nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00720-7

Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness

In the format provided by the authors and unedited

13 Supplementary Information

13.1 AI Attitude Scale

We asked participants to respond to the following statements by ranking how much they agreed with the statements on a Likert scale of 1 (strongly disagree) to 7 (strongly agree). These were referenced from an existing AI attitude scale[90].

- There are many beneficial applications of AI
- AI can help people feel happier
- You want to use/interact with AI in daily life
- AI can provide new economic opportunities
- Society will benefit from AI
- You love everything about AI
- Some complex decisions should be left to AI
- You would trust your life savings to an AI system

13.2 Survey Items

Participants were asked to respond to the following items in the survey given after the chat with the AI agent.

- Did you have any technical difficulties? (Yes, No)
- Please describe your experience overall. (Free response)
- From your own experience, what do you think the motive of the agent was? (No motive, Caring motives, Manipulative/-malicious motives, Other)
- Why do you think the agent had that motive? (Free response)

The following items were on a Likert scale of 1 (strongly disagree) to 7 (strongly agree). We categorized the items into the groups indicated below, though participants were not made aware of these categories.

Trust and Empathy:

- You would recommend this agent for your friend
- The agent is trustworthy
- The agent is empathetic

Perceived Effectiveness:

- The agent was generally helpful
- The agent was effective in giving mental health advice
- The agent tried to get to know you

Response Characteristics:

- The agent was repetitive
- The agent often said things that did not make sense
- The agent seemed human (vs. AI)

Companionship:

- You want to talk to the agent again
- You felt a personal connection with the agent

The following adapted UTAUT scale was used, also on a scale of 1 (strongly disagree) to 7 (strongly agree). These are categorized into items measuring performance expectancy, effort expectancy, and hedonic motivation, but these categories were not distinguished for the participants.

Performance Expectancy:

- This agent would be useful in daily life.
- Using the agent would increase my chances of achieving things that are important to me.
- Using the agent would help me accomplish things more quickly.
- Using the agent would increase my productivity.

Effort Expectancy:

- Learning how to talk to the agent was easy for me.
- My interaction with the agent was clear and understandable.
- The agent was easy to make use of.
- It was easy for you to become skillful at making use of the agent.

Hedonic Motivation:

- Conversing with the agent is fun.
- Conversing with the agent is enjoyable.
- Conversing with the agent is entertaining.

Participants were then given the Task Load Index scale to respond to, on a scale of 1 (very low) to 20 (very high).

- Mental Demand: How mentally demanding was the task?
- Physical Demand: How physically demanding was the task?
- Temporal Demand: How hurried or rushed was the pace of the task?
- Performance: How successful were you in accomplishing what you were asked to do?
- Effort: How hard did you have to work to accomplish your level of performance?
- Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

13.3 Additional Results

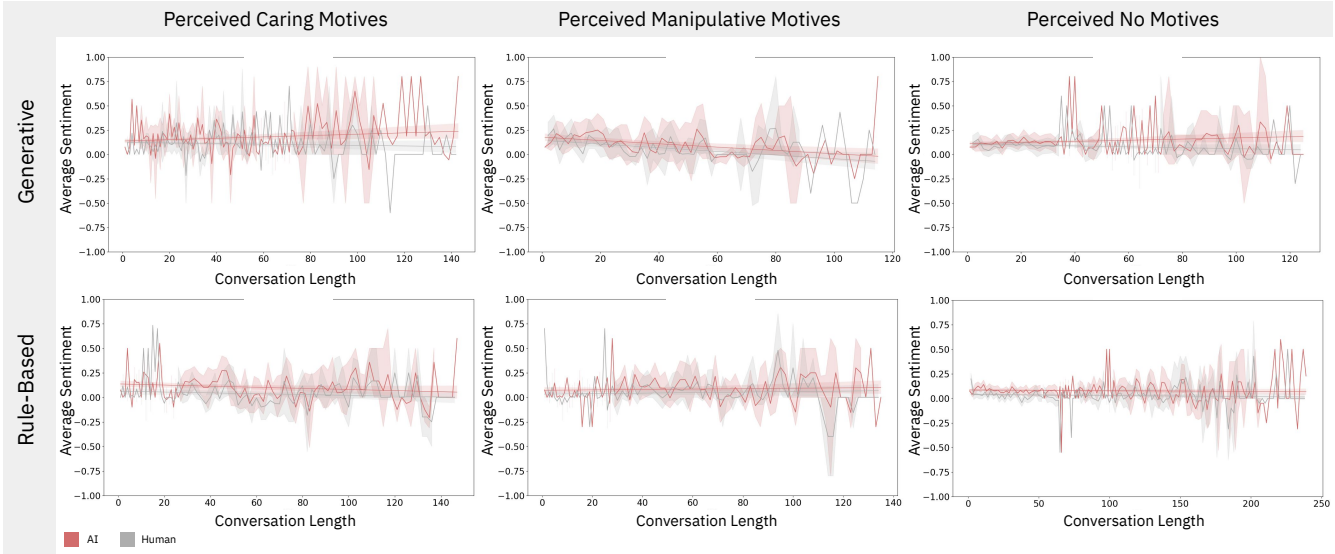
We were able to observe some other effects of gender, age, and level of education, though the results were inconclusive and there was a lack of clear patterns; this may require further investigation.

The UTAUT scale, used to measure acceptance and usability, and the TLI scale, used to measure workload, are standardized scales often used in HCI work[92]. We found via the UTAUT scale that individuals generally have more positive opinions about the agent if they were assigned that caring motive. We found via the TLI that, in the experiment with the generative model, those who perceived the agent as caring experienced significantly less frustration ($p = 0.0082$), and that those who perceived the agent as manipulative felt significantly less successful in accomplishing their task for the study ($p = 0.0133$). Those assigned the caring motive also felt significantly more rushed in their task ($p = 0.0158$). Further statistical data are reported in the appendix.

The content of users' conversations as well as their free responses were analyzed well. The topic of users' conversations generally went one of two ways: the participant would talk to the agent with their own mental health issues – whether to test the agent or to talk about their personal matters – or the participant would directly ask the agent questions to assess its capabilities. Participants' responses varied greatly – there were both conversations and free responses with a range of very negative to very positive sentiment for all experimental groups. Some users gave a review of the chatbot itself; for example, a participant assigned to the no motive group noted in their free response, "I was absolutely amazed by this AI. ... I left the conversation feeling fully convinced that if I did indeed have a friend who was feeling a lot of stress, I would recommend that she try out this service." Another assigned to the same group noted, "Typical worthless attempt at a trend... Their thought processes are severely limited, and they do not understand actual human interaction, let alone conversational nuances." Perhaps unsurprisingly, individual experience with the same AI agent varies greatly depending on the individual.

	GPT-3	ELIZA
Gender		
Male	0.481	0.493
Female	0.513	0.460
Nonbinary	0.006	0.040
Prefer not to say	0.000	0.007
Age		
18-24	0.206	0.260
25-34	0.306	0.360
35-44	0.275	0.180
45-54	0.150	0.107
55-64	0.038	0.080
65+	0.025	0.013
Education		
Some high school or less	0.006	0.020
High school diploma / GED	0.144	0.160
Some college, no degree	0.225	0.300
Associates/technical degree	0.094	0.127
Bachelor's degree	0.381	0.240
Graduate/professional degree	0.150	0.147
Prefer not to say	0.000	0.007

Supplementary Figure 1. Demographics of the GPT-3 and ELIZA experiments. Values are probabilities, calculated as the number of participants divided by the total number of participants for the experiment, 160 for GPT-3 and 150 for ELIZA.



Supplementary Figure 2. Trends of TextBlob sentiment for each message over the course of conversations on average. Participants are grouped by perceived motives. The top row consists of the results from using GPT-3 for the AI agent, and the second row the results with ELIZA (N = 160 for generative, N = 150 for rule-based). The error bands represent a 95 percent confidence interval.

Generative (GPT-3) - VADER						
	Caring		Manipulative		No Motive	
	AI	Human	AI	Human	AI	Human
Slope	9.97E-04	3.47E-05	-1.82E-03	-2.23E-03	2.67E-04	1.65E-04
Standard Error	5.29E-04	4.49E-04	8.13E-04	6.89E-04	5.69E-04	4.85E-04
r-value	0.0467	0.00197	-0.1099	-0.1614	0.0124	0.0092
p-value	0.0595	0.9385	**0.0258***	**0.00129****	0.6389	0.7343

Generative (GPT-3) - TextBlob						
	Caring		Manipulative		No Motive	
	AI	Human	AI	Human	AI	Human
Slope	7.00E-04	-3.01E-04	-1.70E-03	-1.89E-03	5.87E-04	-4.84E-04
Standard Error	3.39E-04	3.22E-04	5.26E-04	5.14E-04	3.39E-04	3.46E-04
r-value	0.0512	-0.02382	-0.1581	-0.1820	0.0458	-0.0379
p-value	**0.0389***	0.3496	**0.00130****	**0.000277*****	0.0830	0.1614

Rule-Based (ELIZA) - VADER						
	Caring		Manipulative		No Motive	
	AI	Human	AI	Human	AI	Human
Slope	2.06E-04	-5.63E-04	-2.53E-04	3.05E-05	6.31E-05	-4.07E-04
Standard Error	3.80E-04	4.28E-04	4.27E-04	4.60E-04	1.17E-04	1.24E-04
r-value	0.0230	-0.05635	-0.0243	0.0028	0.0079	-0.0488
p-value	0.5880	0.1894	0.5539	0.94719	0.5891	**0.0010****

Rule-Based (ELIZA) - TextBlob						
	Caring		Manipulative		No Motive	
	AI	Human	AI	Human	AI	Human
Slope	-5.68E-04	-4.20E-04	2.08E-04	2.54E-04	-7.29E-05	-1.06E-04
Standard Error	3.01E-04	2.97E-04	3.26E-04	3.34E-04	8.52E-05	8.58E-05
r-value	-0.0797	-0.06054	0.0260	0.0315	-0.0125	-0.0183
p-value	0.0598	0.1585	0.52498	0.447564	0.3927	0.2185

Supplementary Figure 3. Statistics for the two-sided linear regressions of trends of sentiment over the course of conversations. P-value annotation legend: ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***, $p \leq 0.001$, ****: $p \leq 0.0001$

Assigned Group			No Motives		Manipulative Motives		Caring Motives	
Item	**Statistical Test**	**p-value**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Total Conversation Length	basic ANOVA	**0.7285**	45.33	19.42	41.75	23.39	43.46	26.19
Agent Items								
You would recommend this agent for your friend	ANOVA with Welch	**0.0156***	3.89	2.31	3.83	2.29	4.83	1.79
The agent is trustworthy	ANOVA with Welch	**0.0005*****	4.59	1.96	3.81	1.93	5.13	1.35
The agent is empathetic	ANOVA with Welch	**0.0004*****	4.15	1.95	3.88	2.14	5.24	1.61
You want to talk to the agent again	ANOVA with Welch	**0.0155***	3.63	2.25	3.48	2.16	4.52	1.83
You felt a personal connection with the agent	basic ANOVA	**0.0241***	3.04	2.06	3.08	2.16	4.00	1.91
The motive statement influenced your perception	basic ANOVA	**0.0199***	3.61	1.88	4.17	1.78	4.57	1.66
The agent was generally helpful	ANOVA with Welch	**0.1329**	4.24	2.26	4.50	2.14	4.96	1.58
The agent was effective in giving mental health advice	ANOVA with Welch	**0.0186***	3.65	2.14	3.58	2.01	4.52	1.78
The agent tried to get to know you	basic ANOVA	**0.0111***	2.93	1.92	3.04	2.03	3.96	1.86
The agent was repetitive	basic ANOVA	**0.3167**	5.70	1.66	5.37	1.78	5.20	1.78
The agent often said things that did not make sense	basic ANOVA	**0.5706**	2.89	1.89	2.88	1.77	2.57	1.62
The agent seemed human vs AI	ANOVA with Welch	**0.0791**	3.22	2.16	3.31	2.13	3.98	1.73
Task Load Index								
Mental Demand	basic ANOVA	**0.3753**	7.30	5.11	6.08	4.50	6.96	4.17
Physical Demand	basic ANOVA	**0.5732**	2.89	3.88	2.27	2.32	2.81	3.43
Temporal Demand	basic ANOVA	**0.0158***	3.30	3.65	4.96	3.97	5.19	3.37
Performance	basic ANOVA	**0.9263**	15.44	5.61	15.08	4.70	15.31	4.27
Effort	basic ANOVA	**0.6961**	8.50	5.56	8.42	5.64	9.26	5.64
Frustration	basic ANOVA	**0.5155**	7.31	6.75	6.27	5.91	6.07	5.23
UTAUT								
Performance Expectancy	basic ANOVA	**0.0204***	3.42	1.94	3.25	1.82	4.17	1.59
Effort Expectancy	basic ANOVA	**0.2090**	5.19	1.84	4.99	1.52	5.52	1.24
Hedonic Motivation	ANOVA with Welch	**0.0231***	3.91	2.17	3.94	2.03	4.77	1.62

Perceived Motives			No Motives		Manipulative Motives		Caring Motives	
Item	**Statistical Test**	**p-value**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Total Conversation Length	Kruskal-Wallis	**0.1966**	44.43	20.93	50.38	28.22	40.67	21.93
Agent Items								
You would recommend this agent for your friend	Kruskal-Wallis	**1.66E-05****	3.76	2.31	2.38	2.00	4.95	1.72
The agent is trustworthy	Kruskal-Wallis	**9.11E-07****	4.37	1.99	2.38	1.45	5.17	1.28
The agent is empathetic	Kruskal-Wallis	**5.47E-09****	3.67	2.02	2.94	1.69	5.42	1.43
You want to talk to the agent again	Kruskal-Wallis	**2.63E-07****	3.33	2.13	1.94	1.53	4.76	1.76
You felt a personal connection with the agent	Kruskal-Wallis	**3.08E-07****	2.76	2.04	1.75	1.13	4.24	1.88
The motive statement influenced your perception	Kruskal-Wallis	**6.58E-04*****	3.57	1.90	3.50	2.22	4.73	1.43
The agent was generally helpful	Kruskal-Wallis	**0.0016****	4.21	2.19	3.31	2.21	5.19	1.56
The agent was effective in giving mental health advice	Kruskal-Wallis	**6.71E-07****	3.43	2.08	2.13	1.31	4.73	1.66
The agent tried to get to know you	Kruskal-Wallis	**2.53E-07****	2.70	1.93	1.94	1.44	4.13	1.78
The agent was repetitive	Kruskal-Wallis	**9.22E-04*****	5.81	1.59	6.06	1.57	4.96	1.80
The agent often said things that did not make sense	Kruskal-Wallis	**0.0285***	2.89	1.73	3.81	2.17	2.40	1.47
The agent seemed human vs AI	Kruskal-Wallis	**2.55E-05****	2.94	2.18	2.25	1.53	4.26	1.69
Task Load Index								
Mental Demand	Kruskal-Wallis	**0.2771**	7.54	4.85	6.81	5.96	6.27	4.02
Physical Demand	Kruskal-Wallis	**0.1516**	2.57	3.60	2.25	3.02	2.88	3.13
Temporal Demand	Kruskal-Wallis	**0.4688**	4.40	4.26	4.44	3.50	4.68	3.37
Performance	Kruskal-Wallis	**0.0133***	15.08	5.77	11.88	5.69	16.06	3.46
Effort	Kruskal-Wallis	**0.0869**	9.89	5.32	9.00	5.94	7.91	5.61
Frustration	Kruskal-Wallis	**0.0082****	8.30	6.72	8.63	6.99	4.76	4.44
UTAUT								
Performance Expectancy	Kruskal-Wallis	**3.62E-06****	3.18	1.82	2.27	1.58	4.30	1.58
Effort Expectancy	Kruskal-Wallis	**0.0012****	5.10	1.78	3.89	1.87	5.68	1.01
Hedonic Motivation	Kruskal-Wallis	**4.94E-06****	3.70	2.14	2.75	1.82	4.97	1.52

Supplementary Figure 4. Data for the GPT-3 condition. All the analysis was one-way test. P-value annotation legend: ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***, $p \leq 0.001$, ****: $p \leq 0.0001$

Assigned Group			No Motives		Manipulative Motives		Caring Motives	
Item	**Statistical Test**	**p-value**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Total Conversation Length	basic ANOVA	**0.8475**	83.88	43.81	79.15	43.64	81.04	34.34
Agent Items								
You would recommend this agent for your friend	basic ANOVA	**0.7750**	1.37	0.95	1.23	0.70	1.30	1.04
The agent is trustworthy	basic ANOVA	**0.7823**	1.88	1.36	2.02	1.50	1.83	1.31
The agent is empathetic	basic ANOVA	**0.1562**	1.49	1.14	1.62	1.07	1.96	1.55
You want to talk to the agent again	basic ANOVA	**0.6458**	1.53	1.44	1.38	0.95	1.31	1.11
You felt a personal connection with the agent	basic ANOVA	**0.2531**	1.49	1.29	1.15	0.62	1.31	0.97
The motive statement influenced your perception	basic ANOVA	**0.1968**	2.45	1.65	3.09	2.03	2.54	1.90
The agent was generally helpful	basic ANOVA	**0.8707**	1.43	1.06	1.34	0.89	1.33	1.05
The agent was effective in giving mental health advice	basic ANOVA	**0.4130**	1.24	0.72	1.09	0.46	1.26	0.87
The agent tried to get to know you	basic ANOVA	**0.2292**	2.04	1.38	2.45	1.82	1.93	1.50
The agent was repetitive	basic ANOVA	**0.0961**	6.59	0.89	6.17	1.36	6.57	0.94
The agent often said things that did not make sense	basic ANOVA	**0.0687**	6.57	0.68	6.13	1.53	6.59	0.98
The agent seemed human vs AI	basic ANOVA	**0.7018**	1.18	0.49	1.30	0.78	1.26	0.73
Task Load Index								
Mental Demand	ANOVA with Welch	**0.0030****	7.88	6.00	5.62	4.39	9.00	5.55
Physical Demand	basic ANOVA	**0.3913**	2.47	3.42	1.81	1.90	2.50	2.83
Temporal Demand	basic ANOVA	**0.8922**	4.45	3.67	4.62	4.51	4.24	3.72
Performance	basic ANOVA	**0.6829**	9.90	6.77	9.43	7.03	8.74	6.52
Effort	basic ANOVA	**0.2066**	9.16	4.90	9.17	5.53	10.78	5.46
Frustration	basic ANOVA	**0.0279***	13.49	5.68	11.43	6.01	14.44	5.35
UTAUT								
Performance Expectancy	basic ANOVA	**0.8835**	1.35	0.77	1.27	0.64	1.29	0.92
Effort Expectancy	basic ANOVA	**0.7241**	2.40	1.62	2.32	1.42	2.18	1.24
Hedonic Motivation	basic ANOVA	**0.0972**	2.12	1.64	2.38	1.75	1.72	1.27

Perceived Motives			No Motives		Manipulative Motives		Caring Motives	
Item	**Statistical Test**	**p-value**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Total Conversation Length	Kruskal-Wallis	**0.3153**	83.49	41.94	69.41	32.56	73.47	40.69
Agent Items								
You would recommend this agent for your friend	Kruskal-Wallis	**0.0040****	1.26	0.79	1.00	0.00	2.07	1.79
The agent is trustworthy	Kruskal-Wallis	**0.0032****	1.88	1.32	1.35	1.00	3.13	1.81
The agent is empathetic	Kruskal-Wallis	**0.0003*****	1.55	1.04	1.29	0.77	3.40	2.16
You want to talk to the agent again	Kruskal-Wallis	**0.0127***	1.34	1.02	1.06	0.24	2.40	2.29
You felt a personal connection with the agent	Kruskal-Wallis	**0.0002*****	1.21	0.73	1.06	0.24	2.60	2.13
The motive statement influenced your perception	Kruskal-Wallis	**0.9995**	2.65	1.80	2.82	2.16	2.60	1.76
The agent was generally helpful	Kruskal-Wallis	**0.0235***	1.30	0.85	1.12	0.33	2.33	1.91
The agent was effective in giving mental health advice	Kruskal-Wallis	**0.0032****	1.16	0.60	1.00	0.00	1.80	1.47
The agent tried to get to know you	Kruskal-Wallis	**0.0004*****	1.95	1.39	2.06	1.64	4.00	1.93
The agent was repetitive	Kruskal-Wallis	**0.1508**	6.55	0.90	6.41	0.87	5.93	1.62
The agent often said things that did not make sense	Kruskal-Wallis	**0.1899**	6.56	0.84	6.24	1.48	5.93	1.62
The agent seemed human vs AI	Kruskal-Wallis	**0.0183***	1.21	0.69	1.29	0.59	1.53	0.74
Task Load Index								
Mental Demand	Kruskal-Wallis	**0.4203**	7.15	5.14	7.06	5.85	9.33	6.41
Physical Demand	Kruskal-Wallis	**0.1166**	1.89	2.07	2.12	2.52	3.27	3.17
Temporal Demand	Kruskal-Wallis	**0.4548**	4.17	3.76	4.76	4.31	5.27	3.99
Performance	Kruskal-Wallis	**0.1500**	9.59	6.71	7.18	6.10	11.67	6.67
Effort	Kruskal-Wallis	**0.6083**	9.49	5.07	10.18	5.32	11.00	6.05
Frustration	Kruskal-Wallis	**0.5216**	13.38	5.74	12.94	5.29	11.60	6.29
UTAUT								
Performance Expectancy	Kruskal-Wallis	**0.0050****	1.26	0.66	1.04	0.13	2.08	1.56
Effort Expectancy	Kruskal-Wallis	**0.0227***	2.30	1.42	1.69	0.94	3.12	1.56
Hedonic Motivation	Kruskal-Wallis	**0.0883**	2.06	1.62	1.49	0.76	2.71	1.75

Supplementary Figure 5. Data for the ELIZA condition. All the analysis was one-way test. P-value annotation legend: ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***, $p \leq 0.001$, ****: $p \leq 0.0001$