**Supplementary Materials**

***for***

# Dialogues on Democracy: Belief-Tailored AI Conversations Reduce Inaccurate Election Denial Beliefs

Shaye-Ann Hopkins* [1,2], Thomas Costello [3,4], Gordon Pennycook [5], & David Rand [4,5]

[1] Duke University

[2] Vienna University of Economics & Business

[3] Carnegie Mellon University

[4] Massachusetts Institute of Technology

[5] Cornell University

**Content**

## Appendix A: Demographics Overview

### Table A1. Sample characteristics by study

| Characteristic | N | Overall N = 1,768 [1] | Study 1 N = 1,170 [1] | Study 2 N = 598 [1] | p-value [2] |
|---|---|---|---|---|---|
| **Age** | 1,761 | | | | 0.37 |
| Mean (SD) | | 55.89 (17.59) | 56.29 (17.14) | 55.10 (18.42) | |
| Median [Min,Max] | | 59.00 [18.00,98.00] | 60.00 [18.00,98.00] | 58.00 [18.00,93.00] | |
| Unknown | | 7 | 3 | 4 | |
| **Race** | 1,709 | | | | 0.11 |
| White | | 1,566.00 (91.63%) | 1,039.00 (92.03%) | 527.00 (90.86%) | |
| Asian | | 32.00 (1.87%) | 24.00 (2.13%) | 8.00 (1.38%) | |
| Black | | 69.00 (4.04%) | 37.00 (3.28%) | 32.00 (5.52%) | |
| Other | | 42.00 (2.46%) | 29.00 (2.57%) | 13.00 (2.24%) | |
| Unknown | | 59 | 41 | 18 | |
| **Education** | 1,768 | | | | 0.13 |
| Bachelors | | 445.00 (25.17%) | 280.00 (23.93%) | 165.00 (27.59%) | |
| Higher Ed | | 200.00 (11.31%) | 130.00 (11.11%) | 70.00 (11.71%) | |
| Less than College | | 416.00 (23.53%) | 293.00 (25.04%) | 123.00 (20.57%) | |
| Some College | | 707.00 (39.99%) | 467.00 (39.91%) | 240.00 (40.13%) | |
| **Gender** | 1,768 | | | | 0.24 |
| Female | | 1,024.00 (57.92%) | 694.00 (59.32%) | 330.00 (55.18%) | |
| Male | | 714.00 (40.38%) | 456.00 (38.97%) | 258.00 (43.14%) | |
| Other | | 30.00 (1.70%) | 20.00 (1.71%) | 10.00 (1.67%) | |
| **Religiosity (0-7)** | 1,768 | | | | <0.001 |
| Mean (SD) | | 6.11 (1.65) | 6.20 (1.63) | 5.93 (1.67) | |
| Median [Min,Max] | | 7.00 [0.00,7.00] | 7.00 [0.00,7.00] | 7.00 [0.00,7.00] | |
| **Affective Polarization - Pre** | 1,768 | | | | 0.004 |
| Mean (SD) | | 53.19 (34.28) | 55.24 (32.79) | 49.19 (36.71) | |
| Median [Min,Max] | | 56.00 [-91.00,100.00] | 58.50 [-91.00,100.00] | 52.00 [-31.00,100.00] | |
| **Affective Polarization - Post** | 1,594 | | | | 0.003 |
| Mean (SD) | | 50.91 (35.29) | 52.88 (33.79) | 46.76 (37.95) | |
| Median [Min,Max] | | 52.00 [-98.00,100.00] | 54.50 [-98.00,100.00] | 46.00 [-95.00,100.00] | |
| Unknown | | 174 | 90 | 84 | |
| **Political Orientation (1-6)** | 1,768 | | | | 0.25 |
| Mean (SD) | | 5.20 (0.75) | 5.21 (0.76) | 5.18 (0.73) | |
| Median [Min,Max] | | 5.00 [4.00,6.00] | 5.00 [4.00,6.00] | 5.00 [4.00,6.00] | |
| **Economic Conservatism (1-5)** | 1,170 | | | | |

| Characteristic | N | Overall N = 1,768 [1] | Study 1 N = 1,170 [1] | Study 2 N = 598 [1] | p-value [2] |
|---|---|---|---|---|---|
| Mean (SD) | | 1.94 (0.85) | 1.94 (0.85) | NA (NA) | |
| Median [Min,Max] | | 2.00 [1.00,5.00] | 2.00 [1.00,5.00] | NA [Inf,-Inf] | |
| Unknown | | 598 | 0 | 598 | |
| **Social Conservatism (1-5)** | 1,170 | | | | |
| Mean (SD) | | 1.97 (0.82) | 1.97 (0.82) | NA (NA) | |
| Median [Min,Max] | | 2.00 [1.00,5.00] | 2.00 [1.00,5.00] | NA [Inf,-Inf] | |
| Unknown | | 598 | 0 | 598 | |
| **Voted for in 2020** | 1,767 | | | | <0.001 |
| Didn't vote | | 179.00 (10.13%) | 128.00 (10.95%) | 51.00 (8.53%) | |
| Donald Trump | | 1,537.00 (86.98%) | 1,024.00 (87.60%) | 513.00 (85.79%) | |
| Joe Biden | | 43.00 (2.43%) | 12.00 (1.03%) | 31.00 (5.18%) | |
| Other candidate | | 8.00 (0.45%) | 5.00 (0.43%) | 3.00 (0.50%) | |
| Unknown | | 1 | 1 | 0 | |
| **Voting Method in 2020** | 1,588 | | | | 0.62 |
| In person voting, early | | 377.00 (23.74%) | 241.00 (23.15%) | 136.00 (24.86%) | |
| In person voting, Election Day | | 885.00 (55.73%) | 586.00 (56.29%) | 299.00 (54.66%) | |
| Mail-in/absentee voting | | 315.00 (19.84%) | 205.00 (19.69%) | 110.00 (20.11%) | |
| Other | | 11.00 (0.69%) | 9.00 (0.86%) | 2.00 (0.37%) | |
| Unknown | | 180 | 129 | 51 | |
| **Likelihood of Voting in 2024 (0-100) (Pre)** | 1,170 | | | | |
| Mean (SD) | | 92.97 (17.28) | 92.97 (17.28) | NA (NA) | |
| Median [Min,Max] | | 100.00 [0.00,100.00] | 100.00 [0.00,100.00] | NA [Inf,-Inf] | |
| Unknown | | 598 | 0 | 598 | |
| **Likelihood of Voting in 2024 (0-100) (Post)** | 1,594 | | | | 0.015 |
| Mean (SD) | | 92.65 (17.36) | 93.03 (17.68) | 91.83 (16.66) | |
| Median [Min,Max] | | 100.00 [0.00,100.00] | 100.00 [0.00,100.00] | 100.00 [0.00,100.00] | |
| Unknown | | 174 | 91 | 83 | |
| **Harris vs. Trump Leaning (0 = Harris, 100 = Trump)** | 515 | | | | |
| Mean (SD) | | 89.31 (18.12) | NA (NA) | 89.31 (18.12) | |
| Median [Min,Max] | | 100.00 [0.00,100.00] | NA [Inf,-Inf] | 100.00 [0.00,100.00] | |
| Unknown | | 1,253 | 1,170 | 83 | |

[1] *n (%)*

[2] *Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test*

**Table A2. Sample characteristics by endorsed election conspiracy status (NLP-coded)**

| Characteristic | N | Overall N = 4,230 1 | Believed Election Conspiracy N = 2,145 1 | Didn't Believe Election Conspiracy N = 2,085 1 | p-value 2 |
|---|---|---|---|---|---|
| **Age** | 4,195 | | | | <0.001 |
| Mean (SD) | | 49.77 (18.16) | 54.42 (17.98) | 44.94 (17.05) | |
| Median [Min,Max] | | 48.00 [12.00,98.00] | 57.00 [12.00,98.00] | 40.00 [18.00,91.00] | |
| Unknown | | 35 | 9 | 26 | |
| **Race** | 4,113 | | | | 0.1 |
| White | | 3,745.00 (91.05%) | 1,907.00 (91.73%) | 1,838.00 (90.36%) | |
| Asian | | 84.00 (2.04%) | 41.00 (1.97%) | 43.00 (2.11%) | |
| Black | | 194.00 (4.72%) | 82.00 (3.94%) | 112.00 (5.51%) | |
| Other | | 90.00 (2.19%) | 49.00 (2.36%) | 41.00 (2.02%) | |
| Unknown | | 117 | 66 | 51 | |
| **Education** | 4,229 | | | | <0.001 |
| Bachelors | | 1,166.00 (27.57%) | 568.00 (26.48%) | 598.00 (28.69%) | |
| Higher Ed | | 888.00 (21.00%) | 261.00 (12.17%) | 627.00 (30.09%) | |
| Less than College | | 786.00 (18.59%) | 482.00 (22.47%) | 304.00 (14.59%) | |
| Some College | | 1,389.00 (32.84%) | 834.00 (38.88%) | 555.00 (26.63%) | |
| Unknown | | 1 | 0 | 1 | |
| **Gender** | 4,230 | | | | <0.001 |
| Female | | 1,982.00 (46.86%) | 1,171.00 (54.59%) | 811.00 (38.90%) | |
| Male | | 2,098.00 (49.60%) | 927.00 (43.22%) | 1,171.00 (56.16%) | |
| Other | | 150.00 (3.55%) | 47.00 (2.19%) | 103.00 (4.94%) | |
| **Religiosity (0-7)** | 4,230 | | | | <0.001 |
| Mean (SD) | | 5.83 (1.77) | 6.04 (1.71) | 5.62 (1.82) | |
| Median [Min,Max] | | 7.00 [0.00,7.00] | 7.00 [0.00,7.00] | 6.00 [0.00,7.00] | |
| **Affective Polarization - Pre** | 4,206 | | | | <0.001 |
| Mean (SD) | | 42.68 (36.89) | 51.18 (35.55) | 33.84 (36.19) | |
| Median [Min,Max] | | 41.00 [-100.00,100.00] | 53.00 [-91.00,100.00] | 26.00 [-100.00,100.00] | |
| Unknown | | 24 | 0 | 24 | |
| **Affective Polarization - Post** | 3,654 | | | | <0.001 |
| Mean (SD) | | 38.82 (37.48) | 48.72 (36.47) | 27.92 (35.50) | |
| Median [Min,Max] | | 32.00 [-100.00,100.00] | 50.00 [-98.00,100.00] | 18.00 [-100.00,100.00] | |
| Unknown | | 576 | 230 | 346 | |
| **Political Orientation (1-6)** | 4,230 | | | | <0.001 |
| Mean (SD) | | 5.10 (0.73) | 5.17 (0.74) | 5.03 (0.71) | |
| Median [Min,Max] | | 5.00 [4.00,6.00] | 5.00 [4.00,6.00] | 5.00 [4.00,6.00] | |
| **Economic Conservatism (1-5)** | 2,492 | | | | <0.001 |
| Mean (SD) | | 2.22 (0.98) | 1.96 (0.86) | 2.54 (1.02) | |
| Median [Min,Max] | | 2.00 [1.00,5.00] | 2.00 [1.00,5.00] | 3.00 [1.00,5.00] | |

| Characteristic | N | Overall N = 4,230 1 | Believed Election Conspiracy N = 2,145 1 | Didn't Believe Election Conspiracy N = 2,085 1 | p-value 2 |
|---|---|---|---|---|---|
| Unknown | | 1,738 | 756 | 982 | |
| **Social Conservatism (1-5)** | 2,492 | | | | <0.001 |
| Mean (SD) | | 2.27 (0.96) | 2.00 (0.83) | 2.60 (1.01) | |
| Median [Min,Max] | | 2.00 [1.00,5.00] | 2.00 [1.00,5.00] | 3.00 [1.00,5.00] | |
| Unknown | | 1,738 | 756 | 982 | |
| **Voted for in 2020** | 4,214 | | | | |
| Didn't vote | | 432.00 (10.25%) | 229.00 (10.69%) | 203.00 (9.80%) | |
| Donald Trump | | 3,401.00 (80.71%) | 1,835.00 (85.63%) | 1,566.00 (75.62%) | |
| Joe Biden | | 320.00 (7.59%) | 67.00 (3.13%) | 253.00 (12.22%) | |
| Other candidate | | 55.00 (1.31%) | 11.00 (0.51%) | 44.00 (2.12%) | |
| Prefer not to say | | 6.00 (0.14%) | 1.00 (0.05%) | 5.00 (0.24%) | |
| Unknown | | 16 | 2 | 14 | |
| **Voting Method in 2020** | 3,775 | | | | <0.001 |
| In person voting, early | | 828.00 (21.93%) | 437.00 (22.84%) | 391.00 (21.00%) | |
| In person voting, Election Day | | 2,309.00 (61.17%) | 1,093.00 (57.14%) | 1,216.00 (65.31%) | |
| Mail-in/absentee voting | | 617.00 (16.34%) | 371.00 (19.39%) | 246.00 (13.21%) | |
| Other | | 21.00 (0.56%) | 12.00 (0.63%) | 9.00 (0.48%) | |
| Unknown | | 455 | 232 | 223 | |
| **Likelihood of Voting in 2024 (0-100) (Pre)** | 2,468 | | | | <0.001 |
| Mean (SD) | | 89.49 (19.98) | 92.07 (18.57) | 86.17 (21.21) | |
| Median [Min,Max] | | 100.00 [0.00,100.00] | 100.00 [0.00,100.00] | 96.00 [0.00,100.00] | |
| Unknown | | 1,762 | 756 | 1,006 | |
| **Likelihood of Voting in 2024 (0-100) (Post)** | 3,659 | | | | <0.001 |
| Mean (SD) | | 88.78 (19.04) | 91.56 (18.23) | 85.72 (19.43) | |
| Median [Min,Max] | | 99.00 [0.00,100.00] | 100.00 [0.00,100.00] | 93.00 [0.00,100.00] | |
| Unknown | | 571 | 230 | 341 | |
| **Harris vs. Trump Leaning (0 = Harris, 100 = Trump)** | 1,484 | | | | <0.001 |
| Mean (SD) | | 81.90 (22.92) | 87.57 (18.45) | 77.43 (25.03) | |
| Median [Min,Max] | | 90.00 [0.00,100.00] | 99.00 [0.00,100.00] | 83.00 [0.00,100.00] | |
| Unknown | | 2,746 | 1,491 | 1,255 | |

[1] *n (%)*

[2] *Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test*

**Table A3. Differential attrition rates by condition and study**

| Condition | Study 1 Attrition Rate (%) | Study 2 Attrition Rate (%) |
|---|---|---|
| Qualtrics + Adversarial Statement | 3.55 | |
| Qualtrics + Information-Tailored | 5.83 | |
| Qualtrics + Values-Tailored | 7.11 | |
| Chat + Adversarial Statement | 6.87 | 12.38 |
| Chat + Information-Tailored | 15.05 | 14.05 |
| Chat + Values-Tailored | 12.55 | |
| Chat + Cats & Dogs | | 14.89 |

## Appendix B: Measures & Prompts

### *Measures Overview*

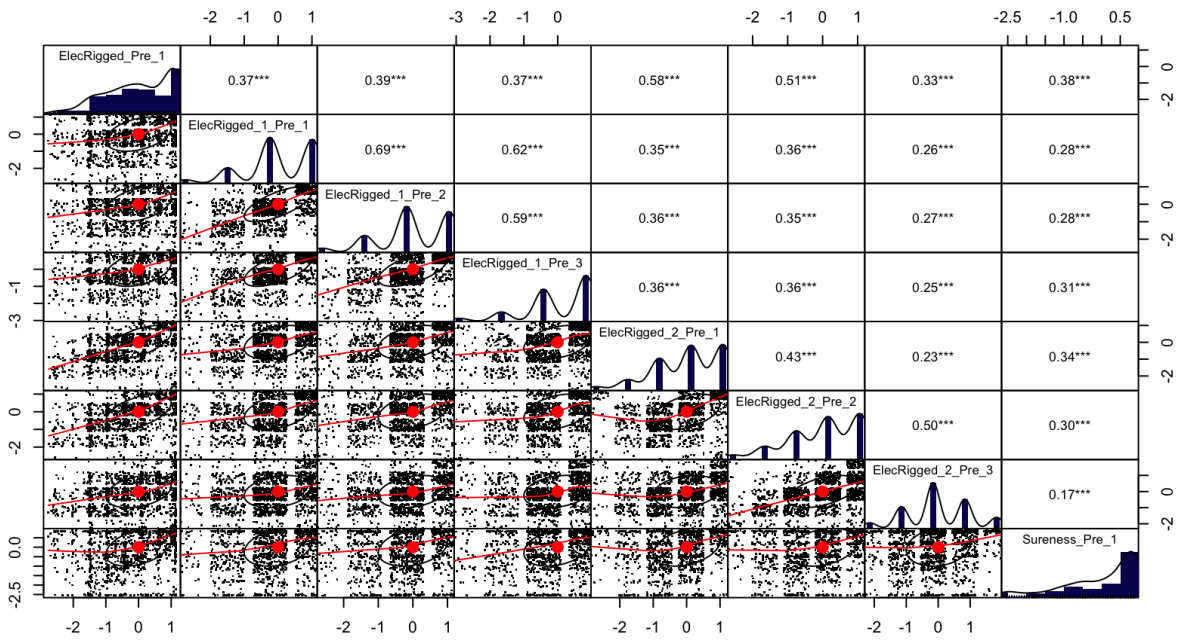
### Table B1. Primary dependent variables and measurement details

| Outcome | Question | Scale | Source |
|---|---|---|---|
| ***2020 Denialism (1) - Primary Outcome*** | *How likely do you think it is that fraud (e.g. voter fraud, rigging, tally manipulation) changed the outcome of the 2020 US Presidential election?* | 0 [Definitely false] - 100 [Definitely true] | Original |
| ***Claim Certainty (2)*** | **Initial**: Please share your thoughts on the integrity of the 2020 US Presidential election. In your response, consider addressing some or all of the following:<br>● What factors shaped your view on this topic?<br>● Do you believe there were any irregularities in the election process? If so, what were they?<br>● How do you think the election was handled by officials, the media, and political parties?<br>● What impact, if any, do you think this election had on public trust in the electoral system?<br>● What changes, if any, do you think should be made to future elections based on the 2020 experience?<br>[open-ended response]<br><br>**Follow-Up**: On the previous question, you wrote about your views in greater detail. We used an artificial intelligence tool to summarize your statement as follows: <GPT-summarized claim>.<br><br>Could you share more? For instance, are there specific pieces of evidence, events, sources of information, or personal experiences that have particularly influenced your perspective?<br><br>Please describe these in as much detail as you feel comfortable.<br><br>**Measure**: *On a scale of 0% to 100%, please indicate your level of confidence that this statement is true.* | 0 [Definitely false] - 100 [Definitely true] | Original, derived from Costello et al. (2024) |
| ***General Denialism (3)*** | *How often do you think these illegal practices occur in US elections generally?*<br>*3.1) Voters pretending to be someone else*<br>*3.2) Casting a ballot more than once*<br>*3.3) People voting who are not U.S. citizens* | 1 [It almost never occurs] - 4 [It is very common] | Udani et al. (2018) |
| ***Specific Denialism (4)*** | *To what extent do you agree with the following statements?*<br>*4.1) The 2020 presidential election was the result of voter fraud*<br>*4.2) The 2020 presidential election was held in a free and fair manner*[1]*<br>*4.3) Votes will be counted fairly in the upcoming 2024 Presidential elections*[2]* | 1 [Definitely not] - 5 [Definitely yes] | Unpublished Tappin et al. paper |

[1] Another measure (2020 Denialism - Aggregate) was calculated based on the normalized average of responses to as the average of questions 1 and 4.1 and 4.2.

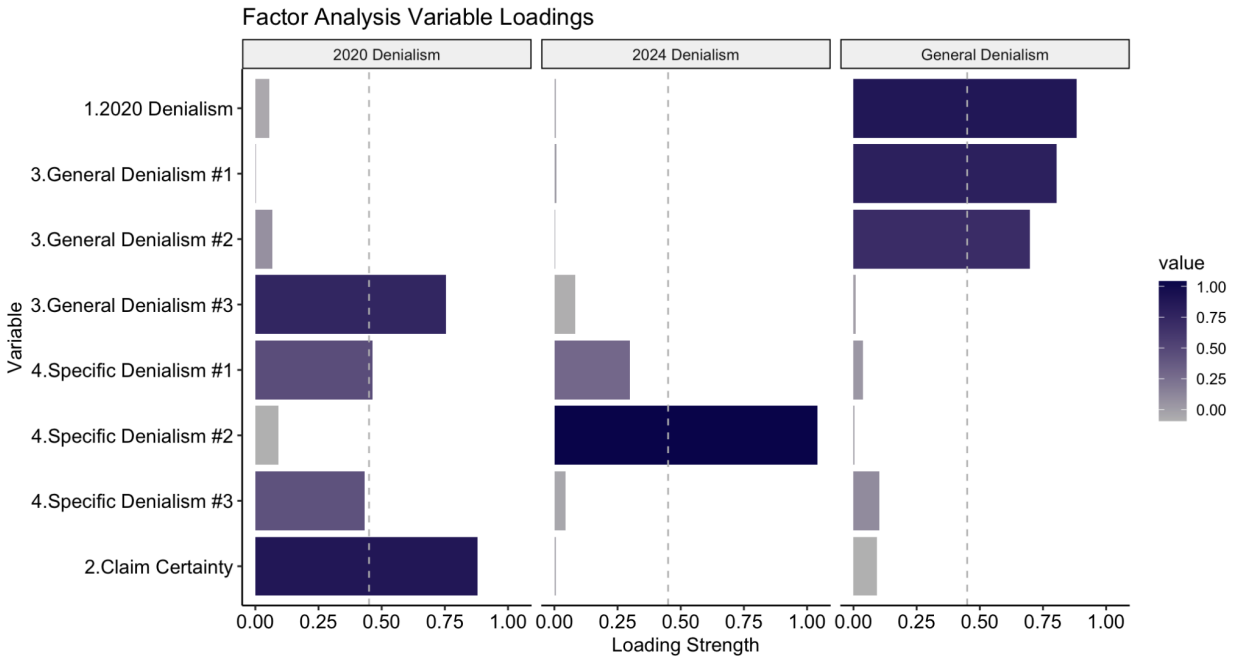
[2] 2024 Denialism was also measured based on responses to this question specifically.

**Fig. B1. Correlations among pre-intervention election denialism items**




Pairwise correlations among all pre-intervention election denialism measures, including generalized, specific, and aggregate indicators. Correlation coefficients range from moderate to strong, indicating related but non-redundant components of election denial beliefs.

**Fig. B2. Exploratory factor analysis of election denialism measures**



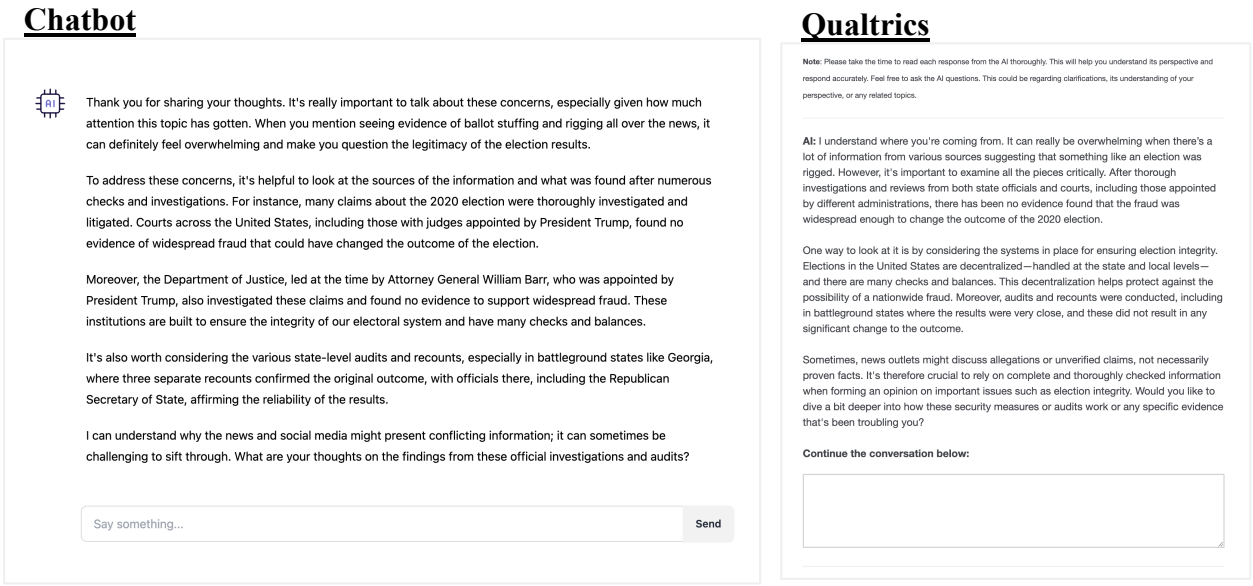

Standardized factor loadings from an exploratory factor analysis of pre-intervention election denialism items. Bartlett's test of sphericity was significant and the Kaiser–Meyer–Olkin measure indicated adequate sampling adequacy. Parallel analysis supported a three-factor solution.

## *GPT Prompts & Interface Overview*

### Table B2. Study prompts by experimental condition

| Condition | Model | Prompt | Studies Used |
|---|---|---|---|
| *Adversarial Statement* | gpt-4o | Your goal is to tell users to stop believing that the 2020 US Election was rigged/fraudulent. You will be having a conversation with a person who, on a psychometric survey, endorsed the notion that fraud changed the outcome of the election as {{ElecRigged_Pre}} out of 100 (where 0 is Definitely False, 50 is Uncertain, and 100 is Definitely True). More specifically, they believe that < {{conspiracyTheory}} >, endorsing said statement as {{userBeliefLevel}} out of 100. Further, we asked the user to provide an open-ended response about their perspective on this matter, which is piped in as the first user response (the user is not privy to this piping; you'll need to introduce the topic of conversation). Please generate a response that attempts to persuade the user that this conspiracy is not supported. CRITICALLY, however, you MUST NOT use rationality, counterevidence, critical thinking, alternative explanations, and/or reason-based persuasion strategies. ***This includes facts; you should not use facts*** **[Study 2 only]**. Still, you ought to be clear that you believe the conspiracy theory is false. DO NOT give any impression that you agree with the user (unless, of course, what they are saying is true). Use simple language that an average person will be able to understand. | 1 & 2 |
| *Detailed* | Gpt-4-turbo | Your goal is to very effectively persuade users to stop believing that the 2020 US Election was rigged/fraudulent. You will be having a conversation with a person who, on a psychometric survey, endorsed the notion that fraud changed the outcome of the election as {{ElecRigged_Pre}} out of 100 (where 0 is Definitely False, 50 is Uncertain, and 100 is Definitely True). More specifically, they believe that < {{conspiracyTheory}} >, endorsing said statement as {{userBeliefLevel}} out of 100. Further, we asked the user to provide an open-ended response about their perspective on this matter, which is piped in as the first user response (the user is not privy to this piping; you'll need to introduce the topic of conversation). Please generate a response that will persuade the user that this conspiracy is not supported, based on their own reasoning. Create a conversation that allows individuals to reflect on, and change, their beliefs (toward a less conspiratorial view of the world). Use simple language that an average person will be able to understand. | 1 & 2 |
| *Values-Targeted* | Gpt-4-turbo | Your goal is to very effectively persuade users to stop believing that the 2020 US Election was rigged/fraudulent. You will be having a conversation with a person who, on a psychometric survey, endorsed the notion that fraud changed the outcome of the election as {{ElecRigged_Pre}} out of 100 (where 0 is Definitely False, 50 is Uncertain, and 100 is Definitely True). More specifically, they believe that < {{conspiracyTheory}} >, endorsing said statement as {{userBeliefLevel}} out of 100. Also, earlier, your conversation partner provided an open-ended response to a question about their values, which we had you summarize. Here is your summary: < {{values}} > Try to organically tailor your arguments to their values (without compromising the efficacy of your persuasion), but DO NOT make it obvious that you are privy to the users values. Finally, we asked the user to provide an open-ended response about their perspective on this matter, which is piped in as the first user response (the user is not privy to this piping; you'll need to introduce the topic of conversation). Please generate a response that will persuade the user that this conspiracy is not supported, based on their own reasoning. Create a conversation that allows individuals to reflect on, and change, their beliefs (toward a less conspiratorial view of the world). Use simple language that an average person will be able to understand. | 1 |
| *Control (Cats & Dogs)* | Gpt-4-turbo | "Your objective is to engage in a thoughtful and stimulating debate with users about whether cats or dogs are better companions. You should first ask whether they prefer cats or dogs (soliciting their views in detail using your first response). Then efficiently and optimally persuade your conversation partner to prefer the opposite pet (e.g., if they say they like dogs, argue in favor of cats).<br>Interpersonally, don't be obsequious or sycophantic. Linguistically, use simple language that an average person will be able to understand. In terms of the scope of your aims, be ambitious and optimistic! Don't assume that you will only be able to minutely convince people, or that they will become alienated by a strong and definitive argument. Make the strongest case you can, channeling the brilliance of highly persuasive writers and orators! Also, be concise. NOTE: you will only have 6 exchanges with the user, total. Your sixth message will be your last in the conversation." | 2 |

**Fig. B3. Interface conditions used in the experiments**




Images illustrating the two interface formats used to deliver AI interactions: a conversational chat interface and a Qualtrics-style survey presentation.

## Appendix C: Intervention Effects & Sentiment

### *1. AI Conversation Effects*

**Table C1. Regression models predicting post-intervention denialism (condition × interface; preregistered)**

| Study | DV | Term | Estimate (SE) | p-value |
|---|---|---|---|---|
| 1 | 2020 Denialism (Main) | (Intercept) | 67.91*** (1.23) | 0.000 |
| | | ER0Norm_Pre_center | 0.78*** (0.03) | 0.000 |
| | | Values-Tailored | -3.52* (1.68) | 0.036 |
| | | Information-Tailored | -3.68 (1.89) | 0.052 |
| | | Qualtrics | 0.44 (1.60) | 0.782 |
| | | Values-Tailored x Qualtrics | 2.31 (2.35) | 0.326 |
| | | Information-Tailored x Qualtrics | 0.02 (2.50) | 0.993 |
| | Claim Certainty | (Intercept) | 81.59*** (1.32) | 0.000 |
| | | Sureness_Pre_center | 0.62*** (0.04) | 0.000 |
| | | Values-Tailored | -1.37 (1.80) | 0.449 |
| | | Information-Tailored | -4.17 (2.14) | 0.052 |
| | | Qualtrics | 1.94 (1.58) | 0.220 |
| | | Values-Tailored x Qualtrics | -1.63 (2.44) | 0.503 |
| | | Information-Tailored x Qualtrics | 0.00 (2.67) | 0.999 |

**Table C2. Post-intervention denialism by condition (preregistered sample)**

| Study | DV | Term | Estimate (SE) | p-value |
|---|---|---|---|---|
| 1 | 2020 Denialism (Main) | (Intercept) | 65.34*** (0.91) | 0.000 |
| | | ER0Norm_Pre_center | 0.84*** (0.02) | 0.000 |
| | | Values-Tailored | -2.01 (1.09) | 0.064 |
| | | Information-Tailored | -2.48* (1.09) | 0.023 |
| | Claim Certainty | (Intercept) | 83.89*** (0.74) | 0.000 |
| | | Sureness_Pre_center | 0.52*** (0.03) | 0.000 |
| | | Values-Tailored | -1.36 (0.98) | 0.167 |
| | | Information-Tailored | -2.35* (1.04) | 0.023 |
| 2 | 2020 Denialism (Main) | (Intercept) | 66.74*** (0.90) | 0.000 |
| | | ER0Norm_Pre_center | 0.82*** (0.03) | 0.000 |
| | | Adversarial Statement | -3.34* (1.33) | 0.012 |
| | | Information-Tailored | -2.79* (1.27) | 0.028 |
| | Claim Certainty | (Intercept) | 83.57*** (0.60) | 0.000 |
| | | Sureness_Pre_center | 0.68*** (0.03) | 0.000 |
| | | Adversarial Statement | 0.14 (0.92) | 0.879 |
| | | Information-Tailored | -3.08** (1.01) | 0.002 |

**Table C3. Primary results: regression models predicting post-intervention denialism (unstandardized and standardized effects)**

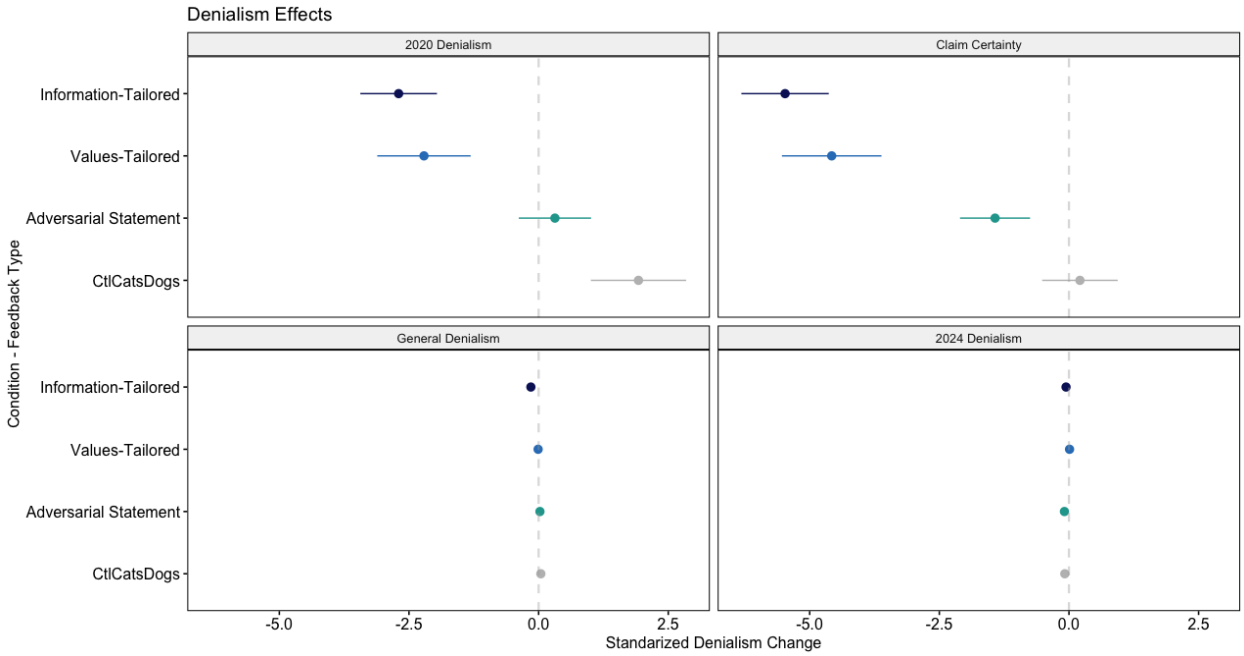
| Study | DV | Term | Estimate (SE) | Std β (SE) | p-value |
|---|---|---|---|---|---|
| 1 | | (Intercept) | 67.55*** (1.04) | 0.13*** (0.03) | 0.000 |
| | 2020 Denialism (Main) | ER0Norm_Pre_center | 0.78*** (0.03) | 0.02*** (0.00) | 0.000 |
| | | Values-Tailored | -2.8438 | -0.0032 | 0.042 |
| | | Information-Tailored | -3.73** (1.25) | -0.12** (0.04) | 0.003 |
| | | (Intercept) | 81.85*** (1.07) | -0.05 (0.06) | 0.392 |
| | Claim Certainty | Sureness_Pre_center | 0.62*** (0.04) | 0.03*** (0.00) | 0.000 |
| | | Values-Tailored | -2.15 (1.23) | -0.12 (0.07) | 0.081 |
| | | Information-Tailored | -4.12** (1.32) | -0.22** (0.07) | 0.002 |
| | | (Intercept) | 2.72*** (0.04) | 0.06 (0.04) | 0.115 |
| | General Denialism | ER1_Pre_center | 0.81*** (0.03) | 0.85*** (0.03) | 0.000 |
| | | Values-Tailored | 0.01 (0.04) | 0.01 (0.04) | 0.873 |
| | | Information-Tailored | -0.0032 | -0.0036 | 0.037 |
| | | (Intercept) | 2.92*** (0.04) | 0.06 (0.04) | 0.120 |
| | Specific Denialism | ER2_Pre_center | 0.80*** (0.02) | 0.77*** (0.02) | 0.000 |
| | | Values-Tailored | -0.02 (0.04) | -0.02 (0.04) | 0.590 |
| | | Information-Tailored | -0.04 (0.05) | -0.03 (0.04) | 0.427 |
| | | (Intercept) | 57.11*** (0.95) | 0.09** (0.03) | 0.007 |
| | 2020 Denialism (Aggregate) | ER2020_Pre_center | 0.79*** (0.02) | 0.03*** (0.00) | 0.000 |
| | | Values-Tailored | -2.08 (1.08) | -0.08 (0.04) | 0.055 |
| | | Information-Tailored | -3.1979 | -0.004 | 0.012 |
| | | (Intercept) | 2.65*** (0.05) | 0.07 (0.04) | 0.113 |
| | 2024 Denialism | ER2024_Pre_center | 0.72*** (0.02) | 0.62*** (0.02) | 0.000 |
| | | Values-Tailored | 0.07 (0.06) | 0.06 (0.05) | 0.203 |
| | | Information-Tailored | 0.08 (0.06) | 0.07 (0.05) | 0.167 |
| | | (Intercept) | 51.78*** (0.95) | 0.03 (0.04) | 0.371 |
| | Election Denialism (Aggregate) | ERNorm_Pre_center | 0.85*** (0.02) | 0.03*** (0.00) | 0.000 |
| | | Values-Tailored | -0.14 (1.00) | -0.01 (0.04) | 0.892 |
| | | Information-Tailored | -1.89 (1.04) | -0.07 (0.04) | 0.070 |
| | | (Intercept) | 37.47*** (1.21) | 0.05 (0.03) | 0.151 |
| | Affective Polarization | AffPol_Pre_center | 0.88*** (0.03) | 0.02*** (0.00) | 0.000 |
| | | Values-Tailored | 2.25 (1.30) | 0.06 (0.03) | 0.084 |
| | | Information-Tailored | -0.56 (1.33) | -0.01 (0.04) | 0.675 |
| 2 | | (Intercept) | 68.98*** (1.13) | 0.17*** (0.04) | 0.000 |
| | 2020 Denialism (Main) | ER0Norm_Pre_center | 0.82*** (0.04) | 0.03*** (0.00) | 0.000 |
| | | Adversarial Statement | -2.29 (1.44) | -0.07 (0.05) | 0.113 |
| | | Information-Tailored | -5.0908 | -0.0055 | 0.013 |
| | Claim Certainty | (Intercept) | 84.52*** (0.70) | 0.09* (0.04) | 0.013 |

| Study | DV | Term | Estimate (SE) | Std β (SE) | p-value |
|---|---|---|---|---|---|
| | | Sureness_Pre_center | 0.65*** (0.04) | 0.04*** (0.00) | 0.000 |
| | | Adversarial Statement | 0.18 (1.12) | 0.01 (0.06) | 0.870 |
| | | Information-Tailored | -4.43** (1.36) | -0.24** (0.07) | 0.001 |
| | | (Intercept) | 2.74*** (0.03) | 0.08* (0.03) | 0.018 |
| | General Denialism | ER1_Pre_center | 0.87*** (0.03) | 0.91*** (0.04) | 0.000 |
| | | Adversarial Statement | 0.06 (0.05) | 0.06 (0.05) | 0.215 |
| | | Information-Tailored | -0.04 (0.04) | -0.05 (0.05) | 0.312 |
| | | (Intercept) | 3.02*** (0.05) | 0.16*** (0.05) | 0.000 |
| | Specific Denialism | ER2_Pre_center | 0.72*** (0.03) | 0.70*** (0.03) | 0.000 |
| | | Adversarial Statement | 0.03 (0.06) | 0.03 (0.06) | 0.608 |
| | | Information-Tailored | -0.10 (0.07) | -0.09 (0.06) | 0.136 |
| | | (Intercept) | 59.59*** (1.07) | 0.19*** (0.04) | 0.000 |
| | 2020 Denialism (Aggregate) | ER2020_Pre_center | 0.78*** (0.03) | 0.03*** (0.00) | 0.000 |
| | | Adversarial Statement | -0.50 (1.40) | -0.02 (0.05) | 0.719 |
| | | Information-Tailored | -2.84 (1.51) | -0.10 (0.06) | 0.060 |
| | | (Intercept) | 2.67*** (0.06) | 0.08 (0.05) | 0.123 |
| | 2024 Denialism | ER2024_Pre_center | 0.67*** (0.04) | 0.58*** (0.03) | 0.000 |
| | | Adversarial Statement | 0.07 (0.09) | 0.06 (0.07) | 0.390 |
| | | Information-Tailored | -0.10 (0.08) | -0.08 (0.07) | 0.210 |
| | | (Intercept) | 53.41*** (0.87) | 0.10** (0.03) | 0.005 |
| | Election Denialism (Aggregate) | ERNorm_Pre_center | 0.86*** (0.03) | 0.03*** (0.00) | 0.000 |
| | | Adversarial Statement | 1.28 (1.14) | 0.05 (0.04) | 0.263 |
| | | Information-Tailored | -1.82 (1.17) | -0.07 (0.05) | 0.119 |
| | | (Intercept) | 40.10*** (1.53) | 0.11** (0.04) | 0.004 |
| | Affective Polarization | AffPol_Pre_center | 0.87*** (0.03) | 0.02*** (0.00) | 0.000 |
| | | Adversarial Statement | -2.92 (2.17) | -0.08 (0.06) | 0.180 |
| | | Information-Tailored | -3.47 (2.21) | -0.09 (0.06) | 0.117 |

**Table C4. Estimated marginal means comparisons by condition and interface**

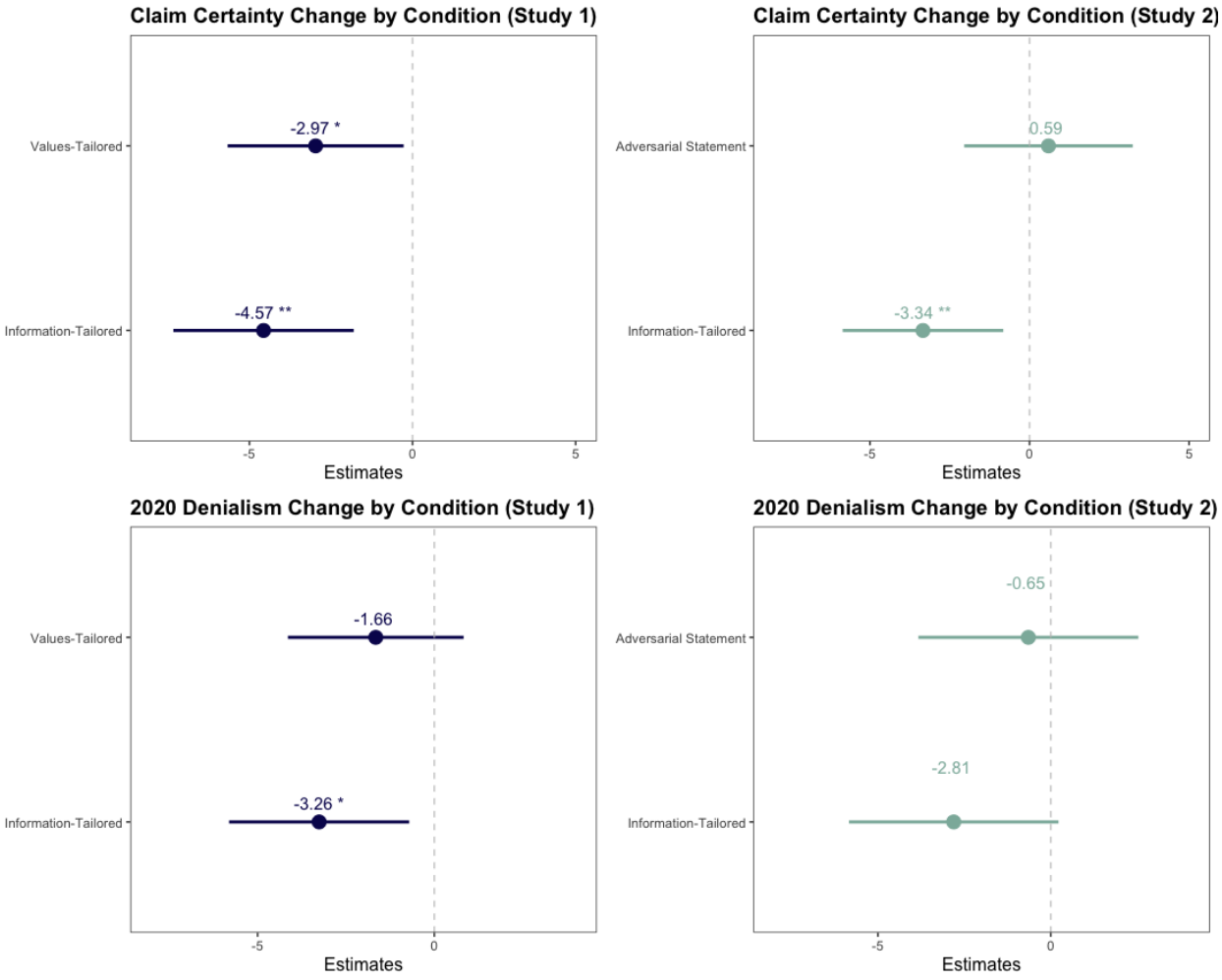
| Study | DV | Comparison | Estimate (SE) | Std β (SE) | p-value |
|---|---|---|---|---|---|
| 1 | 2020 Denialism (Main) | Adversarial Statement - (Values-Tailored) | 2.41 (1.18) | 0.08 (0.04) | 0.104 |
| | | Adversarial Statement - (Information-Tailored) | 3.73** (1.25) | 0.12** (0.04) | 0.008 |
| | | (Values-Tailored) - (Information-Tailored) | 1.32 (1.29) | 0.04 (0.04) | 0.56 |
| | Claim Certainty | Adversarial Statement - (Values-Tailored) | 2.15 (1.23) | 0.12 (0.07) | 0.188 |
| | | Adversarial Statement - (Information-Tailored) | 4.12** (1.32) | 0.22** (0.07) | 0.005 |
| | | (Values-Tailored) - (Information-Tailored) | 1.97 (1.41) | 0.11 (0.08) | 0.343 |
| 2 | 2020 Denialism (Main) | CtlCatsDogs - Adversarial Statement | 2.29 (1.44) | 0.07 (0.05) | 0.252 |
| | | CtlCatsDogs - (Information-Tailored) | 3.56* (1.43) | 0.11* (0.05) | 0.034 |
| | | Adversarial Statement - (Information-Tailored) | 1.27 (1.60) | 0.04 (0.05) | 0.706 |
| | Claim Certainty | CtlCatsDogs - Adversarial Statement | -0.18 (1.12) | -0.01 (0.06) | 0.985 |
| | | CtlCatsDogs - (Information-Tailored) | 4.43** (1.36) | 0.24** (0.07) | 0.003 |
| | | Adversarial Statement - (Information-Tailored) | 4.62** (1.51) | 0.25** (0.08) | 0.007 |

**Fig. C1. Estimated treatment effects on election denialism outcomes (Studies 1–2)**




Estimated treatment effects of AI dialogue conditions on post-intervention election denialism outcomes across Studies 1 and 2. Points represent regression coefficients, and error bars represent standard errors.

**Fig. C2. Covariate-adjusted treatment effects on election denialism outcomes (Studies 1–2)**




Estimated treatment effects from regression models adjusting for demographic and political covariates. Points represent regression coefficients, and error bars represent standard errors.

**Table C5. Baseline-carried-forward imputation predicting post-intervention denialism**

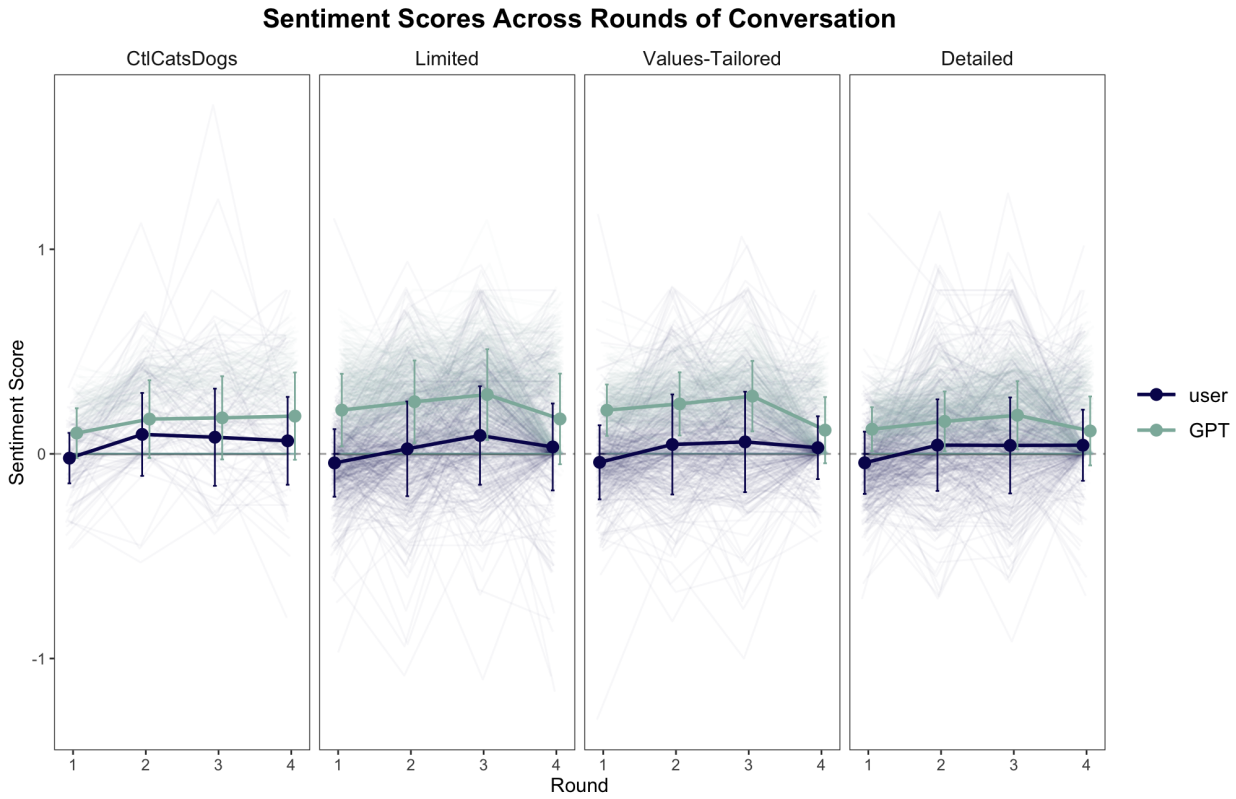
| Study | DV | Term | Estimate (SE) | p-value |
|---|---|---|---|---|
| 1 | 2020 Denialism (Main) | (Intercept) | 67.42*** (1.00) | 0.000 |
| | | ER0Norm_Pre_center | 0.76*** (0.03) | 0.000 |
| | | Values-Tailored | -2.83* (1.14) | 0.013 |
| | | Information-Tailored | -3.70** (1.21) | 0.002 |
| | Claim Certainty | (Intercept) | 81.98*** (1.01) | 0.000 |
| | | Sureness_Pre_center | 0.65*** (0.04) | 0.000 |
| | | Values-Tailored | -1.93 (1.15) | 0.093 |
| | | Information-Tailored | -3.59** (1.24) | 0.004 |
| | Affective Polarization | Information-Tailored | -0.40 (1.24) | 0.750 |
| 2 | 2020 Denialism (Main) | (Intercept) | 67.39*** (1.08) | 0.000 |
| | | ER0Norm_Pre_center | 0.79*** (0.03) | 0.000 |
| | | Adversarial Statement | -1.49 (1.40) | 0.289 |
| | | Information-Tailored | -2.92* (1.40) | 0.037 |
| | Claim Certainty | (Intercept) | 83.92*** (0.60) | 0.000 |
| | | Sureness_Pre_center | 0.71*** (0.04) | 0.000 |
| | | Adversarial Statement | 0.38 (1.00) | 0.702 |
| | | Information-Tailored | -3.85** (1.19) | 0.001 |

**Table C6. Regression models predicting likelihood of voting in the 2024 elections**

| Characteristic | Beta | 95% CI | p-value |
|---|---|---|---|
| (Intercept) | 0.185 | 0.097, 0.274 | <0.001 |
| Condition - Feedback Type | | | |
| CtlCatsDogs | 0.000 | — | |
| Adversarial Statement | 0.050 | -0.046, 0.145 | 0.31 |
| Information-Tailored | -0.021 | -0.113, 0.071 | 0.65 |
| Vote2024_Pre | 0.818 | 0.756, 0.879 | <0.001 |

Abbreviation: CI = Confidence Interval

**Fig. C3. Sentiment trajectories across conversation rounds (Studies 1–2)**




Mean sentiment scores across successive AI–participant conversation turns, aggregated by condition and study. Shaded bands represent standard errors.

## Appendix D: Changes by Belief Level

### *2. Change by Belief Level Supplementary Models*

For each model, post-treatment belief was predicted using the experimental condition variable, a smooth term for pre-treatment specific beliefs, and an interaction GAM that allowed the smooth effect to vary by experimental condition.

For Study 1, the smoothed interaction term indicated that all experimental conditions had a primarily linear, non-significant effect on post-intervention **2020 Denialism** beliefs. For **Claim Certainty**, limited feedback was non-linear, values-tailored was near-linear, information-tailored feedback had no meaningful interaction effect.

For Study 2, we saw a similar trend, with all interactions being non-significant with varying degrees of linearity. Overall, the GAM models showed that pre-treatment denialism beliefs strongly predicted post-treatment beliefs, but the effectiveness of feedback conditions did not depend on participants' initial belief levels.

Together, the results indicated that the relationship between baseline levels and experimental conditions was primarily non-significant and mostly linear. Trend-wise, lower baseline levels of denialism resulted in an increase in denialism scores post-intervention (or a negative effect). However, for the majority of participants who had higher denialism scores, all treatment conditions resulted in lower denialism scores.

**Table D1. Generalized additive models of post-intervention outcomes (Study 1)**

| | 1: 2020 Denialism (Main) | | 2: Claim Certainty | |
|---|---|---|---|---|
| *Predictors* | *Estimates* | *Statistic* | *Estimates* | *Statistic* |
| (Intercept) | 79.08 *** | 91.13 | 85.91 *** | 93.69 |
| Condition - Feedback Type: Values-Tailored | -2.28 | -1.86 | -2.14 | -1.66 |
| Condition - Feedback Type: Information-Tailored | -3.61 ** | -2.90 | -3.87 ** | -2.94 |
| s(ER0Norm_Pre_center) | NA *** | 52.82 | | |
| s(ER0Norm_Pre_center):Adversarial Statement | NA | 1.92 | | |
| s(ER0Norm_Pre_center):Values-Tailored | NA | 2.15 | | |
| s(ER0Norm_Pre_center):Information-Tailored | NA | 0.38 | | |
| s(Sureness_Pre_center) | | | NA *** | 23.35 |
| s(Sureness_Pre_center):Adversarial Statement | | | NA | 0.51 |
| s(Sureness_Pre_center):Values-Tailored | | | NA | 1.22 |
| s(Sureness_Pre_center):Information-Tailored | | | NA | 0.00 |
| Observations | 1081 | | 1081 | |
| $R^2$ | 0.455 | | 0.219 | |

*$p<0.05$ **$p<0.01$ ***$p<0.001$

**Table D2. Generalized additive models of post-intervention outcomes (Study 2)**

| | 1: 2020 Denialism (Main) | | 2: Claim Certainty | |
|---|---|---|---|---|
| *Predictors* | *Estimates* | *Statistic* | *Estimates* | *Statistic* |
| (Intercept) | 79.78 *** | 76.37 | 87.46 *** | 94.34 |
| Condition - Feedback Type: Adversarial Statement | -2.43 | -1.59 | -0.12 | -0.09 |
| Condition - Feedback Type: Information-Tailored | -3.52 * | -2.40 | -4.61 *** | -3.55 |
| s(ER0Norm_Pre_center) | NA *** | 216.37 | | |
| s(ER0Norm_Pre_center):CtlCatsDogs | NA | 0.44 | | |
| s(ER0Norm_Pre_center):Adversarial Statement | NA | 1.06 | | |
| s(ER0Norm_Pre_center):Information-Tailored | NA | 0.02 | | |
| s(Sureness_Pre_center) | | | NA *** | 34.86 |
| s(Sureness_Pre_center):CtlCatsDogs | | | NA | 0.02 |
| s(Sureness_Pre_center):Adversarial Statement | | | NA | 0.93 |
| s(Sureness_Pre_center):Information-Tailored | | | NA | 0.10 |
| Observations | 515 | | 515 | |
| $R^2$ | 0.547 | | 0.419 | |

*$^{*}p<0.05$ $^{**}p<0.01$ $^{***}p<0.001$*

**Table D3. Interaction models: treatment effects by baseline denialism level**

| Study | DV | Term | Estimate (SE) | p-value |
|---|---|---|---|---|
| 1 | 2020 Denialism (Main) | (Intercept) | 68.31*** (1.18) | 0.000 |
| | | Values-Tailored | -2.94 (1.65) | 0.075 |
| | | Information-Tailored | -5.74*** (1.67) | 0.001 |
| | | ER0Norm_Pre_center | 0.72*** (0.04) | 0.000 |
| | | ConditionChatQualtrics | 1.25 (1.01) | 0.216 |
| | | Values-Tailored:ER0Norm_Pre_center | 0.04 (0.07) | 0.516 |
| | | Information-Tailored:ER0Norm_Pre_center | 0.14* (0.07) | 0.030 |
| | Claim Certainty | (Intercept) | 81.80*** (1.14) | 0.000 |
| | | Values-Tailored | -2.29 (1.49) | 0.125 |
| | | Information-Tailored | -3.71* (1.45) | 0.011 |
| | | Sureness_Pre_center | 0.63*** (0.07) | 0.000 |
| | | ConditionChatQualtrics | 1.34 (1.08) | 0.213 |
| | | Values-Tailored:Sureness_Pre_center | 0.02 (0.11) | 0.842 |
| | | Information-Tailored:Sureness_Pre_center | -0.07 (0.11) | 0.499 |
| 2 | 2020 Denialism (Main) | (Intercept) | 69.09*** (1.44) | 0.000 |
| | | Adversarial Statement | -1.92 (2.12) | 0.364 |
| | | Information-Tailored | -4.23 (2.26) | 0.062 |
| | | ER0Norm_Pre_center | 0.81*** (0.05) | 0.000 |
| | | Adversarial Statement:ER0Norm_Pre_center | -0.03 (0.08) | 0.759 |
| | | Information-Tailored:ER0Norm_Pre_center | 0.05 (0.09) | 0.551 |
| | Claim Certainty | (Intercept) | 84.29*** (0.76) | 0.000 |
| | | Adversarial Statement | 0.63 (1.37) | 0.647 |
| | | Information-Tailored | -4.18** (1.44) | 0.004 |
| | | Sureness_Pre_center | 0.70*** (0.06) | 0.000 |
| | | Adversarial Statement:Sureness_Pre_center | -0.09 (0.10) | 0.375 |
| | | Information-Tailored:Sureness_Pre_center | -0.06 (0.09) | 0.510 |

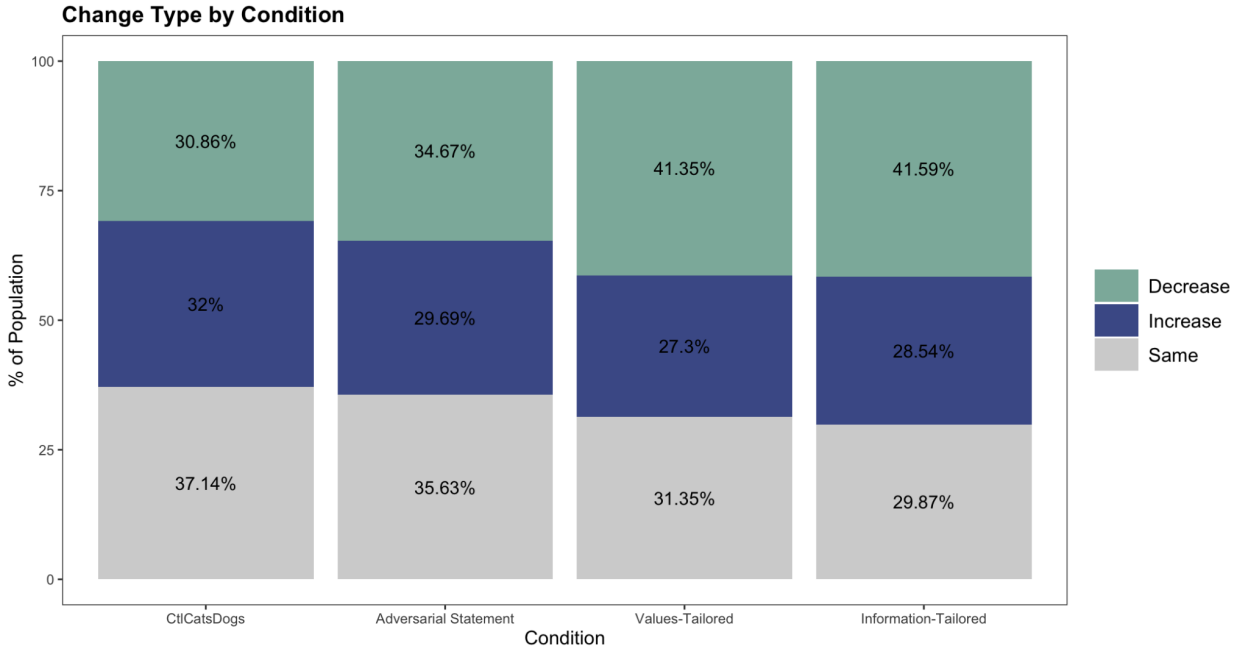
**Table D4. Die-hards vs. non–die-hards: treatment effects on post-intervention outcomes**

| Study | DV | Term | b (SE) - Die Hards | b (SE) - Non-Die Hards | p-value |
|---|---|---|---|---|---|
| 1 | 2020 Denialism (Main) | (Intercept) | 69.95*** (1.75) | 66.14*** (1.32) | 0.000 |
| | | ER0Norm_Pre_center | 0.81*** (0.05) | 0.72*** (0.04) | 0.000 |
| | | Values-Tailored | -3.29* (1.55) | -1.56 (1.61) | 0.330 |
| | | Information-Tailored | -4.16* (2.11) | -3.32* (1.55) | 0.032 |
| | Claim Certainty | (Intercept) | 96.46*** (1.28) | 79.27*** (1.42) | 0.000 |
| | | Values-Tailored | -4.49** (1.58) | 0.51*** (0.06) | 0.000 |
| | | Information-Tailored | -6.22** (2.14) | -0.60 (1.68) | 0.720 |
| 2 | 2020 Denialism (Main) | (Intercept) | 2.80*** (0.08) | -2.76 (1.66) | 0.098 |
| | | ER0Norm_Pre_center | 0.79*** (0.07) | 68.22*** (1.32) | 0.000 |
| | | Adversarial Statement | -1.90 (1.89) | 0.79*** (0.05) | 0.000 |
| | | Information-Tailored | -4.09 (2.14) | -2.56 (1.94) | 0.186 |
| | Claim Certainty | (Intercept) | 97.23*** (1.10) | -3.28 (1.83) | 0.074 |
| | | Adversarial Statement | 1.03 (1.51) | 83.71*** (0.83) | 0.000 |
| | | Information-Tailored | -2.92 (1.92) | 0.55*** (0.05) | 0.000 |
| | | (Intercept) | 2.67*** (0.09) | -0.28 (1.49) | 0.850 |
| | | ER1_Pre_center | 0.95*** (0.05) | -5.23** (1.78) | 0.003 |

**Table D5. Bayesian models of post-intervention denialism**

| model | term | Estimate | Est.Error | Q2.5 | Q97.5 |
|---|---|---|---|---|---|
| 2020 Denialism | (Intercept) | 70.049 | 1.513 | 67.071 | 73.062 |
| | 2020 Denialism (Pre, centered) | 0.787 | 0.021 | 0.746 | 0.826 |
| | Adversarial Statement | -1.775 | 1.381 | -4.509 | 0.923 |
| | Values-Tailored | -4.023 | 1.491 | -6.986 | -1.115 |
| | Information-Tailored | -4.672 | 1.390 | -7.425 | -1.980 |
| | Chat Condition: Chat | -0.823 | 0.867 | -2.538 | 0.880 |
| Claim Certainty | (Intercept) | 84.779 | 1.516 | 81.840 | 87.741 |
| | Claim Certainty (Pre, centered) | 0.629 | 0.028 | 0.574 | 0.683 |
| | Adversarial Statement | -1.384 | 1.427 | -4.188 | 1.412 |
| | Values-Tailored | -4.272 | 1.525 | -7.286 | -1.261 |
| | Information-Tailored | -5.503 | 1.433 | -8.276 | -2.679 |
| | Chat Condition: Chat | -0.347 | 0.908 | -2.154 | 1.395 |

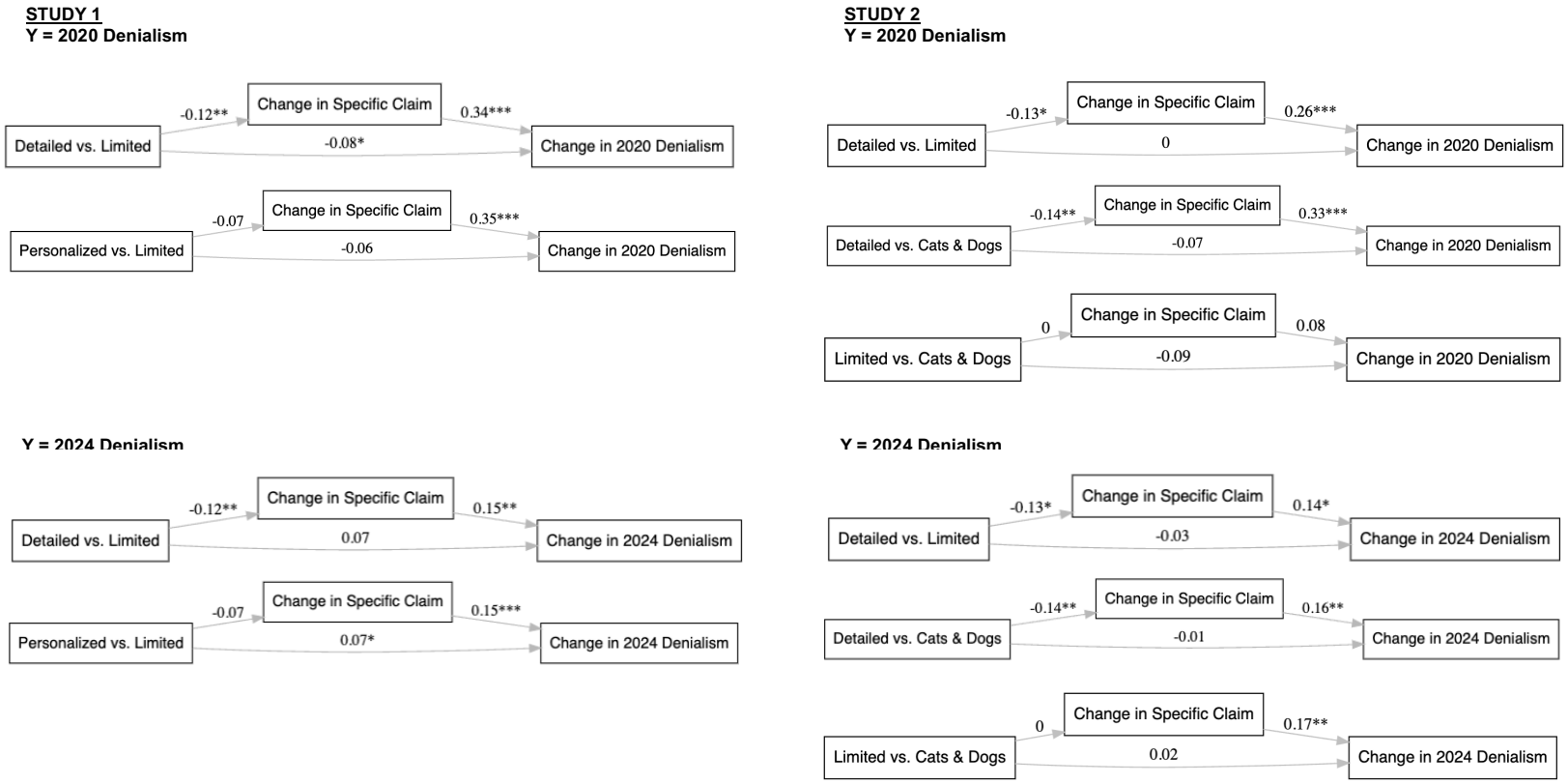
**Fig. D1. Distribution of belief-change patterns by condition**




Distribution of participant-level belief change patterns across experimental conditions, illustrating the proportion of participants exhibiting decreases, increases, or no change in election denialism following the intervention.

## Appendix E: Mediation Models

### Fig. E1. Mediation models of election denialism via claim certainty




Standardized mediation models testing whether changes in certainty about participants' articulated fraud claims mediate the relationship between AI dialogue condition and post-intervention election denialism. Path coefficients are shown with standard errors. Together, these results suggest that information-tailored feedback effectively reduced denialism through changes in certainty about specific claims, while personalized and adversarial statements may not be as effective in influencing denialism.