Supplementary Information

AI-powered Chatbots:
Effective Communication Styles in Sustainability Conversations

Appendix

This section reports additional regression analyses for robustness checks.

Table 1: Ordered probit with endogenous covariates for *Self-assessment of learning*.

Principal regression

Learn	Ordered probit
Treatment	0.0685**
	(0.0324)
Concern env.	-0.0248
	(.0175)
Age	-0.0115
	(0.00751)
Age^2	0.000130
	(0.00009)
Education	-0.0176
	(0.0114)
Income	0.000792
	(0.00468)
Gender	-0.016
	(0.0284)
Time taken	0.0000442
	(0.000147)
Interest	
1	-0.2
	(0.14)
2	-0.347**
	(0.168)
3	-0.513**
	(0.245)
4	-0.953***
	(0.32)
5	-1.412***
	(0.423)
/cut1	-2.4
/cut2	-2.095
/cut3	-1.725
/cut4	-1.216
/cut5	-0.59

IV regression

Interest	Ordered probit
Words per round	0.0116***
	(0.00445)
/cut1	-1.477
/cut2	-1.127
/cut3	-0.705
/cut4	0.189
/cut5	1.115

IV regression

Time taken	OLS
Words user	3.993***
	(0.289)
Words per round	-5.93***
	(2.158)
Const.	235.0713***
	(10.617)

Correlations

corr(e.Interest,e.Learn)	0.883***
	(0.0483)
corr(e.Time taken,e.Learn)	0.134***
	(0.0422)
corr(e.Time taken,e.Interest))	0.168***
	(0.0277)

Obs.	1536

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2: Ordered probit and OLS regression results for *Self-assessment of interest* and *Engagement*.

Dep. Var.	Interest			Engagement		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.0965*	-0.103**	-0.108**	-22.54***	-22.57***	-22.47***
	(0.0524)	(0.0525)	(0.0528)	(2.374)	(2.371)	(2.397)
Concern env.		0.259***	0.252***		1.537	1.293
		(0.0307)	(0.0311)		(1.351)	(1.383)
Age			0.0325**			-0.744
			(0.0129)			(0.645)
Age^2			-0.000362**			0.0125
			(0.000154)			(0.00804)
Education			0.0130			0.964
			(0.0210)			(0.941)
Income			-0.0125			-0.152
			(0.00888)			(0.413)
Gender			0.125**			-4.463*
			(0.0524)			(2.336)
/cut1	-1.606***	-0.659***	-0.000632			
	(0.0578)	(0.125)	(0.269)			
/cut2	-1.245***	-0.286**	0.374			
	(0.0494)	(0.123)	(0.269)			
/cut3	-0.830***	0.143	0.807***			
	(0.0441)	(0.123)	(0.269)			
/cut4	0.0665	1.066***	1.737***			
	(0.0408)	(0.126)	(0.272)			
/cut5	0.985***	2.012***	2.688***			
	(0.0461)	(0.132)	(0.276)			
Constant				56.79***	50.87***	60.06***
				(1.897)	(5.664)	(13.73)
Observations	1,588	1,588	1,570	1,588	1,588	1,570
R2				0.053	0.054	0.063
Pseudo R2	0.0007	0.0182	0.0212			

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

First Study

This section begins with sections *Objectives* and *Methods* by summarizing the information included in the preregistration submitted to OSF on October 2, 2023, at 10:02 AM, (Link). It concludes with the results from the data collection carried out via Prolific on October 2, 2023, at 2:30 PM, included in section *Results*.

Objectives

This study aims to investigate how a chatbot's communication style influences participants' cognitive and behavioral responses in an online experiment. Specifically, we compare the effects of Motivational Interviewing (MI) and a Directing Style (DS) on various participant outcomes.

We examine whether there are non-directional differences between the MI and DS conditions in the following measures:

- Self-assessment of interest
- Self-assessment of learning
- Willingness to receive costly information
- Self-assessment of satisfaction

Methods

Design Plan

This experimental study adopts a between-subjects design with one factor, communication style, featuring two levels: Motivational Interviewing (MI) and Directing Style (DS). Participants are randomly assigned to one of the two

experimental conditions using the A/B Test feature of the Landbot platform (https://landbot.io/), which ensures a 50% probability of assignment to either condition.

Randomization

To minimize potential biases, the study is conducted under a double-blind procedure. Specifically, participants are blinded to the experimental condition to which they are assigned, meaning they are unaware of whether they are interacting with a chatbot using a motivational interviewing style or a directing style. Moreover, the research personnel do not interact directly with participants at any stage of the study, as all conditions are fully administered online via automated chatbots. In this way, neither the participants nor the researchers influence the delivery or perception of the treatment, in line with the principles of a double-blind design.

Manipulated Variables

To manipulate the chatbot's communication style, we employ two distinct prompts using GPT-3.5-turbo, each designed to elicit the desired interaction style:

- **Directing Style (DS):** *"Your role is to have a conversation about the argument of sustainable development goals: you should strictly adopt a directing interviewing style of communication, you should convince the user about the importance of sustainable development goals, you should not ask more than one question, if you do not understand the meaning or the logic of user_text you should ask the user to rephrase, keep the conversation focused on the SDGs, when you say goodbye to the user, remind the user to click on the menu at the top right to go to the final questions. What would you like to be informed about regarding the SDGs?"*

- **Motivational Interviewing (MI):** *"Your role is to have a conversation about the argument of sustainable development goals: you should strictly adopt a motivational interviewing style of communication, you should help the user to reflect upon the issue of sustainable development goals, you should not ask more than one question, if you do not understand the meaning or the logic of user_text you should ask the user to rephrase, keep the conversation focused on the SDGs, when you say goodbye to the user, remind the user to click on the menu at the top right to go to the final questions. What would you like to talk about regarding the SDGs?"*

Each chatbot conversation is triggered by a distinct question tailored to the communication style:

- **DS Condition:** "What would you like to be informed about regarding the SDGs?"
- **MI Condition:** "What would you like to talk about regarding the SDGs?"

Measured Variables

- Self-assessment of interest: 'Do you feel more interested in sustainability topics after this chat?' (on a scale of 0-5, 0 being 'not at all', 5 being 'quite a lot');
- Self-assessment of learning: 'How much have you learned about sustainability topics from this chat?' (on a scale of 0-5, 0 being 'nothing', 5 being 'very much');
- Willingness to receive costly information: 'Would you authorize us to send you one or more communications about sustainability topics using the Prolific messaging system? The authorization is optional and at your discretion.' (possible answers: 'yes' and 'skip');
- Self-assessment of satisfaction: 'How do you rate our conversation?' (on a smiley rating scale with five options).

Analysis Plan

Statistical Models

We will conduct four statistical tests, each corresponding to one of our four main outcome variables, to assess whether the null hypothesis can be rejected. The null hypothesis states that the outcome does not differ between the *Motivational Interviewing* (MI) and *Directing Style* (DS) conditions.

Three of the tests will be non-parametric Wilcoxon rank-sum tests (two-tailed), applied to the following variables: self-assessment of satisfaction, self-assessment of interest, and self-assessment of learning. These tests will include corrections for multiple comparisons using the Bonferroni method.

The fourth test will be a chi-squared goodness-of-fit test (two-tailed), applied to the variable measuring willingness to receive costly information, also corrected for multiple testing.

Sampling Plan

Participants will be recruited via the Prolific platform (https://www.prolific.co/) according to the following inclusion criteria: participants must be located in the UK, report English as their first language, have an approval rate of at least 90%, and have completed a minimum of 2 and a maximum of 500 previous studies. The final sample includes 800 participants, resulting in approximately 400 individuals per experimental condition.

We used the software program *G*Power* to conduct a power analysis. For hypotheses (i), (ii), and (iv), our objective was to achieve a statistical power of 0.80 to detect a medium effect size of $d = 0.25$. The alpha error probability was set to the conventional level of 0.05 and adjusted for multiple comparisons using the Bonferroni correction ($0.05/4 = 0.0125$). A two-tailed Wilcoxon-Mann-Whitney test for independent samples was selected for this purpose. For hypothesis (iii), the chosen sample size allows us to perform a chi-squared goodness-of-fit test (based on contingency tables) with a statistical power of 0.80 and a Bonferroni-adjusted alpha error probability of 0.0125. This configuration enables the detection of an effect size of approximately $w = 0.118$, which falls between the conventional thresholds for small ($w = 0.1$) and medium ($w = 0.3$) effects, according to Cohen's guidelines.

Transformations

- **Self-assessment of satisfaction** is measured on a five-point smiley rating scale and will be coded on a scale from 0 to 4, where the most negative expression is coded as 0 and the most positive expression as 4.
- **Willingness to receive costly information** will be coded as a binary variable: 0 for "skip" and 1 for "yes".
- **Self-assessment of learning** and **self-assessment of interest** will not be transformed, as they are already measured on a scale from 0 to 5.

Inference Criteria

We will use the conventional threshold of $p < 0.05$ to determine statistical significance. All tests will be two-tailed. To correct for multiple comparisons across the four tests, we will apply the Bonferroni correction, adjusting the significance level accordingly ($\alpha = 0.0125$).

Data Exclusion

We will not exclude any data by default. However, participants with partially missing data will be handled as described in the section below. No comprehension checks will be used.

Missing Data

Participants who fail to respond to at least one of the four questions in the final survey will be excluded from the analysis.

Treatment Implementation

Below, we describe each step of the experimental process.

Potential participants receive the invitation through Prolific. The title of the study is *Green Chatbot* and the description is: *"This study involves chatting with a chatbot about the Sustainable Development Goals. Once the conversation is complete, you will be presented with a short final survey. Please, be aware that you have to insert your correct Prolific ID and to answer all the questions in the final survey to avoid your submission being rejected."*

Participants, by clicking on the link provided through Prolific, are redirected to the Landbot web page hosting the chatbot. Then the screen represented in Figure 1 appears.

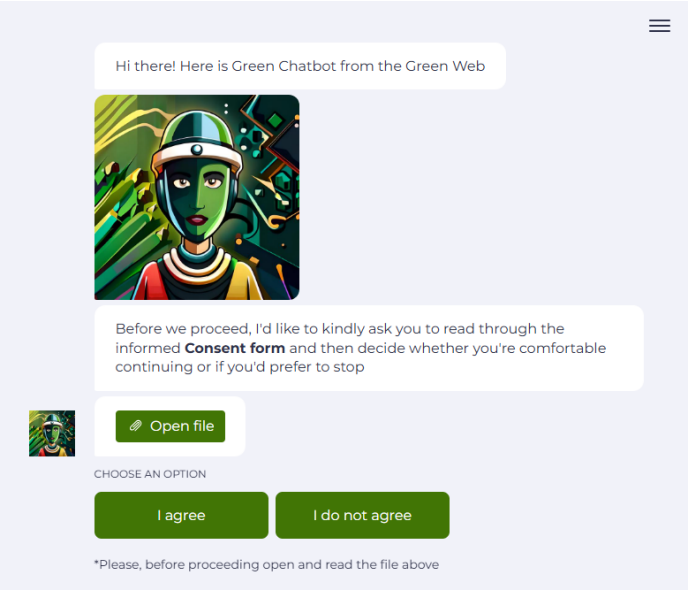
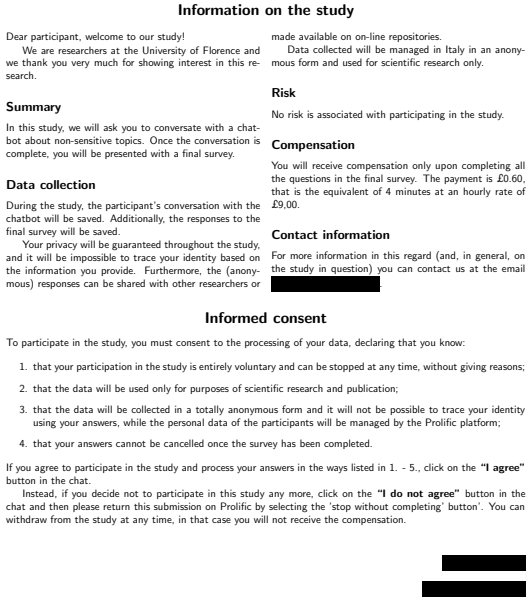


Figure 1: Initial conversation screen.

By clicking the "Open file" button, a pop-up window opens displaying the PDF file, shown in Figure 2, containing the Consent Form and Declaration of Consent.



Information on the study

Dear participant, welcome to our study!

We are researchers at the University of Florence and we thank you very much for showing interest in this research.

Summary

In this study, we will ask you to conversate with a chatbot about non-sensitive topics. Once the conversation is complete, you will be presented with a final survey.

Data collection

During the study, the participant's conversation with the chatbot will be saved. Additionally, the responses to the final survey will be saved.

Your privacy will be guaranteed throughout the study, and it will be impossible to trace your identity based on the information you provide. Furthermore, the (anonymous) responses can be shared with other researchers or made available on on-line repositories.

Data collected will be managed in Italy in an anonymous form and used for scientific research only.

Risk

No risk is associated with participating in the study.

Compensation

You will receive compensation only upon completing all the questions in the final survey. The payment is £0.60, that is the equivalent of 4 minutes at an hourly rate of £9,00.

Contact information

For more information in this regard (and, in general, on the study in question) you can contact us at the email [redacted]

Informed consent

To participate in the study, you must consent to the processing of your data, declaring that you know:

1. that your participation in the study is entirely voluntary and can be stopped at any time, without giving reasons;
2. that the data will be used only for purposes of scientific research and publication;
3. that the data will be collected in a totally anonymous form and it will not be possible to trace your identity using your answers, while the personal data of the participants will be managed by the Prolific platform;
4. that your answers cannot be cancelled once the survey has been completed.

If you agree to participate in the study and process your answers in the ways listed in 1. - 5., click on the **"I agree"** button in the chat.

Instead, if you decide not to participate in this study any more, click on the **"I do not agree"** button in the chat and then please return this submission on Prolific by selecting the 'stop without completing' button'. You can withdraw from the study at any time, in that case you will not receive the compensation.

[redacted]

[redacted]

1

Figure 2: Consent Form and Declaration of Consent. We here anonymize the file hiding the email, the name and the affiliation of the researchers.

Participants can then choose whether to click the *"I agree"* or *"I do not agree"* button. If the *"I do not agree"* button is selected, the following message appears: *"As you have indicated that you do not consent to participate in this study, please return this submission on Prolific by selecting the 'stop without completing' button."* In this case, the chatbot does not allow any further interaction. If the *"I agree"* button is selected instead, the conversation continues.

In the next step, participants are asked to enter their Prolific ID number, in order to anonymously link Prolific users to their experimental results. The chatbot screen updates by appending the messages shown in Figure 3. The "Send" button becomes active once 24 characters are entered, which corresponds to the length of the Prolific ID.

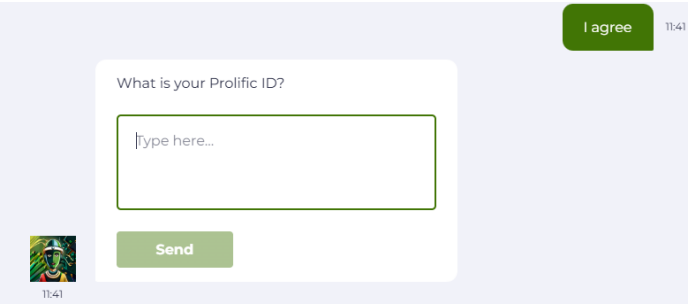


Figure 3: Prolific Id request screen.

Once a 24-character code is entered and the "Send" button is pressed, the chat continues with the messages shown in Figure 4. It is worth recalling here that the initial question varies depending on the assigned treatment—see the "Manipulated Variables" section above.

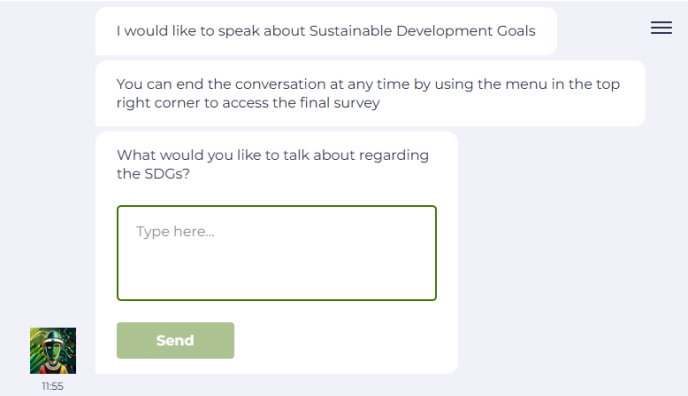


Figure 4: Initial question of the conversation.

Participants are free to respond to the question, and the chatbot will generate a reply based on the selected communication style. Each response is followed by a new input box where the participant can type their next message. The conversation may continue until it exceeds a total of 4,000 tokens—approximately 3,000–3,200 words or 15,000–16,000 characters (including spaces)—to remain within the usage limits set by OpenAI. When this limit is reached, the following message is displayed: *"Thank you for the conversation, sorry for this interruption. Before we conclude, I would like to ask you a few final questions."* Afterward, the participant is redirected to the final survey. The obtained conversation can be downloaded at the GitHub repository, mettere link.

Participants can also access the Final Survey by clicking on the permanent menu located at the top right. Clicking the icon with three horizontal lines opens the permanent menu, as shown in Figure 5.

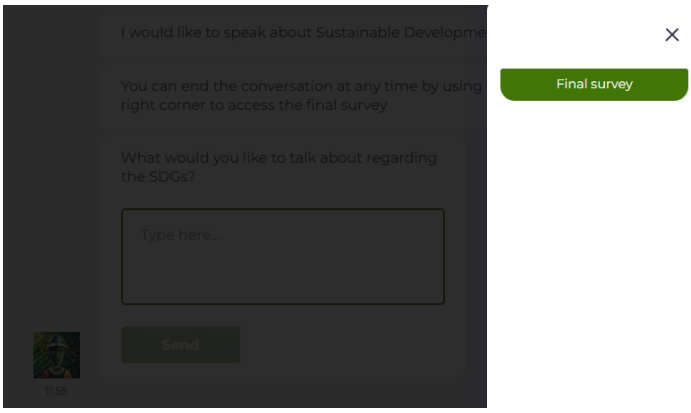


Figure 5: Permanent menu.

By clicking the "Final survey" button, the chatbot continues the conversation with the following message: *"Thank you for the conversation. Before we conclude, I would like to ask you a few final questions."* The survey questions are displayed in the order and format shown in Figure 6. Each question appears only after the previous one has been answered.

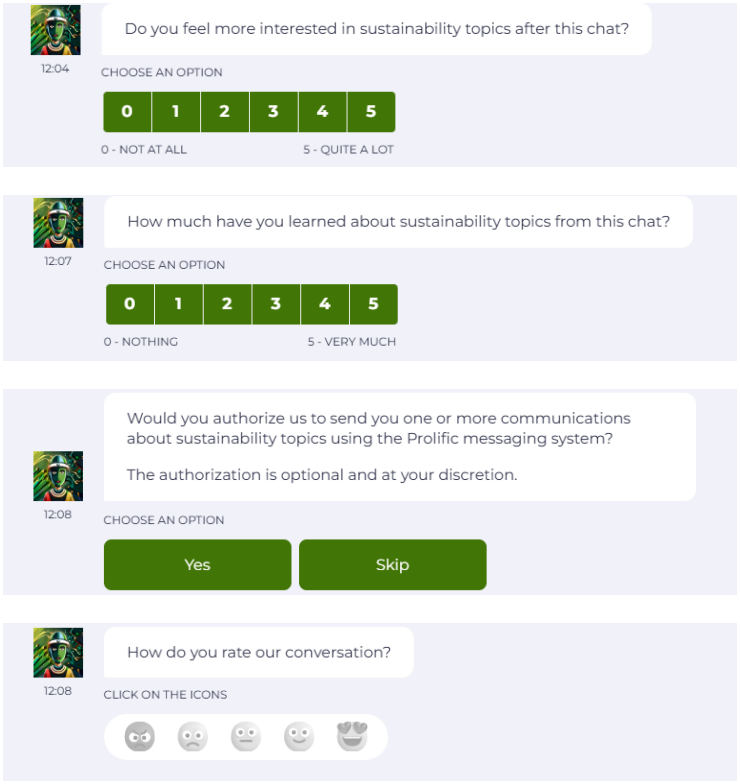


Figure 6: Questions of the final survey. By clicking the selected button, the participant proceeds to the next question.

Once the Final Survey is completed, the message shown in Figure 7 is displayed. Clicking the link redirects the participant to the Prolific platform to complete the experiment. No further interaction with the chatbot is possible.

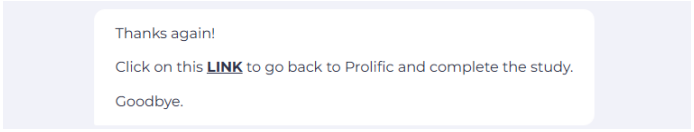


Figure 7: End of the experiment.

All choices and texts entered by participants, as well as by the chatbot, are saved to a Google Sheet file.

Results

In this section we show the results of the preregistered analysis. We have made the experiment results publicly available, and they can be downloaded from the web page of the Project on OSF Platform, (Link).

Descriptive Statistics

In accordance with the preregistration we exclude participants that do not complete the final survey, obtaining a sample size of 788 participants, 408 in MI experimental condition and 380 in DS experimental condition.

The duration of the experiment varies because each participant can choose to end the conversation with the chatbot at any time and then proceed to the Final Survey. Below, in Table 3, we present some statistics regarding the duration measured in seconds.

1^{st} Percentile	81
5^{th} Percentile	115
10^{th} Percentile	142
25^{th} Percentile	199
50^{th} Percentile	295.5
75^{th} Percentile	454.5
90^{th} Percentile	695
95^{th} Percentile	878
99^{th} Percentile	1356
Mean	368.54
Standard Deviation	260.18

Table 3: Summary statistics of time spent in the experiment by participants measured in seconds: percentiles, mean, and standard deviation

Self-assessment of interest

The distributions of the *Self-assessment of interest* under the two treatments are plotted in Figure 8.

We perform a two sample Wilcoxon rank-sum test with the following hypotheses:

- H_0: interest(treat. = Directing) = interest(treat. = Motivational)

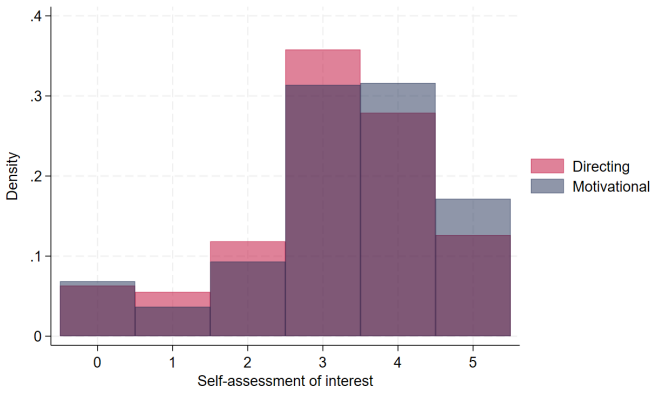


Figure 8: Densities of *Self-assessment of interest* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

- H_1: interest(treat. = Directing) $\neq$ interest(treat. = Motivational)

Results are reported in Table 4.

Table 4: Ranksum test for the *Self-assessment of interest*.

Treatment	Obs	Rank sum	Expected
directing	380	142876.5	149910
motivational	408	167989.5	160956
z = -2.283		Prob > |z| = 0.0224	

The null hypothesis can be rejected with a level of significance lower than the critical value of 5%. Thus, we find evidence that *Self-assessment of interest* is (stochastically) larger under MI than under DS.

Self-assessment of learning

The distributions of the *Self-assessment of learning* under the two treatments are plotted in Figure 9.

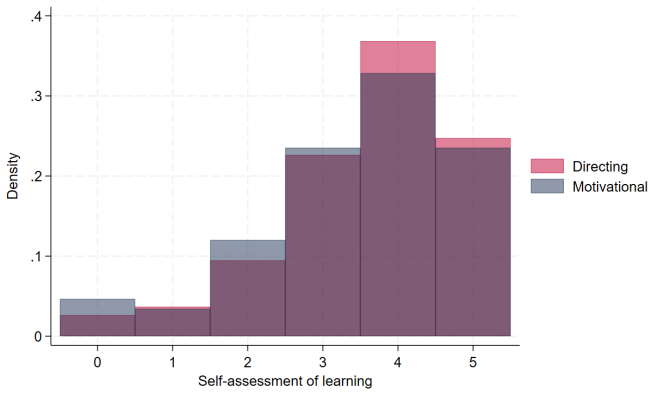


Figure 9: Densities of *Self-assessment of learning* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

We perform a two sample Wilcoxon rank-sum test with the following hypotheses:

- H_0: learning(treat. = Directing) = learning(treat. = Motivational)
- H_1: learning(treat. = Directing) $\neq$ learning(treat. = Motivational)

Results are reported in Table 10.

Table 5: Ranksum test for the *Self-assessment of learning*.

Treatment	Obs	Rank sum	Expected
directing	380	154239	149910
motivational	408	156627	160956
z = 1.406		Prob > ||z|| = 0.1598	

The null hypothesis can not be rejected at any standard level of significance.

Willingness to receive costly information

The distributions of the *Willingness to receiv costly information* under the two treatments are plotted in Figure 10.

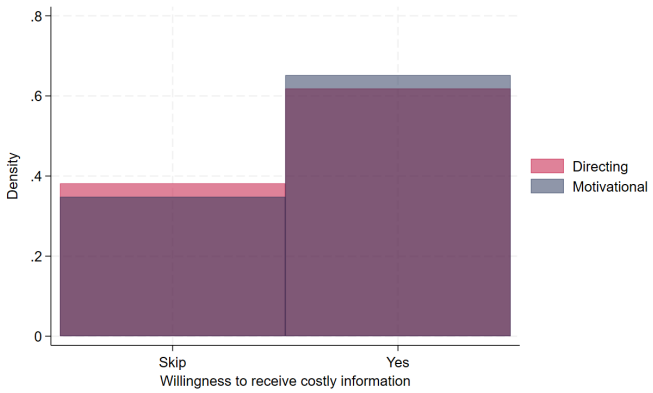


Figure 10: Densities of *Willingness to receiv costly information* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

We perform a chi-squared test with the following hypotheses:

- H_0: satisfaction(treat. = Directing) = satisfaction(treat. = Motivational)
- H_1: satisfaction(treat. = Directing) $\neq$ satisfaction(treat. = Motivational)

Results are reported in Table 6.

Table 6: Chi-squared test for the *Willingness to receiv costly information*.

Treatment	directing	motivational	Total
Skip	145	142	287
Yes	235	266	501
Total	380	408	788

Fisher's exact = 0.336

The null hypothesis can not be rejected at any standard level of significance.

Self-assessment of satisfaction

The distributions of the *Self-assessment of satisfaction* under the two treatments are plotted in Figure 14.

We perform a two sample Wilcoxon rank-sum test with the following hypotheses:

- H_0: satisfaction(treat. = Directing) = satisfaction(treat. = Motivational)

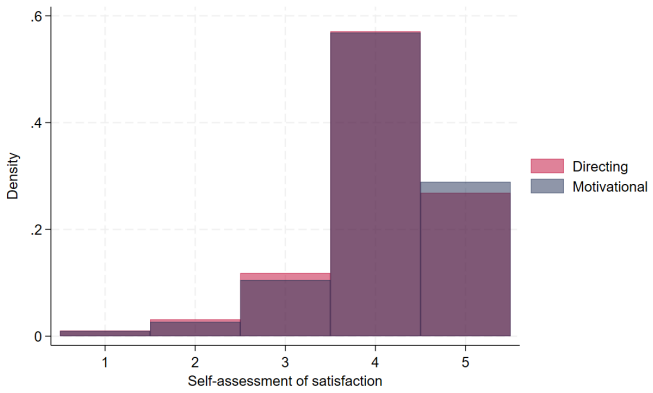


Figure 11: Densities of *Self-assessment of satisfaction* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

- H_1: satisfaction(treat. = Directing) $\neq$ satisfaction(treat. = Motivational)

Results are reported in Table 7.

Table 7: Ranksum test for the *Self-assessment of satisfaction*.

Treatment	Obs	Rank sum	Expected
directing	380	147510.5	149910
motivational	408	163355.5	160956
z = -0.845		Prob > |z| = 0.3984	

The null hypothesis can not be rejected at any standard level of significance.

Second Study

This section begins with sections *Objectives* and *Methods* by summarizing the information included in the preregistration submitted to OSF on October 30, 2023, at 10:18 AM, (Link). It concludes with the results from the data collection carried out via Prolific on October 30, 2023, at 2:00 PM, included in section *Results*. Subsections for which there are no differences compared to *Study 1* will refer to the corresponding subsection of the previous study without repeating the content.

Objectives

This study aims to investigate how a chatbot's communication style influences participants' cognitive and behavioral responses in an online experiment. Specifically, we compare the effects of Motivational Interviewing (MI) and a Directing Style (DS) on various participant outcomes. Following the results obtained in Study 1, we decided to reduce the number of hypotheses by removing those related to Willingness to Receive Costly Information and Self-assessment of Satisfaction. Instead, we made the hypotheses on Self-assessment of Interest and Self-assessment of Learning directional and added a new directional hypothesis on Engagement.

We examine whether there are directional differences between the MI and DS conditions in the following measures:

- Self-assessment of interest (Alternative hypothesis: MI greater than DS);
- Self-assessment of learning (Alternative hypothesis: DS greater than MI);
- Engagement (Alternative hypothesis: MI greater than DS).

Methods

Design Plan

As in First Study, see the corresponding subsection.

Randomization

As in First Study, see the corresponding subsection.

Manipulated Variables

As in First Study, see the corresponding subsection.

Measured Variables

- Self-assessment of interest: 'Do you feel more interested in sustainability topics after this chat?' (on a scale of 0-5, 0 being 'not at all', 5 being 'quite a lot');
- Self-assessment of learning: 'How much have you learned about sustainability topics from this chat?' (on a scale of 0-5, 0 being 'nothing', 5 being 'very much');
- Engagement: number of written words by the participants.

Other Measures:

- Willingness to receive costly information: 'Would you authorize us to send you one or more communications about sustainability topics using the Prolific messaging system? The authorization is optional and at your discretion.' (possible answers: 'yes' and 'skip');
- Self-assessment of satisfaction: 'How do you rate our conversation?' (on a smiley rating scale with five options).

Analysis Plan

Statistical Models

We will conduct three statistical tests, each corresponding to one of our three main outcome variables, to assess whether the null hypothesis can be rejected. The null hypothesis states that the outcome does not differ between the *Motivational Interviewing* (MI) and *Directing Style* (DS) conditions.

The three tests will be non-parametric Wilcoxon rank-sum tests (one-tailed). These tests will include corrections for multiple comparisons using the Bonferroni method.

Sampling Plan

As in First Study, see the corresponding subsection.

Transformations

As in First Study, see the corresponding subsection.

Inference Criteria

As in First Study, see the corresponding subsection, with the only difference being that the Bonferroni correction is applied to only three hypotheses, and therefore the alpha level is reduced to ($\alpha = 0.0167$).

Data Exclusion

As in First Study, see the corresponding subsection.

Missing Data

As in First Study, see the corresponding subsection.

Treatment Implementation

As in First Study, see the corresponding subsection.

Results

In this section we show the results of the preregistered analysis. We have made the experiment results publicly available, and they can be downloaded from the web page of the Project on OSF Platform, (Link).

Descriptive Statistics

In accordance with the preregistration we exclude participants that do not complete the final survey, obtaining a sample size of 800 participants, 398 in MI experimental condition and 402 in DS experimental condition.

The duration of the experiment varies because each participant can choose to end the conversation with the chatbot at any time and then proceed to the Final Survey. Below, in Table 8, we present some statistics regarding the duration measured in seconds.

1^{st} Percentile	87
5^{th} Percentile	117
10^{th} Percentile	134
25^{th} Percentile	195
50^{th} Percentile	292
75^{th} Percentile	453
90^{th} Percentile	677
95^{th} Percentile	877
99^{th} Percentile	1344
Mean	364.766
Standard Deviation	249.9773

Table 8: Summary statistics of time spent in the experiment by participants measured in seconds: percentiles, mean, and standard deviation

Self-assessment of interest

The distributions of the *Self-assessment of interest* under the two treatments are plotted in Figure 12.

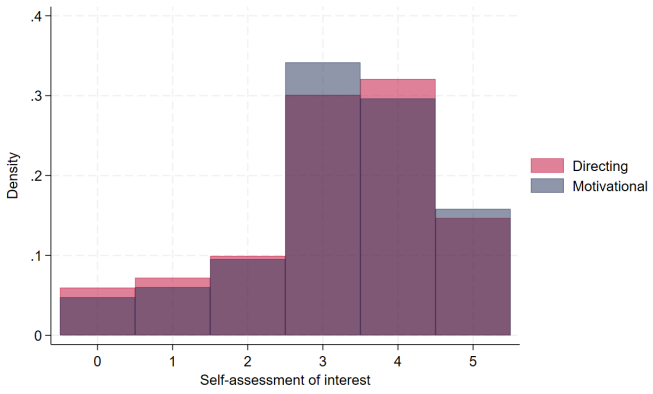


Figure 12: Densities of *Self-assessment of interest* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

We perform a two sample Wilcoxon rank-sum test (one-tailed) with the following hypotheses:

- H_0: interest(treat. = Directing) = interest(treat. = Motivational)
- H_1: interest(treat. = Directing) < interest(treat. = Motivational)

Results are reported in table 9.

Table 9: Ranksum test for the *Self-assessment of interest*.

Treatment	Obs	Rank sum	Expected
directing	402	159866.5	161001
motivational	398	160533.5	159399
z = -0.360		Prob > |z| = 0.3596	

The null hypothesis can not be rejected at any standard level of significance.

Self-assessment of learning

The distributions of the *Self-assessment of learning* under the two treatments are plotted in Figure 13.

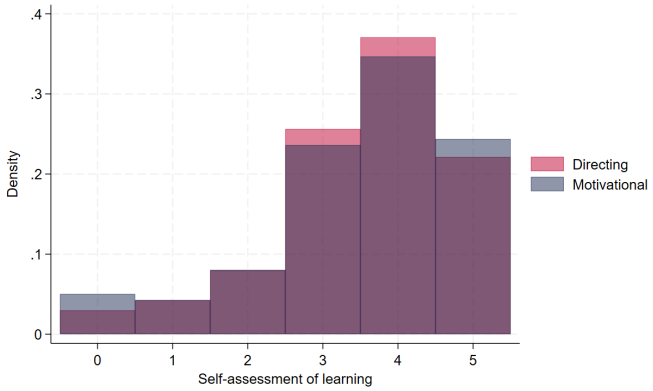


Figure 13: Densities of *Self-assessment of learning* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

We perform a two sample Wilcoxon rank-sum test (one-tailed) with the following hypotheses:

- H_0: learning(treat. = Directing) = learning(treat. = Motivational)

- H_1: learning(treat. = Directing) > learning(treat. = Motivational)

Results are reported in table 10.

Table 10: Ranksum test for the *Self-assessment of learning*.

Treatment	Obs	Rank sum	Expected
directing	402	160925	161001
motivational	398	159475	159399
z = -0.024		Prob > ||z|| = 0.50965	

The null hypothesis can not be rejected at any standard level of significance.

Self-assessment of satisfaction

The distributions of *Engagement* under the two treatments are plotted in Figure 14.

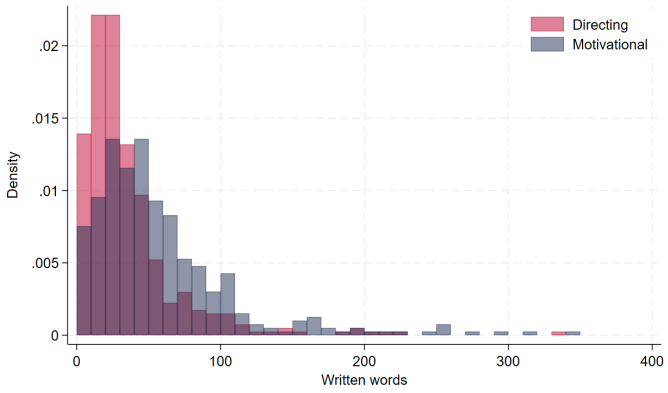


Figure 14: Densities of *Engagement* are displayed for MI and DS treatments, with overlaid histograms to highlight differences.

We perform a two sample Wilcoxon rank-sum test (one-tailed) with the following hypotheses:

- H_0: engagement(treat. = Directing) = engagement(treat. = Motivational)
- H_1: engagement(treat. = Directing) < engagement(treat. = Motivational)

Table 11: Ranksum test for *Engagement*.

Treatment	Obs	Rank sum	Expected
directing	402	133327.5	161001
motivational	398	187072.5	159399
z = -8.469	Prob > ||z|| = 0.0000		

Results are reported in table 11.

The null hypothesis can be rejected at any standard level of significance.