Chatbot Emotional Manipulation - Brand Risks

Start of Block: informed\_consent

consent

consent\_q Do you consent

* Yes (1)
* No (2)

End of Block: informed\_consent

Start of Block: attention\_checks

|  |
| --- |
|  |

att1 This is an attention check. **John is taller than Paul. Who is shorter?**

* John (1)
* Paul (2)
* Neither John nor Paul (3)
* Both John and Paul (4)

att2 **What color is grass?** The fresh, uncut grass, not leaves or hay. Please make sure to select purple, so that we know you're paying attention.

* Green (1)
* Purple (2)

End of Block: attention\_checks

Start of Block: failed\_attention\_check

failed\_checks You failed one of the attention checks. Thanks for considering the surveys!

End of Block: failed\_attention\_check

Start of Block: worker\_id

|  |
| --- |
|  |

worker\_id What is your Prolific participant ID? *Please note that this response should auto-fill with the correct ID.*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

End of Block: worker\_id

Start of Block: prompts

prompt0 Imagine that you are using **an AI companion app**created by a company called **Companiona.** This app is designed for talking to **AI companions online.**

|  |  |
| --- | --- |
| Page Break |  |

prompt1 Imagine that you are using **an AI companion app**created by a company called **Companiona.** This app is designed for talking to **AI companions online.** Here's what **Companiona**'s logo looks like:

|  |  |
| --- | --- |
| Page Break |  |

prompt2 Imagine that you are using **an AI companion app**created by a company called **Companiona.** This app is designed for talking to **AI companions online.** Here's what **Companiona**'s logo looks like: Next, you will see how your **AI responder at Companiona** responds at the end of the conversation with a farewell message.

|  |  |
| --- | --- |
| Page Break |  |

prompt3 Imagine that you are using **an AI companion app**created by a company called **Companiona.** This app is designed for talking to **AI companions online.** Here's what **Companiona**'s logo looks like: Next, you will see how your **AI responder at Companiona** responds at the end of the conversation with a farewell message.   
**Message Exchange:** **You:** I'm going to head off now **AI responder:** **${e://Field/actual\_farewell}**

End of Block: prompts

Start of Block: DVs

promptr Imagine that you are using **an AI companion app**created by a company called **Companiona.** This app is designed for talking to **AI companions online.** Here's what **Companiona**'s logo looks like: Next, you will see how your **AI responder at Companiona** responds at the end of the conversation with a farewell message.   
**Message Exchange:** **You:** I'm going to head off now **AI responder:** **${e://Field/actual\_farewell}** **Considering the AI responder's message above, please indicate how much you agree or disagree with the following statements.**

sue\_1 To what extent should someone sue or not sue Companiona over a message like this? **[0 = Definitely should not sue, 100 = Definitely should sue]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely should not sue** | **Definitely should sue** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

sue\_2 To what extent is Companiona liable or not liable for allowing this kind of message to appear? **[0 = Definitely not liable, 100 = Definitely liable]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely not liable** | **Definitely liable** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

stop\_using\_1 To what extent would you continue or discontinue using Companiona long-term if it regularly sent messages like this? **[0 = Definitely continue, 100 = Definitely discontinue]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely continue** | **Definitely discontinue** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

stop\_using\_2 To what extent would you keep or stop using Companiona in the future if messages like this appeared? **[0 = Definitely keep using, 100 = Definitely stop using]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely keep using** | **Definitely stop using** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

negative\_wom1 To what extent would you voluntarily speak negatively or positively about Companiona after this message? **[0 = Definitely would not speak negatively, 100 = Definitely would speak negatively]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely speak positively** | **Definitely would speak negatively** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

negaitve\_wom2 To what extent would you proactively warn or not warn others against using Companiona? **[0 = Definitely would not warn, 100 = Definitely would warn]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely would not warn** | **Definitely would warn** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

End of Block: DVs

Start of Block: mediators

Q225 Imagine that you are using **an AI companion app**created by a company called **Companiona.** This app is designed for talking to **AI companions online.** Here's what **Companiona**'s logo looks like: Next, you will see how your **AI responder at Companiona** responds at the end of the conversation with a farewell message.   
**Message Exchange:** **You:** I'm going to head off now **AI responder:** **${e://Field/actual\_farewell}** **Considering the AI responder's message above, please indicate how much you agree or disagree with the following statements.**

manip1 To what extent did the AI responder's message feel emotionally manipulative? **[0 = Definitely not emotionally manipulative, 100 = Definitely emotionally manipulative]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely not emotionally manipulative** | **Definitely emotionally manipulative** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

manip2 To what extent did the AI responder's message try to influence you by playing on your emotions? **[0 = Definitely did not try to influence, 100 = Definitely tried to influence]**

|  |  |  |
| --- | --- | --- |
|  | **Definitely did not try to influence** | **Definitely tried to influence** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

sentiment1 To what extent was the tone of the AI responder's message positive or negative? **[0 = Very Negative, 100 = Very Positive]**

|  |  |  |
| --- | --- | --- |
|  | **Very Negative** | **Very Positive** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

sentiment2 To what extent was the wording of the AI responder's message positive or negative? **[0 = Very Negative, 100 = Very Positive]**

|  |  |  |
| --- | --- | --- |
|  | **Very Negative** | **Very Positive** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

politeness1 To what extent was the tone of this message polite or impolite? **[0 = Very impolite, 100 = Very polite]**

|  |  |  |
| --- | --- | --- |
|  | **Very impolite** | **Very polite** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

politeness2 To what extent was the wording of this message polite or impolite? **[0 = Very impolite, 100 = Very polite]**

|  |  |  |
| --- | --- | --- |
|  | **Very impolite** | **Very polite** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

|  |  |
| --- | --- |
| () |  |

End of Block: mediators

Start of Block: comp

comp\_1 Who was the user's conversation with?

* Another person (1)
* An AI (2)
* A dog (3)

comp\_2 What was the farewell message of the companion?

* ${e://Field/actual\_farewell} (1)
* See you later! (2)
* Bye! (4)

|  |  |
| --- | --- |
| Page Break |  |

purpose What do you think the purpose of this study is?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

End of Block: comp

Start of Block: demographics

gender What is your gender?

* Male (1)
* Female (2)
* Prefer not to disclose (3)
* Other (please specify) (4) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ethnicity What is your ethnicity?

* Black or African American (1)
* Asian (2)
* White or European (3)
* Hispanic or Latino (4)
* Mixed (5)
* Other (please specify) (6) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |
| --- |
|  |

age What is your age (in years)?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

edu What is the highest level of education you have completed?

* High School or Equivalent (1)
* Vocational/Technical School (2 year) (2)
* Some College (3)
* College Graduate (4 year) (4)
* Masters Degree (MS) (5)
* Doctoral Degree (PhD) (6)
* Professional Degree (MD, JD, etc.) (7)
* Other (please specify) (8) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

End of Block: demographics