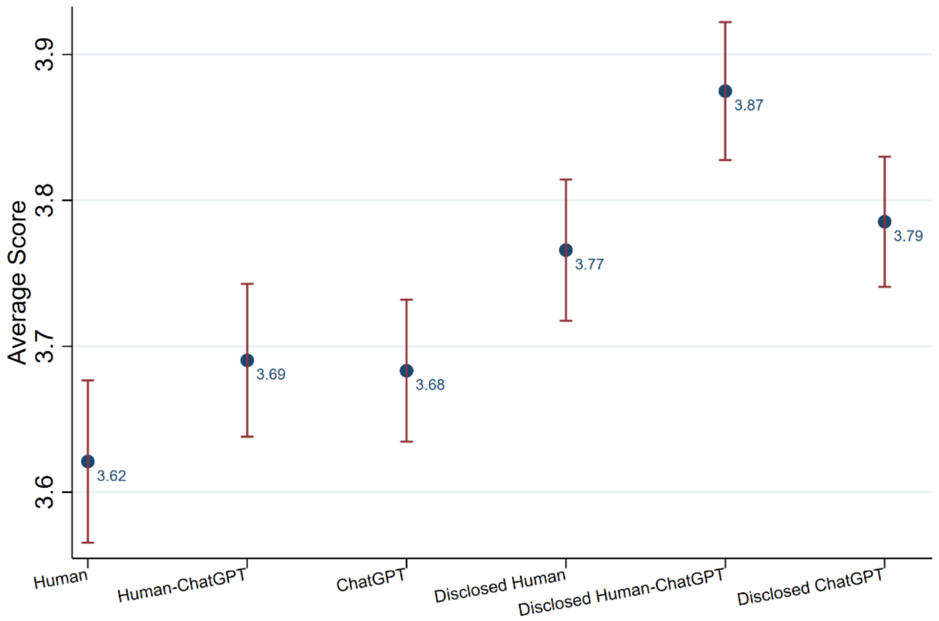
- **Summary Statistics**

Over 40% respondents are male, with an average age slightly above 30 years old. The balance check shows that there are no significant differences between respondents of undisclosed and disclosed questionnaires (Appendix, Table S8). As previously mentioned, attention is more abundant and uniform across treatments in the survey, which is evidenced by the insignificant difference in completion time for both questionnaires.

Figure 2 displays the average score for copies in different groups. When undisclosed, the average score for copies in the Human-ChatGPT Group and the ChatGPT Group are 3.69 and 3.68 respectively, both slightly higher than 3.62 in the Human Group. Comparing disclosed groups with undisclosed groups, we find positive disclosure effect for all three groups. When disclosed, the score of copies from all groups increase, with the Human-ChatGPT Group have the highest score 3.87, while the score for the Human Group is 3.77 and the ChatGPT Group is 3.79. The score increase after human disclosure is interesting but nonetheless consistent with the findings of human favoritism in Yang et al. (2023) and Zhang and Gosline (2023), wherein simply knowing that the content is generated by a human increases the perceived quality.[2] So it is likely that with AI implicit in background, people tend to demonstrate human favoritism.

[2] Another reason for positive human disclosure effect is that, the disclosure of independent ChatGPT writing and human-ChatGPT collaboration may have a spillover effect on disclosing human in our within-subject design. This possibility is less likely because in a between-individual design in Zhang and Gosline (2023), where one group views the content without disclosing human authorship and the other group with such disclosure (i.e. no ChatGPT is explicitly involved in the comparison), an increase in scores after disclosing human is also observed.

Figure 2. Survey Experiment: Average Score by Group.



Note: This figure displays the average score in each group with 95% confidence interval.

- **Regression Results**

We utilize OLS regressions and first analyze the undisclosed and disclosed context separately. Table 2, column (1) shows the result for undisclosed groups. Compared with copies from the Human Group, the Human-ChatGPT Group's copies have an average score that is 1.8% (0.064/3.62) higher, while the ChatGPT Group's copies have an average score 1.7% (0.061/3.62) higher (p<0.10). There is no significant difference between the two groups. Column (2) shows the result for disclosed survey and there has been a change in the score pattern. The copies from Human-ChatGPT Group achieve the highest scores that is 3.3% (0.123/3.77) higher than those from the Human Group (p<0.01), while copies from ChatGPT Group no longer show a significant difference from Human Group.

Columns (3) and (4) estimate the effect of disclosure by combining observations from both undisclosed and disclosed surveys. The estimation of this effect depends on the variations between subjects. Thus, in Column (3), we factor in respondent characteristics as opposed to respondent fixed effects, enabling us to estimate the disclosure effects across all three groups. In contrast, for Column (4), we use the Human Group as the baseline and incorporate respondent fixed effects. This approach allows us to estimate the relative differences in disclosure effects between the Human-ChatGPT and ChatGPT groups in comparison to the Human Group.

Table2 Survey Experiment: Regression Analysis

	(1)	(2)	(3)	(4)
			Disclosure	Disclosure
DV: Average Score	Undisclosed	Disclosed	Effect	Effect
Human-ChatGPT	0.064*	0.123***	0.074**	0.067**
	(0.034)	(0.035)	(0.032)	(0.034)
ChatGPT	0.061*	0.016	0.061**	0.063*
	(0.034)	(0.034)	(0.031)	(0.034)
Disclose			0.151***	
			(0.044)	
Disclose*Human-ChatGPT			0.043	0.056
			(0.045)	(0.049)
Disclose*ChatGPT			-0.042	-0.047
			(0.044)	(0.048)
Constant	4.099***	4.138***	3.307***	4.040***
	(0.114)	(0.036)	(0.135)	(0.073)
	Undisclosed	Disclosed	Undisclosed	Undisclosed
Base Group	Human	Human	Human	Human
Topic FE (13 dummies)	Yes	Yes	Yes	Yes
Order FE (5 dummies)	Yes	Yes	Yes	Yes
Respondent FE	Yes	Yes	No	Yes
Respondent Control	No	No	Yes	No
p_value(Human-ChatGPT=ChatGPT)	0.925	0.000476		
p_value(Disclose*Human-ChatGPT=Disclose*ChatGPT)			0.0321	0.0185
R-squared	0.510	0.506	0.039	0.511
Observations	2,912	2,908	5,820	5,820

Note: This table displays the OLS regression results of survey experiment. The dependent variable is the average score each respondent gives to each copy. *Human-ChatGPT, ChatGPT* and *Disclose* are all dummy variables. A value of 1 for each represents that the copy was co-written by human-ChatGPT, written by ChatGPT alone, and disclosed when displaying, respectively. We additionally control for a series of fixed effects, including copy topics, display orders, and respondents in Column (1), (2) and (4). In Column (3), we drop *Respondent FE* and add *Respondent Control*, they are all eight variables presented in balance check, so we can estimate the coefficient before *Disclose*. Standard errors are clustered at respondent level in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

In column (3), we observe significantly positive disclosure effect for all three groups. The coefficient before *Disclose* reveals that disclosing the authorship of a human significantly increases the average score by 4.17% (0.151/3.62) compared to cases without disclosure (p<0.01). From the two interaction terms, we find that disclosing Human-ChatGPT collaboration increases scores by 5.26% ((0.151+0.043)/3.69), while disclosing independent ChatGPT increases scores by 2.96% ((0.151-0.042)/3.68) compared to corresponding undisclosed contexts. The p-value of the test on the two interaction terms indicates that the higher score increase in disclosing Human-ChatGPT collaboration compared to disclosing independent ChatGPT is statistically significant (p < 0.05). These results are confirmed in column (4) in which respondent fixed effects were included.[3]

The heterogeneity analysis is in Appendix 5, which shows that the positive impact of disclosure, as well as the degree of its influence across all groups, is primarily attributed to female participants and those aged over 30.

Finding 2. *In the survey experiment, when undisclosed, both human-ChatGPT collaboration and independent ChatGPT writing achieve marginally significantly higher scores compared to independent human writing, with increases of 1.8% and 1.7%, respectively (p<0.10). Disclosure the authorship of the posts leads to significant increase in the score, with the largest increase being 5.26% in the Human-ChatGPT Group, followed by 4.17% in the Human Group and 2.96% in the ChatGPT Group (p<0.01). The difference between disclosing Human-ChatGPT and ChatGPT is statistically significant (p < 0.05).*

4. Mechanism

This section presents a summary of our empirical findings and delves into potential mechanisms. Our two experiments yield both consistent and divergent findings. Firstly, they consistently show that in an undisclosed context, both the collaboration of humans with ChatGPT and independent ChatGPT writing match or even outperform independent human writing. Secondly, both experiments reveal a positive disclosure effect for ChatGPT and Human-ChatGPT collaborations, contrasting with the negative

[3] The regression results of both the Weibo and survey experiments, after correcting for multiple-hypothesis testing with false discovery rate adjusted q-values (Anderson 2008), are presented in Appendix, Table S5 and S10. The results remain robust: the positive disclosure effect of ChatGPT and Human-ChatGPT remains significant at the 1% level, and the distinction between disclosing Human-ChatGPT and ChatGPT is significant at least at the 10% level.

disclosure effect and algorithm aversion noted in prior studies. However, a discrepancy was observed in the prominence of disclosure effects: ChatGPT had a more pronounced disclosure effect than Human-ChatGPT collaborations on social media, while the inverse was found in the survey experiment. Below we discuss the potential mechanisms for these findings based on copy quality, consumer attention, and ChatGPT preference.

- **Copy quality**

The influence of copy quality on response outcomes and survey ratings is evident (Lee et al. 2018, Chen and Chan 2023). When undisclosed, both Weibo responses and survey ratings primarily reflect copy quality. These collective findings imply that, in the context of promotional product copies, both collaborative writing with Human-ChatGPT and independent ChatGPT writing are at least on par with, if not superior to, independent human writing.

In the analysis of the disclosure effect, our methodology ensures control over the copy quality. This is because the content of the copies remains the same, regardless of whether the authorship is disclosed or not. We then resort to attention and preference in understanding the positive disclosure effect and their different magnitude we observed.

- **Attention**

Consumer attention is frequently limited in social media (Weng et al. 2012). As is common practice on many social media platforms, our advertising copies are displayed among a stream of posts while users are browsing, making them easy to overlook. In fact, the coefficients of *Figure* in Table 2 show that posts with figures have approximately 28.4%-57.9% higher responses on average compared to those without figures, suggesting that figures can potentially draw higher attention and limited attention is indeed an issue.

This attention issue can affect our estimated disclosure effect in social media. On Weibo, there's a default five-line space limit on the mobile screen (see Appendix, Figure S2). Displaying posts within space constraints is a common feature shared by nearly all major social media platforms, including Twitter, Facebook, and Instagram. In such context, users need to click anywhere in the body text to expand a post and this behavior is part of responses. If post authorship is disclosed, users may click to see the complete content due to heightened attention and curiosity, especially with ChatGPT's prominence today.

To further explore this possibility, we examined the numbers of “expand clicks” in evaluating the disclosure effect in Weibo. Interestingly, we noticed that the disclosure effect on the number of "expand clicks" mirrors the significant disclosure effect previously identified on total responses, while other sub-indices, including "likes", which are more likely to indicate preferences, did not exhibit such a significant pattern. This finding suggests that the attention drawn to ChatGPT-related content may play a crucial role for the outcome when disclosing ChatGPT involvement in social media platforms like Weibo.

In the survey, the problem of insufficient attention is mitigated by the context of filling out questionnaire with monetary incentives. Because we can present the copies in their entirety to the respondents, curiosity to click and see the entire content in the Weibo setting does not play a role here. Even if some level of inattention exists, there is no significant difference in attention between the undisclosed and disclosed groups, because the completion time difference between the two surveys us statistically insignificant, as shown in Appendix, Table S8. Therefore, attention does not appear to be a primary mechanism contributing to the positive disclosure effect in the survey.

- **Preference**

Aside from copy quality and attention, disclosure also stimulates users' preferences towards algorithms. On Weibo, both attention and preference may drive the disclosure effect, but we lack sufficient independent variation to distinguish between them. The aforementioned distinction between "click" to expand and "likes" suggests that the positive disclosure effect on Weibo may primarily result from increased attention. However, our Weibo setting may lack the power to identify preference due to the low level of attention.

Our survey experiment should predominantly reflect the preference effect, as the issue of limited attention is mitigated as discussed above. Thus, the fact that we observe positive disclosure effects, as well as a larger disclosure effect for Human-ChatGPT compared to ChatGPT alone in the survey experiment, suggests that consumer preferences yield positive disclosure effects favoring Human-ChatGPT.[4]

Overall, our findings indicate that both attention and preference channels contribute to the positive disclosure effect. Consumers appear to favor the collaboration of Human-ChatGPT over the standalone use of ChatGPT. In terms of attention, disclosing the use of independent ChatGPT on social media may attract more interest than revealing a Human-ChatGPT collaboration. These differences in attention and preference can account for the observed variances in the magnitude of the disclosure effect across social media and survey platforms.

5. Conclusion

Our research employs a combination of field experiment and survey experiment to evaluate the writing proficiency of ChatGPT and to examine the effects of disclosing ChatGPT involvement. Our findings indicate that in terms of copy quality, copies written by ChatGPT and Human-ChatGPT are comparable, if not superior, to those written by humans. When comparing disclosed and undisclosed contexts to examine the disclosure effect, we observe a positive disclosure effect in both experiments — responses on Weibo and survey scores significantly increase with disclosure. This contrasts with the negative or insignificant results found in existing literature (Luo et al., 2019; Zhang and Gosline, 2023). The discrepancy between our study and previous literature can be attributed to both preference and attention. Specifically, individuals may hold different attitudes towards traditional algorithms and large language models (LLMs) such as ChatGPT because the latter exhibits more human-like characteristics. Furthermore, consumers appear to demonstrate heightened attention towards ChatGPT-related content on social media in recent times. However, in terms of magnitudes of the disclosure effect, consumers tend to favor the collaboration of Human-ChatGPT over the standalone use of ChatGPT. For attention,

[4] Aside from preference, another possible explanation for the observed disclosure effect in survey experiment could be signaling of product quality, which suggests that revealing the use of GPT or Human-GPT collaboration might indicate superior product quality or advanced technology, thus increasing consumer scores. We tested this hypothesis with a survey asking respondents how much they agreed with two statements related to the perceived creativity and advancement of a company using GPT. Contrary to the signaling explanation, our data showed that disclosing the use of GPT had a similar positive effect on both those who agreed and disagreed with the statements. Even more unexpected was that the positive effect of disclosing Human-GPT was larger among those who disagreed. Therefore, the signaling of product quality is not likely to explain our findings here.

disclosing the use of independent ChatGPT on social media may attract more interest than revealing a Human-ChatGPT collaboration.

Our research contrasts with the algorithm aversion and negative AI disclosure effect in literature and highlights a novel change in how consumers feel about AI-assisted content, especially when it behaves more like a human. The complex interplay between GPT involvement, quality, consumer preferences, and attention, though subtle, is crucial for understanding consumer behavior in real-world settings. The attention towards ChatGPT may wane over time, but preference for Human-ChatGPT collaborations may be more enduring. This could potentially shape social choices concerning the form of AI adoption, offering valuable insights for future AI policy design and implementation.

Acknowledgments

We acknowledge the financial support from the National Natural Science Foundation of China (Grant No. 72225002).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4.0 for grammar editing. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Reference

Anderson ML (2008) Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, Perry preschool, and early training projects. *J. Am. Stat. Assoc.* 103(484):1481–1495.

Burton JW, Stein MK, Jensen TB (2020) A systematic review of algorithm aversion in augmented decision making. *J. Behav. Decis. Mak.* 33(2):220–239.

Castelo N, Boegershausen J, Hildebrand C, Henkel AP (2023) Understanding and improving consumer reactions to service bots. *J. Consum. Res.* 50(4):848–863.

Castelo N, Bos MW, Lehmann DR (2019) Task-dependent algorithm aversion. *J. Mark. Res.* 56(5):809–825.

Chen Y, Konstan J (2015) Online field experiments: a selective survey of methods. *J. Econ. Sci. Assoc.* 1(1): 29–42.

Chen Y, Liu TX, Shan Y, Zhong S (2023) The emergence of economic rationality of GPT. *Proc. Natl. Acad. Sci. U. S. A.* 120(51): e2316205120.

Chen Z, Chan J (2023) Large language model in creative work: The role of collaboration modality and user expertise. Working Paper.

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* 144(1):114–126.

Dietvorst BJ, Simmons JP, Massey C (2018) Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Manage. Sci.* 64(3):1155–1170.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. Working Paper.

Gilardi F, Alizadeh M, Kubli M (2023) ChatGPT outperforms crowd-workers for text-annotation tasks. *Proc. Natl. Acad. Sci. U. S. A.* 120(30): e2305016120.

Karataş M, Cutright KM (2023) Thinking about God increases acceptance of artificial intelligence in decision-making. *Proc. Natl. Acad. Sci. U. S. A.* 120(33):e2218961120.

Lambrecht A, Tucker C (2019) Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Manage. Sci.* 65(7):2966–2981.

Lee D, Hosanagar K, Nair HS (2018) Advertising content and consumer engagement on social media: Evidence from Facebook. *Manage. Sci.* 64(11):5105–5131.

Liu Y, Mittal A, Yang D, Bruckman A (2022) Will AI console me when I lose my pet? Understanding perceptions of AI-mediated email writing. *CHI Conference on Human Factors in Computing Systems*. (ACM, New York, NY, USA).

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* 151:90–103.

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical Artificial intelligence. *J. Consum. Res.* 46(4):629–650.

Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. Humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Mark. Sci.* 38(6):937-947.

Magni F, Park J, Chao MM (2023) Humans as creativity gatekeepers: Are we biased against AI creativity? *J. Bus. Psychol.*: 1-14.

Mei Q, Xie Y, Yuan W, Jackson MO (2023) A Turing Test: Are AI Chatbots behaviorally similar to humans? *Proc. Natl. Acad. Sci. U. S. A.* 121(9): e2313925121.

Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381(6654):187–192.

Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv [cs.SE].*

Tong S, Jia N, Luo X, Fang Z (2021) The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Manage. J.* 42(9): 1600-1631.

Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. *Sci. Rep.* 2(1):335.

Xie Q, Han W, Lai Y, Peng M, Huang J (2023) The Wall Street Neophyte: A zero-shot analysis of ChatGPT over MultiModal stock movement prediction challenges. *arXiv [cs.CL].*

Zhang Y, Gosline R (2023) Human favoritism, not AI aversion: People’s perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgm. Decis. Mak.* 18: e41.