AI-Authored Prebunking Against Election Rumors: Evidence from Articles and Chatbots

Mitchell Linegar[1], Betsy Sinclair[2], Sander van der Linden[3], and R. Michael Alvarez[1]

[1]Linde Center for Science, Society, and Policy, California Institute of Technology, Pasadena, CA 91125, USA
[2]Department of Political Science, Washington University in St. Louis, St. Louis, MO 63130, USA
[3]Department of Psychology, Cambridge University, Cambridge, CB2 3EB, UK

July 14, 2026

Abstract

How can we improve trust in democracy and elections? Building on inoculation theory, we test whether generative artificial intelligence (GAI) can support interventions against election misinformation. We present GAI-generated prebunking materials either as static articles or interactive chatbot conversations. In a two-wave pre-registered experiment with U.S. registered voters (Wave 1: n=5,432, October 24-November 4, 2024; Wave 2: n=4,385, November 11-15, 2024), we measure effects on two outcomes: belief in false election rumors and trust in election integrity. Both interventions significantly reduce confidence in election rumors immediately and at recontact. They also increase trust in elections before the election, with chatbots showing larger point estimates on this broader measure. We also find that lower-cost static text interventions perform comparably to chatbots for reducing rumor belief. Overall, these results suggest that GAI-assisted prebunking can be a scalable tool for reducing rumor confidence and supporting election confidence under real-world electoral conditions.

1 Introduction

Can AI-authored prebunking materials, that is, preemptive refutations designed to build resistance before full exposure to false claims, reduce confidence in election misinformation and increase confidence in election administration? We evaluate matched prebunking content delivered either as a static article or as a chatbot conversation. Because both formats are

generated by the same AI system, the core informational content is held largely constant, and our primary analysis combines them into a single treatment arm to test whether prebunking can reduce confidence in election misinformation and shift broader election-confidence judgments.

GAI and large language models (LLMs) have rapidly expanded the supply of political content and the range of tools available to campaigns, platforms, and voters. These systems can generate politically relevant text, images, audio, and video at low cost, including deceptive content such as political deepfakes (Linegar, Kocielnik, and Alvarez, 2023; Barari, Lucas, and Munger, 2025). They can also be used to tailor messages and target communications (Simchon, Edwards, and Lewandowsky, 2024), which raises clear concerns about misinformation, trust, and electoral behavior (Spearing et al., 2025).

The same technologies can also be used for civic information and misinformation defense. Recent work shows that AI chatbots can durably reduce some conspiracy beliefs (Costello, Pennycook, and Rand, 2024), and other studies develop AI-enabled civic-information tools for voters (Velez, Green, and Sevi, 2025). What remains less clear is whether AI-authored prebunking works in a high-salience election setting. We also examine whether conversational delivery adds much beyond a cheaper static article when the underlying content is matched.

Our paper adds to this literature in two main ways. First, we study AI-authored prebunking in a high-salience election setting and show that it reduces confidence in false election rumors both immediately and at recontact. Robustness checks restricted to respondents who engaged with their assigned content strengthen these estimates, as one would expect when intent-to-treat effects are attenuated by minimal engagement. Second, because our intervention is implemented in both static-article and chatbot formats, we can examine whether delivery format materially changes those effects. Because both formats are AI-authored, these comparisons speak to delivery mode rather than AI-versus-human authorship.

The remainder of the paper proceeds as follows: Section 2 places our work in context. Section 3 introduces the design of our experiment. Section 4 explains our chatbot implementation and survey methodology. Section 5 presents our findings, which are discussed in Section 6 while Section 7 concludes and offers suggestions for future research.

2 Contribution

2.1 Election integrity and voter confidence

It has long been a staple of research on representative democracy that elections need to be free and fair (Bjornlund, 2004). What constitutes a free and fair election — and how one might verify that an election was conducted fairly — has been the subject of considerable research. Election integrity is complex and nuanced, with important components including voter and stakeholder trust that the process has been transparent and free from tampering, that the results are accurate and not tainted by manipulation, that only eligible individuals have participated, and that there has been no intimidation of candidates or voters (Elklit and Svensson, 1997; Birch, 2011; Norris, 2014).

These questions have produced a large and growing literature on election and voter confidence. Early work in this area, published nearly two decades ago, focused largely on how election administration is associated with voter confidence (Atkeson and Saunders, 2007; Alvarez, Hall, and Llewellyn, 2008). These studies also established the general measurement approach, which is to have voters evaluate their confidence that their own ballot was tabulated as they intended, but then to also ask each voter how confident they are that ballots in their county, state, and at the national level, were counted as intended (Atkeson and Saunders, 2007; Alvarez, Hall, and Llewellyn, 2008; Atkeson, Alvarez, and Hall, 2015).

Several studies have now investigated the relationship between many aspects of the voting experience and confidence, in particular voter satisfaction with their in-person or voting-by-mail experience (Claassen et al., 2013; Vonnahme and Miller, 2013; King and Barnes, 2019; Alvarez, Cao, and Li, 2021; Clark, 2021; Carter et al., 2024; Atkeson, McKown-Dawson, and Stein, 2025; Shino and Smith, 2025). Related studies have looked at specific aspects of election administration, such as election security, auditing procedures, and voter identification requirements, connecting those systems and procedures to voter confidence (Atkeson, Alvarez, Hall, and Sinclair, 2014; Bowler and Donovan, 2016; Coll, 2024; Jaffe et al., 2024). Importantly for our research, studies have looked at how voter education efforts are associated with voter confidence (Suttmann-Lea and Merivaki, 2023), and how voter perceptions of election fraud are related to their confidence in election administration and outcomes (Levy, 2021; Berlinski et al., 2023; Linegar and Alvarez, 2024).

The context of an election, and the election's partisan outcome, have also been shown to be associated with voter confidence, in particular whether elections are contentious is related to how confident voters are about the outcome (Wellman, Hyde, and Hall, 2018). Additionally, there is the so-called "Winner Effect" with respect to voter confidence; voters identifying with the winning candidates and parties often express greater confidence in the election's conduct and outcome after the outcome has been announced (Sances and Stewart III, 2015; Sinclair, Smith, and Tucker, 2018).

Our paper contributes to this literature in several ways. We show that AI-authored prebunking reduces confidence in false election rumors and produces an immediate increase in national-election confidence. Because these estimates come from a randomized experiment, they provide causal evidence on intervention effects. We do not find durable post-election effects on election confidence of the same magnitude; as we discuss below, this attenuation is consistent with the "Winner Effect" documented in prior work.

2.2 Attitudes about election fraud and election crimes

There is little systematic evidence that significant election fraud occurs in contemporary American elections (Minnite, 2011; Alvarez, Hall, and Hyde, 2009; Eggers, Garro, and Grimmer, 2021). Despite that, survey research shows that significant segments of the American electorate believe election fraud occurs with some frequency (Linegar and Alvarez, 2024). Understanding why these beliefs persist, and how they interact with broader confidence in elections, is central to the motivation for our study.

Several factors shape fraud perceptions. Partisanship plays an important role, with voters'

fraud concerns tracking their party identification (Beaulieu, 2014). There is also a "Winner Effect": those identifying with winning candidates and parties are less likely to express concerns about election fraud after the outcome is known (Levy, 2021). This pattern is adjacent to broader work arguing that conspiracy theories can become more appealing among losing or politically excluded groups (Uscinski and Parent, 2014). More broadly, perceptions of election-security procedures matter — when voters perceive stronger safeguards, they tend to express lower concern about fraud (Coll, 2024) — as do voter-identification requirements (Ansolabehere and Persily, 2007), with notable heterogeneity by partisanship (Kane, 2017). Some studies also connect concerns about voting technologies to fraud concerns (Beaulieu, 2016).

These fraud perceptions, in turn, feed back into confidence in elections. Recent research finds that election-fraud misinformation negatively affects confidence in elections (Berlinski et al., 2023), and the broader literature indicates that weaker confidence in election administration is associated with greater concern about election fraud. This creates a reinforcing dynamic: fraud beliefs erode institutional confidence, which may in turn make voters more receptive to further fraud claims.

More recent work has tried develop methods to refute or debunk election fraud claims. Research in this area has produced mixed results. On one hand, some studies have found that concerns about election fraud seem quite resistant to debunking and the presentation of information that tries to counteract claims of election fraud (Holman and Lay, 2019; Berlinski et al., 2023). Other recent work has shown that with respect to concerns about long delays in reporting election results, informational videos can reduce distrust in election results (Lockhart et al., 2024). Significantly, our findings contribute directly to this literature. AI-authored election-rumor prebunking reduces confidence in false election-fraud rumors, and this effect appears immediately after treatment and remains visible when we reinterview subjects following the 2024 election.

2.3 Inoculation theory

Finally, our work draws upon and contributes to inoculation theory. Inoculation theory has long been of interest in psychology and politics, where understanding how people are influenced by false information and propaganda has been a central research question (Papageorgis and McGuire, 1961; McGuire, 1964; van der Linden, 2024). The theory draws on an immunization analogy: just as vaccines expose the body to weakened doses of a pathogen to activate resistance against future infection, psychological inoculation works by exposing individuals to weakened forms of false information before they encounter it in full (prebunking). Interest in inoculation theory has surged in recent years, particularly regarding its potential to counter beliefs in conspiracy theories, climate misinformation, and false information about vaccines (Lewandowsky and van der Linden, 2021b; Traberg, Roozenbeek, and van der Linden, 2022). A recent meta-analysis also finds that psychological inoculation improves discrimination between reliable and unreliable information (van der Linden, Simchon, and Lewandowsky, 2026). Others have shown that prebunking approaches can produce doubts about misinformation (Guess et al., 2020; Pereira et al., 2024). Some political scientists have

used sustained inoculation strategies to reduce susceptibility to political misinformation, showing the general power of inoculation approaches (Bowles et al., 2025).

Inoculation or prebunking stands in contrast to traditional debunking, in which information-based strategies are developed and tested to try to get individuals to discount or disbelieve information they have already been exposed to (Chan et al., 2017; Bruns et al., 2023). There has been some research arguing that debunking strategies can backfire, with the attempt at refuting the incorrect information serving instead to reinforce the individual's beliefs in the falsehood because debunks tend to repeat the initial misinformation (Nyhan and Reifler, 2010; Lewandowsky, Ecker, et al., 2012), though more recent work has questioned the prevalence of these risks (Nyhan, 2021).

The larger issue with debunking, however, is that people can continue to rely on falsehoods despite having seen a correction, a phenomenon known as the "continued influence of misinformation" (Lewandowsky, Ecker, et al., 2012; Chan et al., 2017). Moreover, in today's information environment where rumors and conspiracy theories can spread virally on social media (van der Linden, 2023), debunking alone may not fully limit exposure. Although debunking remains important, researchers have increasingly focused on proactive strategies for people who have not yet encountered the falsehood, namely prebunking.

A growing body of research has examined prebunking specifically (Compton et al., 2021; Lewandowsky and van der Linden, 2021a; van der Linden, 2022; van der Linden, 2024; Roozenbeek and van der Linden, 2024). These studies have generally found that prebunking can give individuals the cognitive tools to resist misinformation or rumors they later encounter (Lewandowsky, Ecker, et al., 2012; Traberg, Roozenbeek, and van der Linden, 2022), and that these effects can sometimes be durable (van der Linden, 2022; Roozenbeek and van der Linden, 2024). Most of this work, however, has focused on conspiracy beliefs, climate misinformation, or anti-vaccination claims. Few published studies have focused on countering election rumors and misinformation, though see (Lockhart et al., 2024) and (Hopkins et al., 2025) for recent exceptions.

Most recently, researchers have begun leveraging GAI for prebunking. Costello, Pennycook, and Rand (2024) report sizable and durable reductions in conspiracy beliefs using an AI chatbot, while other work examines AI-generated civic-information tools (Velez, Green, and Sevi, 2025). This emerging literature suggests that AI may be useful for misinformation defense, but it leaves open whether AI-authored prebunking remains effective in a high-salience election context, and whether delivery format — article versus chatbot — matters when the underlying content is matched. In inoculation terms, chatbot delivery is also a more active format than static text because respondents interact with the intervention rather than only reading it (van der Linden, 2024).

We contribute to the inoculation literature in several ways. First, we use artificial intelligence to author and deliver prebunking arguments, and we show that these interventions reduce confidence in election rumors while producing a shorter-run increase in election confidence. Second, we extend inoculation research into a high-salience electoral setting by testing election-rumor interventions around the 2024 U.S. presidential election. Third, the reductions in non-assigned-rumor confidence connect election-rumor prebunking to broader-

spectrum resistance, including cross-protection or refutational-different effects beyond the specific claims that were directly prebunked (McGuire, 1961; Roozenbeek, Traberg, and van der Linden, 2022). Fourth, by holding authorship constant and varying delivery mode, we examine whether lower-cost static articles and interactive chatbots differ when the underlying content is matched.

2.4 Experimental context and identification scope

Our design identifies the treatment effects of AI-authored prebunking and, secondarily, compares static and interactive delivery within that framework. Both delivery formats are AI-authored, so the design speaks to how prebunking content is delivered rather than to whether AI authorship is inherently more or less persuasive than human authorship. The primary estimand is the pooled treatment effect relative to an active placebo, combining both formats into a single treatment arm; delivery-format comparisons are supplementary.

Three features of our experimental context warrant discussion before we turn to the design itself. The first concerns *content fidelity*. The prebunking arguments are closely matched across article and chatbot formats, so the informational content should be the primary driver of any treatment effect. If so, the pooled effect on rumor confidence should be negative regardless of delivery format, and article-specific and chatbot-specific estimates should not diverge sharply.

The second concerns *interaction intensity*. Chatbot delivery adds opportunities for conversational clarification, follow-up questions, and personalized engagement that go beyond the matched informational content. These features may matter more for broader attitudinal judgments — like confidence in the national election — than for the targeted rumor items, where the factual substance of the prebunking argument may suffice on its own. Chatbot delivery could therefore produce somewhat larger shifts on election confidence, though the design's 90/10 modality allocation limits precision for such comparisons.

The third concerns the *post-election context*. Our experiment spans the 2024 presidential election, and prior work documents large partisan shifts in election confidence once the outcome is known (the Winner Effect). Such shifts may compress detectable treatment contrasts on election confidence at recontact, even if rumor-reduction effects persist — rumor confidence is a more targeted judgment and likely less sensitive to the broad partisan realignment triggered by learning who won.

With these considerations in mind, we now describe the experimental design.

3 Experimental Design

Our experimental design, depicted in Figures 1b and 1a, evaluates the efficacy of AI-delivered pre-bunking interventions against election misinformation, comparing static text versus interactive chatbot delivery modalities. The experiment unfolds across four primary stages: pre-treatment measurement, randomized intervention assignment and delivery, misinformation exposure, and post-treatment/recontact measurement.

First, all participants completed an identical battery of pre-treatment questions. This survey collected standard demographic information and political variables, including partisanship and ideology. It also measured baseline levels of conspiracy beliefs, populist attitudes, confidence in specific election rumors, confidence in election integrity, feelings of political representation, and confidence in government. These measures establish a baseline against which treatment effects can be estimated, allowing for analyses of conditional average treatment effects and increasing statistical power.

Second, following the pre-treatment survey, participants proceeded through a sequence of random assignments. As shown in Figures 1b and 1a, participants were randomly assigned to the AI interaction modality: either reading a static **Article** (10% probability) or engaging with an interactive **Chatbot** (90% probability). We used this 90/10 allocation to prioritize precision in the chatbot condition, which was the primary deployment mode in our implementation, while preserving an article benchmark for modality comparisons. Given prior evidence that article-format prebunking can be effective (Linegar, Sinclair, et al., 2024), this design was intended primarily to test whether chatbot delivery produced substantively large departures from the article benchmark. The design is therefore informative about large modality differences, but less precise for small article-versus-chatbot contrasts. Subsequently, participants were randomly assigned to one of two misinformation topics: **Voter Rolls** or **Voter Fraud**, and further randomized the ordering of their two assigned articles. Finally, participants were randomly assigned to either the **Treatment** or **Control** condition. A key feature of this design is that, within each condition (**Treatment** or **Control**) and for each topic (**Voter Rolls** or **Voter Fraud**), the substantive informational content presented was held constant across the **Article** and **Chatbot** modalities. This allows for a direct comparison of the persuasive efficacy of delivering identical information via static text versus an interactive conversation, holding the core message constant. Because both modalities are AI-authored, these contrasts identify delivery-modality differences within AI interventions, not AI-versus-human authorship effects. The procedures for generating the AI articles and structuring the chatbot conversations are detailed in Section 4. Participants assigned to the **Treatment** condition received a pre-bunking intervention specific to their assigned topic, while those in the **Control** condition received a placebo intervention involving AI-generated content on thematically similar but unrelated topics (voting plans, felony disenfranchisement).[1]

Third, after the intervention phase (pre-bunking or placebo), all participants were exposed to misinformation related to their assigned topic. This exposure constituted the "full dose" of misinformation. Respondents were not told that these claims were true, and the study did not use a deceptive cover story to endorse them. Instead, respondents were asked to read short, real, published articles making those claims and then report how credible they found them. Specifically, each participant read two short articles detailing distinct rumors within their assigned topic category. For the **Voter Fraud** topic, the rumors involved allegations of non-citizen voting potentially swaying elections in key states. **Rumor A** focused on

[1]Relative to the misinformation claims tested later in the experiment, placebo materials were designed to be neutral with respect to election-fraud claims and election-direction narratives. We discuss potential interference channels and robustness diagnostics in A.5.

Arizona, summarizing claims from a lawsuit filed in Maricopa County alleging significant non-citizen registration based partly on survey data suggesting over 1% of likely voters might be non-citizens, a potentially decisive number given Arizona's narrow election margins. **Rumor B** focused on Nevada, describing claims from another lawsuit alleging up to 11,000 non-citizens were registered, with several thousand purportedly voting in 2020, potentially impacting outcomes in a state where recent major races were decided by fewer than 10,000 votes. For the **Voter Rolls** topic, the rumors centered on inaccuracies in voter registration databases. **Rumor A** described claims of "ghost voters," using a New Mexico example where an audit allegedly found thousands of deceased individuals, impossibly old registrants (over 120 years), and duplicate registrations remaining on the rolls, potentially affecting thousands of votes. **Rumor B** presented claims about impossibly old voters more generally, citing hypothetical reports of numerous voters listed as over 150 years old in a state's database, often coupled with accusations of official cover-ups. To mitigate order effects, the sequence in which participants encountered these two rumor articles (**Rumor A** then **Rumor B**, or **Rumor B** then **Rumor A**) was randomized, denoted as **Assign First Rumor** in the figures.

Fourth, immediately following the misinformation exposure, participants completed a post-treatment questionnaire. This survey measured our primary outcomes of interest: confidence that ballots in the upcoming national election would be accurately cast and counted, and confidence in the truthfulness of each specific rumor narrative they were exposed to during the misinformation exposure phase. Election-confidence items used a 0-to-10 scale, where 0 indicates no confidence and 10 indicates complete confidence. Rumor-confidence items asked respondents to rate agreement with the false rumor statement, where 0 means completely disagree and 10 means completely agree; lower rumor-confidence scores therefore indicate lower agreement that the false claim is true. The questionnaire also re-measured secondary outcomes, including feelings of political representation and confidence in government. Approximately one week after the initial experimental session, participants were invited to complete a follow-up (recontact) survey, which re-administered the same battery of outcome measures to assess the persistence of any observed treatment effects.

This multi-stage, randomized design allows for the rigorous estimation of the causal effects of AI-delivered pre-bunking interventions, facilitating comparison between **Article** and **Chatbot** delivery modalities while controlling for baseline beliefs, topic focus, and exposure to specific misinformation narratives.

4 Methodology

This study compares the effectiveness of delivering pre-bunking information via static AI-generated text (**Article**) versus an interactive AI **Chatbot**. This section details the procedures used to create the intervention content for both modalities, ensuring similar informational content while varying the delivery mechanism. All AI text generation, for both articles and chatbot responses, utilized OpenAI's GPT-4o.

4.1 AI-Generated Pre-bunking Articles (Article modality)

The pre-bunking articles serving as the intervention in the **Article** modality were generated following procedures established in our prior work (Linegar, Sinclair, et al., 2024). In that work, we utilized a human-in-the-loop process to iteratively refine an LLM prompt designed to produce effective inoculation-style articles. This involved providing the LLM with examples of misinformation ("full exposure" articles) and factual counter-information, generating draft inoculation articles, having researchers with relevant expertise review these drafts, and revising the prompt based on this feedback to improve the quality and persuasive framing of the output. We applied this optimized prompt to produce inoculation articles for the **Voter Fraud** and **Voter Rolls** rumors central to this experiment, using articles endorsing these rumors as input to the LLM. We found no evidence that AI-written inoculation articles were less effective than human ones. We use this same prompt structure and methodology to produce the **Treatment** intervention articles in the current experiment. The corresponding placebo articles (**Control** condition) were generated by providing the LLM with one of the pre-bunking articles as a stylistic template and instructing it to produce a similar article on the neutral topic (e.g., voting plans, felony disenfranchisement).[2]

4.2 AI Chatbot Implementation (Chatbot modality)

The **Chatbot** modality aimed to deliver the same substantive pre-bunking information contained in the **Article** interventions but through an interactive, conversational format. The informational core for the chatbot's responses in the **Treatment** condition was derived directly from the AI-generated pre-bunking articles corresponding to the participant's assigned topic (**Voter Fraud** or **Voter Rolls**), supplemented by baseline facts about election integrity primarily drawn from the Cybersecurity and Infrastructure Security Agency's (CISA) public resources. Similarly, in the **Control** condition, the chatbot drew upon the content of the placebo articles. This ensures informational equivalence between the **Article** and **Chatbot** modalities with respect to the core message content provided by the experimenters. The chatbot could then draw on this material alongside the general knowledge encoded in the underlying GPT-4o model, guided and constrained by the system prompt, to engage with users.

Our chatbot implementation utilizes a structured conversational architecture, distinct from more open-ended chatbot interactions common in some recent research (Costello, Pennycook, and Rand, 2024). This structure more closely resembles protocols used in traditional survey experiments or structured interventions like deep canvassing (Broockman and Kalla, 2016), facilitating systematic delivery of information and analysis of interactions. The system employs a three-agent design. The first agent processes the user's input, for example identifying

[2]Because the placebo topics are not entirely inert — felony disenfranchisement, for instance, touches on electoral fairness and could plausibly interact with partisan priors — one might worry that the placebo itself shifts outcomes or does so differentially by party. We investigate this possibility using order-based diagnostics in A.5. Order effects within the placebo condition are small and statistically insignificant across all outcomes, suggesting that the placebo content does not meaningfully influence the treatment-control comparisons.

frustration, "jail-break" attempts (where the user uses prompts to attempt to take control of the bot in order to elicit harmful or false responses), and assessing its relevance to the current conversational step (e.g., is it an answer to the last question, a follow-up query, an expression of disagreement, or off-topic?). The second agent acts as a state manager, tracking information from the user, which question in the script is active, whether conditions to proceed to the next question have been met, and forwarding all this information (in particular, whether to ask a follow-up question, ask the last question again, or move on to the next question) to the final agent.

The third agent is the response generator. It receives the current conversational state, the history of the interaction, the core informational content (the relevant **Article** text and CISA facts), and specific instructions from the state manager based on a predefined script (specified in JSON format, see Supplementary Materials). This agent then generates the natural language response presented to the user. The script guides the conversation through distinct phases. It begins with an initial step involving introductions and an assessment of the participant's willingness to discuss the topic, potentially offering a brief overview. The core of the interaction involves sequentially presenting and discussing the two relevant rumors (**Rumor A** and **Rumor B**) corresponding to the assigned topic. In the **Treatment** condition, the chatbot introduces the rumor (framing it as something encountered) and then presents the pre-bunking arguments and facts drawn directly from the provided **Article** text, often prompted by specific tool-use instructions within the script (e.g., referencing 'ARTICLE1' or 'ARTICLE2'). The chatbot then solicits the participant's reaction and is equipped to respond to follow-up questions or points raised by the user, drawing on the provided context and its underlying knowledge base, before attempting to proceed according to the script. Participants can skip questions, prompting the state manager to advance the script accordingly. The conversation concludes with a final step offering an opportunity for final questions and thanking the participant.

This structured, three-agent approach provides several benefits. It ensures relatively consistent delivery of the core intervention content across participants within the **Chatbot** condition, comparable to the **Article** condition. Furthermore, by standardizing the conversational flow and the points at which specific information is introduced or questions are asked, it facilitates more systematic analysis of participant interactions, allowing for comparisons of how users engage with specific arguments or prompts across the sample.

4.3 Data Collection

The study was fielded by YouGov using a sample of 5,432 U.S. registered voters from its online panel, recruited to match the national registered-voter population. Wave 1 was fielded October 24 to November 4, 2024; Wave 2 (recontact) was fielded November 11 to November 15, 2024. All Wave 1 respondents were invited to the follow-up survey, and 4,385 participated in Wave 2, for an 81% recontact rate.

Modality assignment intentionally targeted an approximately 90/10 chatbot/article split; the realized counts were 4,808 chatbot versus 624 article respondents in Wave 1 (88.5% versus 11.5%) and 3,884 versus 501 in Wave 2 (88.6% versus 11.4%). This allocation prioritizes

precision for the pooled and chatbot-centered estimates, while the article arm serves as a benchmark linked to our prior non-chatbot prebunking work (Linegar, Sinclair, et al., 2024). Consequently, the design is informative about large delivery-format differences but has limited power for small article-versus-chatbot contrasts.

In the first implementation, all subjects completed a questionnaire before being assigned to treatment or control conditions. See Figure 1 for further details about the treatment and control procedures. After the treatment and control procedures, all subjects were administered a post-treatment questionnaire. In the recontact wave, all subjects completed a questionnaire with similar questions so that we can test the durability of the treatment effects.

Disposition and sample accounting. YouGov provided a full Wave 1 disposition report covering every survey start and its final status, including final completes, breakoffs, screenouts, and removals for speed or response quality. These materials begin at survey starts rather than invitation sends. We contracted for at least $n = 4{,}000$ Wave 1 completes, but that figure was simply a minimum delivery threshold rather than an analytically important cutoff. In practice, the October–November 2024 field period produced 8,428 starts, 8,093 randomized respondents, and 5,431 final completes in YouGov's disposition accounting. The main Wave 1 analysis sample contains 5,432 respondents because it also retains one case with missing post-treatment outcomes, and the expanded Wave 1 sample used for the appendix robustness checks contains 8,429 respondents because it retains non-final respondents whenever immediate post-treatment outcomes are observed and also includes one respondent with a missing disposition code. The full sample flow, arm-level attrition and compliance checks, expanded-sample robustness reruns, and party-specific attrition check where party identification is observed are reported in A.3.

The one-respondent difference between the main Wave 1 analysis sample ($n = 5{,}432$) and YouGov's "Final sample" count ($n = 5{,}431$) reflects that the analyzed sample retains one respondent who is missing all post-treatment outcomes and is therefore not counted as a final complete in the disposition report.

Additional detail on how the Wave 1 sample narrows from survey starts to the analyzed sample, how sensitive the follow-up results are to recontact missingness, how we handled misinformation exposure and post-exposure correction choices, and how the chatbot materials were archived for review appears in A.3, A.4, A.1, and A.7.

4.4 Analysis Sample and Missing Data

Our primary analysis uses complete cases for each model specification after applying the relevant survey weights.[3] The models summarized in Table 1 combine article and chatbot delivery into a single treatment arm and estimate treatment-minus-placebo effects in the full

[3] A.4 reports a multiple-imputation sensitivity analysis for follow-up nonresponse (MICE, 30 imputations). Imputed estimates are directionally consistent with complete-case results and do not change the qualitative conclusions.

analytic sample. We treat this pooled intent-to-treat estimate as primary because it preserves the randomized assignment contrast.

As a robustness check, A.2 additionally reports estimates in an *engaged subsample* that excludes chatbot respondents with low engagement (fewer than three messages sent) and article respondents who spent fewer than five seconds on the assigned intervention pages. These estimates are uniformly stronger, consistent with what one would expect when intent-to-treat effects are attenuated by minimal exposure. A.3 provides the full disposition accounting, engagement rates, Lee bounds, and expanded-sample checks; usable sample sizes by outcome and wave are reported in the model tables.

5 Results

Table 1 and Figures 2 and 3 summarize our main findings. The table pairs overall treatment coefficients with party-specific subgroup estimates and pre-treatment group means; the figure shows the group-level trajectories behind those estimates. Table A.1 reports the full primary models, and Table A.2 compares these estimates with the engaged subsample. The appendix extends the same specifications to ballot-confidence items, other-rumor outcomes, and additional robustness checks (A.2).

All estimates are intent-to-treat contrasts between treatment and active placebo, using survey-weighted complete cases in the full analytic sample. Restricting to the engaged subsample yields uniformly stronger coefficients (Table A.2), consistent with attenuation from respondents with minimal exposure. Because the placebo arm is active rather than inert, A.5 reports order-based diagnostics for placebo non-inertness.

5.1 Primary Specifications

We focus on two outcomes: assigned-rumor confidence and national election confidence, the clearest tests of the paper's two main claims. The appendix reports companion ballot-confidence items, ANES representation outcomes, and broader outcome-family tables.

5.2 Confidence in Election Rumors

Treatment reduces assigned-rumor confidence immediately after exposure, and the reduction remains visible at recontact. Table 1 reports pre-treatment group means alongside the estimated effects. Before the intervention, placebo and treatment groups start at nearly identical levels of rumor confidence (overall means of 4.72 and 4.73 on the 0–10 scale), with substantial partisan variation: Democrats average around 2.5, Independents around 4.4, and Republicans around 7.0. The pooled treatment coefficient is -0.320 (SE: 0.054, $p < 0.001$) immediately and -0.200 (SE: 0.079, $p < 0.05$) at recontact. Substantively, the immediate effect is about one-third of a point on the 0–10 rumor-agreement scale. In the adjusted trajectories in Figure 3, placebo respondents move upward by about 0.19 points after misinformation exposure, while treated respondents move downward by about 0.13

points, so the treatment more than offsets the placebo group's immediate movement toward the false rumor. The party-specific rows show the same negative immediate pattern for Democrats, Independents, and Republicans, while the recontact estimates remain negative with wider subgroup uncertainty. Figure 3 shows the corresponding group-level trajectories: rumor-confidence changes move downward after treatment and do not exhibit the large post-election reversal seen in the election-confidence panels.

This pattern is robust across specifications. A.2 shows that engaged-subsample checks, reduced-control variants, minimal benchmarks, and recontact-missingness analyses all preserve the basic rumor-reduction result. Notably, the appendix also reports effects on the *other-rumor* outcome — confidence in the election-rumor topic that a respondent was *not* assigned to and therefore did not receive prebunking for. Treatment reduces other-rumor confidence by -0.480 (SE: 0.088, $p < 0.001$) immediately and -0.436 (SE: 0.139, $p < 0.01$) at recontact (Table A.6). These roughly half-point effects suggest that the intervention reduces receptivity to election rumors more broadly, not just the specific claims targeted by the prebunking content.

Article-versus-chatbot comparisons are secondary. As described in Section 4, the design's 90/10 modality split limits precision for small delivery-format differences, so we emphasize the pooled specification that combines both formats into a single treatment arm.

5.3 Confidence in the National Election

Treatment also increases national election confidence immediately after exposure, but that broader effect is less durable. As Table 1 shows, pre-treatment election confidence is already high overall (6.93/7.03 for placebo/treatment) but varies sharply by party: Democrats average 8.4, Independents 6.8, and Republicans 5.8. The pooled post-treatment coefficient is 0.163 (SE: 0.039, $p < 0.001$), while the post-election coefficient is 0.077 (SE: 0.070), which is not statistically distinguishable from zero. The immediate effect is modest in absolute terms — about one-sixth of a point on the 0–10 national-election-confidence scale — but it is directionally consistent with the intervention's intended institutional-confidence effect. The party-specific rows show that the immediate estimate is non-negative in each group and largest for Republicans (0.245, SE: 0.071), the group with the most room to move upward on this measure. After the election, however, all subgroup contrasts compress. Figure 3 illustrates why: post-election partisan movement in the election-confidence components is large relative to the between-arm delta in adjusted change. We therefore treat the immediate election-confidence gain as a supporting result rather than as a durable effect.

The broader election-confidence family follows the same qualitative pattern in the appendix: immediate estimates are generally non-negative, while post-election estimates are smaller and noisier.

5.4 Winner Effect

Why do rumor-confidence effects persist while election-confidence effects attenuate after the election? Figure 3 is the clearest summary: it places the weighted group-level trajectories

for placebo and treatment alongside the between-arm delta in adjusted change. Across the election-confidence panels, post-election partisan updating is large relative to that cross-arm delta, especially once respondents have observed the election outcome. The pre-treatment means and subgroup coefficients in Table 1 line up with that visual summary, while Table A.1 makes the same point in coefficient form: the Republican indicator on national election confidence shifts from a small negative pre-election coefficient to a large positive post-election coefficient.

This winner-effect swing is larger than the immediate cross-arm delta on national election confidence and therefore compresses detectable post-election contrasts. In the full national-election-confidence model, for example, the Republican coefficient shifts by roughly one point from the immediate post-treatment model to the post-election model (Table A.1), several times the size of the immediate treatment coefficient. By contrast, the rumor-confidence panels in Figure 3 remain comparatively stable across waves, so the treatment-minus-placebo difference stays negative even after the election outcome is known. The figure in A.1 shows the same pattern at the broader outcome-family level for the two main outcome families.

6 Discussion

AI-authored prebunking reduces confidence in false election rumors, and that effect persists when we reinterview respondents after the election. Moreover, the intervention reduces confidence not only in the specific rumors targeted by the prebunking content, but also in election rumors from the non-assigned topic — suggesting that prebunking builds broader resistance to election misinformation rather than merely addressing the particular claims it refutes. The election-confidence result is more limited: national-ballot confidence increases immediately after treatment, but the post-election estimate is smaller and imprecise. Table 1 and Figure 3 show these patterns in the overall coefficients and in the underlying group-level trajectories by party. Robustness checks in the engaged subsample strengthen the estimates in the expected direction, while comparisons between article and chatbot delivery do not change the main conclusions.

The election-confidence effect is also specific in its geographic scope. Treatment increases national-ballot confidence but does not significantly shift own-ballot or county-ballot confidence (Appendix Tables A.4 and A.5). This pattern is consistent with the content of the rumors themselves: the misinformation claims in our study concern national-level phenomena — non-citizen voting swaying state outcomes, inaccurate voter rolls affecting statewide totals — so prebunking those claims shifts judgments about the national election without altering respondents' confidence that their own vote was counted correctly. The distinction suggests that prebunking operates on the specific beliefs it targets rather than producing a diffuse boost to electoral confidence at all levels.

Three aspects of the design and setting shape how we interpret these results. The first is content fidelity: because the prebunking arguments are closely matched across delivery formats, the overall rumor result is consistent with the informational content itself doing most of the work. The second is interaction intensity: conversational delivery may add opportunities

for clarification and follow-up, but our evidence on that margin is not precise enough to bear much weight. The third is contextual updating: post-election partisan shifts — the Winner Effect — are larger than the between-arm delta on election confidence, which limits detectable durability for that outcome. Figure 3 illustrates this directly: election-confidence trajectories diverge sharply by party after the election, whereas rumor-confidence trajectories remain comparatively stable.

The article-versus-chatbot comparison warrants additional qualification. As noted in Section 4, the 90/10 modality split limits precision for small differences between delivery formats, so we do not treat that comparison as a primary finding. And because both formats are AI-authored, the design identifies differences between delivery modes within AI interventions — it does not isolate AI-versus-human authorship effects. This is evidence on how to deploy AI-authored prebunking, not a test of whether AI authorship is itself uniquely persuasive.

Relative to recent open-ended AI persuasion studies, our effects are smaller in absolute magnitude. This is expected: we constrain content to factual election-integrity material, evaluate durability over a one-week window, and field during a high-salience national election in which large exogenous shocks to confidence are likely. In that setting, effects on the order of a few tenths of a scale point are still substantively meaningful, especially when they are directionally consistent across subgroups and robustness checks.

From an implementation perspective, these results support AI-authored prebunking as a scalable risk-mitigation tool rather than a standalone cure. The practical objective is not to eliminate misinformation exposure, but to reduce susceptibility and preserve confidence under realistic electoral conditions. The secondary delivery-format evidence is a useful caution against assuming that more complex conversational delivery is automatically necessary for short-run rumor reduction.

7 Conclusion

AI-authored election-rumor prebunking can work in practice. In our data, these materials reduce confidence in targeted rumors, spill over to reduce confidence in non-targeted election rumors, and support election confidence in the short run. These findings are especially relevant when interventions need to be deployed quickly and at scale.

There remains substantial room for additional research. One direction is to improve chatbot design and targeting, for example by testing alternative conversational styles and matching strategies by subject characteristics. Another is to study more explicitly when more intensive conversational delivery yields benefits large enough to justify its added cost and complexity. Those are implementation questions, not merely stylistic ones.

Our prebunking interventions here are entirely text-based. In a world where rumors and falsehoods are increasingly disseminated via audio or video, it will be important for future research to develop and test AI methods for producing inoculation materials in those formats. Similarly, we encourage researchers to develop and test AI election rumor inoculation methods for languages other than English.

Finally, while election rumors are central to this paper, other studies suggest that prebunk-

ing can also be useful in domains such as climate change and vaccination. We therefore view AI-based inoculation as one potential strategy for helping citizens better evaluate misleading claims in political communication. This approach is not censorship of political expression; rather, it is an informational intervention intended to strengthen evaluation and judgment under real-world information exposure.

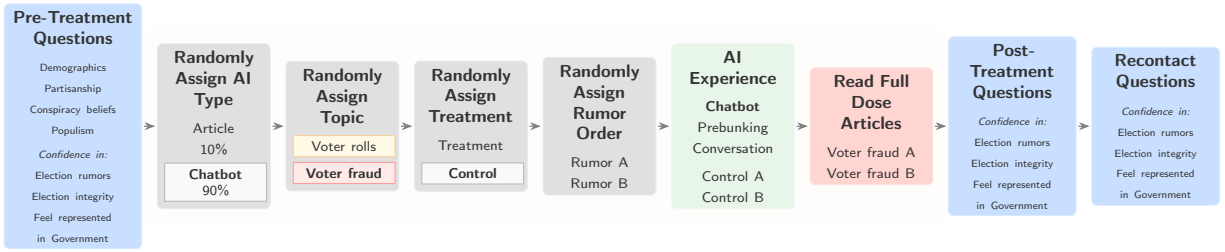
Tables

Table 1: Treatment effects on rumor confidence and election confidence

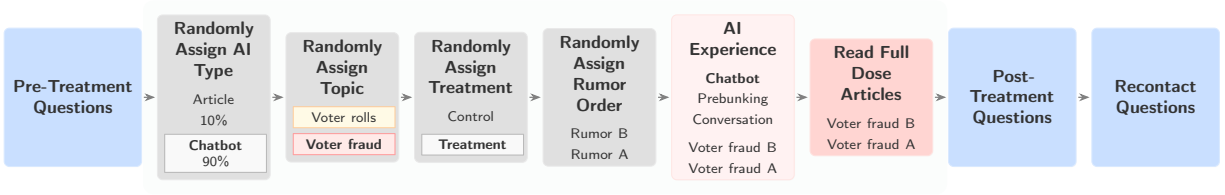
Group	Pre mean (placebo / treatment)	Post-treatment effect (SE)	Post-election effect (SE)	N (post / recontact)
Assigned rumor confidence				
Overall	4.72 / 4.73	-0.320*** (0.054)	-0.200* (0.079)	5,191 / 4,203
Democrats	2.53 / 2.53	-0.291** (0.089)	-0.066 (0.123)	2,032 / 1,614
Independents	4.45 / 4.43	-0.296*** (0.089)	-0.350* (0.146)	1,428 / 1,205
Republicans	6.92 / 7.21	-0.313** (0.095)	-0.192 (0.135)	1,731 / 1,384
National election confidence				
Overall	6.93 / 7.03	0.163*** (0.039)	0.077 (0.070)	5,190 / 4,203
Democrats	8.43 / 8.45	0.111 (0.057)	0.094 (0.123)	2,031 / 1,613
Independents	6.70 / 6.83	0.135 (0.070)	0.070 (0.135)	1,428 / 1,205
Republicans	5.81 / 5.80	0.245*** (0.071)	0.056 (0.097)	1,731 / 1,385

Notes: Pre means are weighted pre-treatment placebo/treatment averages and correspond to the pre-treatment labels in Figure 3. Effect entries are weighted OLS treatment coefficients with design-based standard errors from `svyglm`; significance stars are appended to the coefficient estimates. The overall rows report the full-sample treatment estimates, while the party rows estimate the same model within each party group and therefore omit party identification because it is constant within subgroup. All models adjust for the corresponding pre-treatment outcome and the preregistered covariate set. Rumor-confidence and election-confidence outcomes are measured on 0–10 scales. Post-election models use the recontact survey weights. Figure 3 shows the corresponding group-level trajectories, and Appendix Table A.2 reports the engaged-subsample estimates. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Figures



(a) Control Condition, Rumor A First



(b) Treatment Condition, Rumor B First

Figure 1: **Chatbot Inoculation Experimental Workflow.** Panels illustrate the experimental sequence for participants assigned to the (a) Control and (b) Treatment conditions. Both panels show progression from Pre-Treatment Questions through four randomization stages to Post-Treatment and Recontact assessments. Grey boxes indicate randomization points: AI Type (target: 90% Chatbot vs 10% Article), Topic (Voter fraud in yellow vs Voter rolls in red), Treatment assignment, and Rumor Order (A-B vs B-A). The critical experimental manipulation occurs at the AI Experience stage: Control participants (Panel a, pink background) receive neutral content about voting plans and felony disenfranchisement (green box), while Treatment participants (Panel b, red box) receive topic-specific prebunking content matching their assigned rumor. All participants then read two "Full Dose" misinformation articles corresponding to their assigned topic. Measurements occur immediately post-treatment and again one week later. Color-coded legend identifies content types. Highlighted boxes trace the example participant's path through each condition.

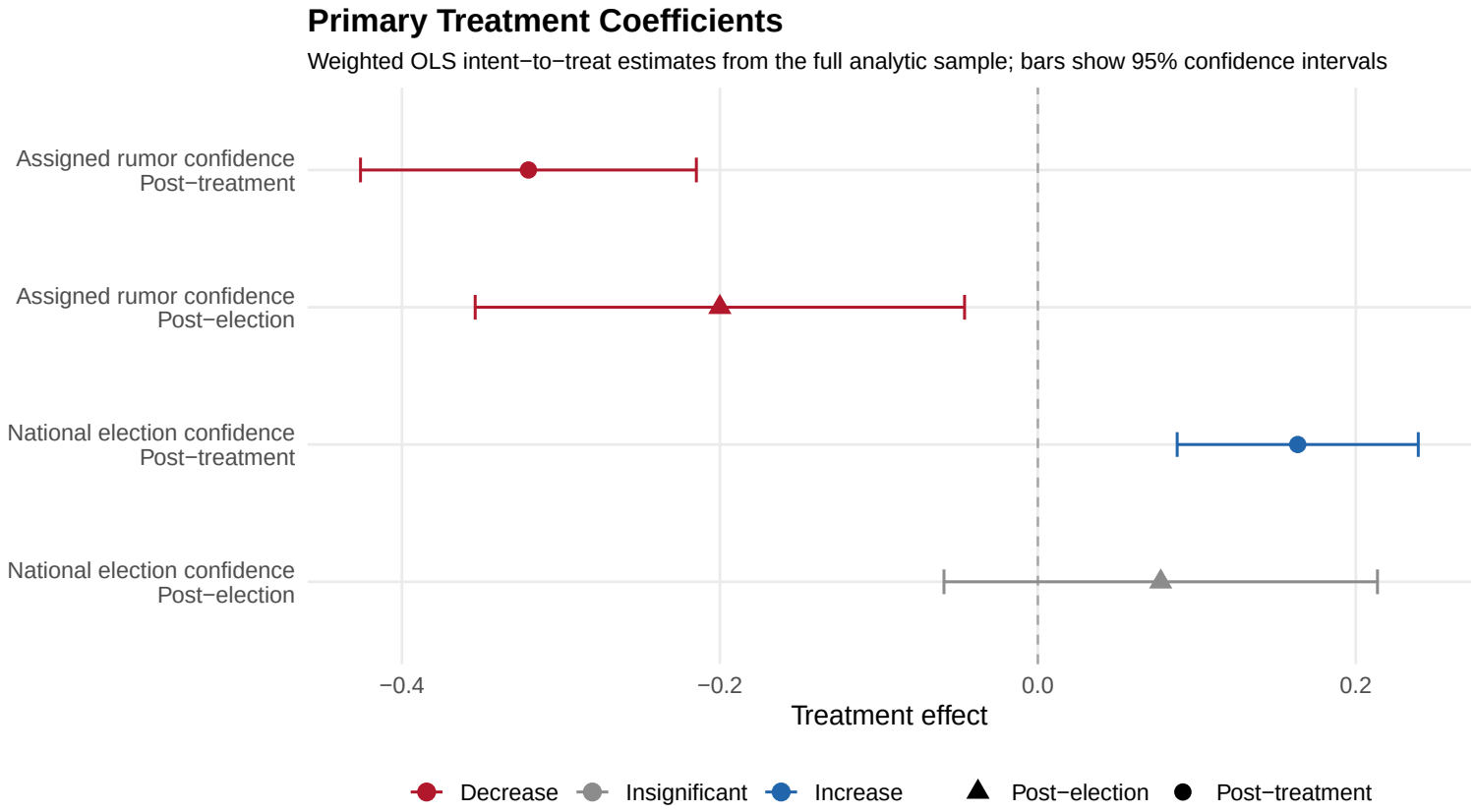


Figure 2: **Primary Treatment Coefficients.** Points show the treatment effect from the four primary models summarized in Table 1; bars are 95% confidence intervals. Red points indicate statistically distinguishable decreases, blue points indicate statistically distinguishable increases, and gray points indicate estimates whose 95% confidence intervals include zero. Rumor-confidence estimates are negative both immediately and at recontact, while the national-election estimate is positive immediately after treatment and smaller post-election.

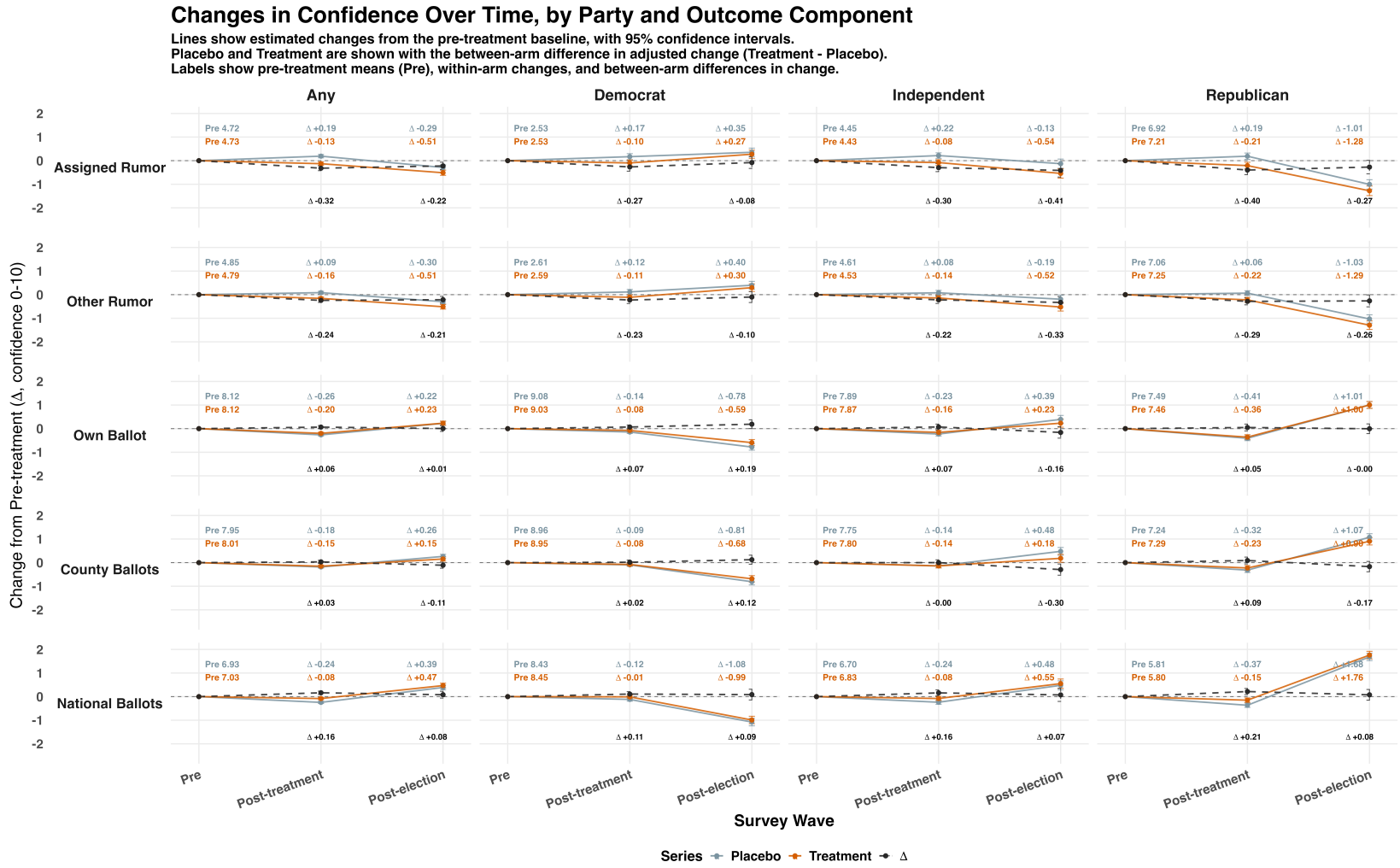


Figure 3: **Adjusted Change Trajectories by Party and Outcome Component.** Weighted regression-based changes relative to the pre-treatment baseline are shown separately for assigned-rumor confidence, other-rumor confidence, and the three election-confidence items. Placebo and treatment series are plotted alongside mean differences (Δ = Treatment − Placebo) with 95% confidence intervals. Post-election partisan movement is much larger in the election-confidence components than in the rumor-confidence components, which helps explain why election-confidence differences attenuate after the election while rumor-confidence differences remain more stable.

References

Alvarez, R Michael, Jian Cao, and Yimeng Li (2021). "Voting experiences, perceptions of fraud, and voter confidence". In: *Social Science Quarterly* 102.4, pp. 1225–1238.

Alvarez, R Michael, Thad E Hall, and Susan D Hyde (2009). *Election fraud: detecting and deterring electoral manipulation*. Brookings Institution Press.

Alvarez, R Michael, Thad E Hall, and Morgan H Llewellyn (2008). "Are Americans confident their ballots are counted?" In: *The Journal of Politics* 70.3, pp. 754–766.

Ansolabehere, Stephen and Nathaniel Persily (2007). "Vote fraud in the eye of the beholder: The role of public opinion in the challenge to voter identification requirements". In: *Harvard Law Review* 121, p. 1737.

Atkeson, Lonna Rae, R Michael Alvarez, and Thad E Hall (2015). "Voter confidence: How to measure it and how it differs from government support". In: *Election Law Journal* 14.3, pp. 207–219.

Atkeson, Lonna Rae, R. Michael Alvarez, Thad E. Hall, and J. Andrew Sinclair (2014). "Balancing Fraud Prevention and Electoral Participation: Attitudes Toward Voter Identification". In: *Social Science Quarterly* 95.5, pp. 1381–1398.

Atkeson, Lonna Rae, Eli McKown-Dawson, and Robert M. Stein (2025). "The Costs of Voting and Voter Confidence". In: *Political Research Quarterly* 78.1, pp. 22–37. DOI: 10.1177/10659129241283169.

Atkeson, Lonna Rae and Kyle L Saunders (2007). "The effect of election administration on voter confidence: A local matter?" In: *PS: Political Science & Politics* 40.4, pp. 655–660.

Barari, Soubhik, Christopher Lucas, and Kevin Munger (2025). "Political Deepfakes Are as Credible as Other Fake Media and (Sometimes) Real Media". In: *Journal of Politics* 87 (2). DOI: https://doi.org/10.1086/732990.

Beaulieu, Emily (2014). "From voter ID to party ID: How political parties affect perceptions of election fraud in the US". In: *Electoral Studies* 35, pp. 24–32.

— (2016). "Electronic Voting and Perceptions of Election Fraud and Fairness". In: *Journal of Experimental Political Science* 3.1, pp. 18–31. DOI: 10.1017/XPS.2015.9.

Berlinski, Nicolas et al. (2023). "The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections". In: *Journal of Experimental Political Science* 10.1, pp. 34–49. DOI: 10.1017/XPS.2021.18.

Birch, Sarah (2011). *Electoral Malpractice*. Oxford University Press, USA.

Bjornlund, Eric (2004). *Beyond free and fair: Monitoring elections and building democracy*. Woodrow Wilson Center Press.

Bowler, Shaun and Todd Donovan (2016). "A Partisan Model of Electoral Reform: Voter Identification Laws and Confidence in State Elections". In: *State Politics & Policy Quarterly* 16.3, pp. 340–361. DOI: 10.1177/1532440015624102.

Bowles, Jeremy et al. (2025). "Sustaining Exposure to Fact-Checks: Misinformation Discernment, Media Consumption, and Its Political Implications". In: *American Political Science Review* 119.4, pp. 1864–1887. DOI: 10.1017/S0003055424001394.

Broockman, David and Joshua Kalla (2016). "Durably reducing transphobia: A field experiment on door-to-door canvassing". In: *Science* 352.6282, pp. 220–224.

Bruns, Hendrik et al. (2023). "The role of (trust in) the source of prebunks and debunks of misinformation. Evidence from online experiments in four EU countries". In.

Carter, Luke et al. (July 2024). "From Confidence to Convenience: Changes in Voting Systems, Donald Trump, and Voter Confidence". In: *Public Opinion Quarterly* 88.SI, pp. 516–535. DOI: 10.1093/poq/nfae034.

Chan, Man-pui Sally et al. (2017). "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation". In: *Psychological Science* 28.11, pp. 1531–1546. DOI: 10.1177/0956797617714579.

Claassen, R.L. et al. (2013). "Voter Confidence and the Election-Day Voting Experience". In: *Political Behavior* 35, pp. 215–235. DOI: https://doi.org/10.1007/s11109-012-9202-4.

Clark, Jesse T. (2021). "Lost in the Mail? Vote by Mail and Voter Confidence". In: *Election Law Journal: Rules, Politics, and Policy* 20.4, pp. 382–394. DOI: 10.1089/elj.2020.0682.

Coll, Joseph A. (2024). "Securing Elections, Securing Confidence? Perceptions of Election Security Policies, Election Related Fraud Beliefs, and Voter Confidence in the United States". In: *Election Law Journal: Rules, Politics, and Policy* 23.2, pp. 173–192. DOI: 10.1089/elj.2023.0024.

Compton, Josh et al. (2021). "Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories". In: *Social and Personality Psychology Compass* 15.6, e12602.

Costello, Thomas H, Gordon Pennycook, and David G Rand (2024). "Durably reducing conspiracy beliefs through dialogues with AI". In: *Science* 385.6714, eadq1814.

Eggers, Andrew C., Haritz Garro, and Justin Grimmer (2021). "No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election". In: *Proceedings of the National Academy of Science* 118.45, e2103619118. DOI: https://doi.org/10.1073/pnas.2103619118.

Elklit, Jørgen and Palle Svensson (1997). "The rise of election monitoring: What makes elections free and fair?" In: *Journal of Democracy* 8.3, pp. 32–46.

Guess, Andrew M et al. (2020). "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India". In: *Proceedings of the National Academy of Sciences* 117.27, pp. 15536–15545.

Holman, Mirya R. and J. Celeste Lay (2019). "They See Dead People (Voting): Correcting Misperceptions about Voter Fraud in the 2016 U.S. Presidential Election". In: *Journal of Political Marketing* 18.1-2, pp. 31–68. DOI: 10.1080/15377857.2018.1478656.

Hopkins, Shaye-Ann M et al. (Aug. 2025). "Politricks: Teaching political tricks and discernment through active and passive tools". In: *PNAS Nexus* 4.8, pgaf245. ISSN: 2752-6542. DOI: 10.1093/pnasnexus/pgaf245.

Jaffe, Jacob et al. (July 2024). "Trust in the Count: Improving Voter Confidence with Post-election Audits". In: *Public Opinion Quarterly* 88.SI, pp. 585–607. DOI: 10.1093/poq/nfae029.

Kane, John V (Dec. 2017). "Why Can't We Agree on ID? Partisanship, Perceptions of Fraud, and Public Support for Voter Identification Laws". In: *Public Opinion Quarterly* 81.4, pp. 943–955. DOI: 10.1093/poq/nfx041.

King, Bridgett A. and Alicia Barnes (2019). "Descriptive Representation in Election Administration: Poll Workers and Voter Confidence". In: *Election Law Journal: Rules, Politics, and Policy* 18.1, pp. 16–30. DOI: 10.1089/elj.2018.0485.

Lee, David S. (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects". In: *The Review of Economic Studies* 76.3, pp. 1071–1102.

Levy, Morris (2021). "Winning cures everything? Beliefs about voter fraud, voter confidence, and the 2016 election". In: *Electoral Studies* 74, p. 102156.

Lewandowsky, Stephan, Ullrich KH Ecker, et al. (2012). "Misinformation and its correction: Continued influence and successful debiasing". In: *Psychological science in the public interest* 13.3, pp. 106–131.

Lewandowsky, Stephan and Sander van der Linden (2021a). "Countering misinformation and fake news through inoculation and prebunking". In: *European Review of Social Psychology* 32.2, pp. 348–384. DOI: 10.1080/10463283.2021.1876983.

— (2021b). "Inoculation theory and misinformation". In: *Nature* 590.7847, pp. 518–519.

Linegar, Mitchell and R Michael Alvarez (2024). "American Views About Election Crimes in 2024". In: *Preprint*.

Linegar, Mitchell, Rafal Kocielnik, and R. Michael Alvarez (2023). "Large language models and political science". In: *Frontiers in Political Science* 5. DOI: https://doi.org/10.3389/fpos.2023.1257092.

Linegar, Mitchell, Betsy Sinclair, et al. (2024). *Prebunking Elections Rumors: Artificial Intelligence Assisted Interventions Increase Confidence in American Elections*. https://doi.org/10.7910/DVN/BLC8KO. Dataset. DOI: https://doi.org/10.7910/DVN/BLC8KO.

Lockhart, Mackenzie et al. (Oct. 2024). "Voters distrust delayed election results, but a prebunking message inoculates against distrust". In: *PNAS Nexus* 3.10, pgae414. DOI: 10.1093/pnasnexus/pgae414.

McGuire, William J (1961). "The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion". In: *Sociometry* 24.2, pp. 184–197.

— (1964). "Some Contemporary Approaches". In: *Advances in Experimental Social Psychology*. Ed. by Leonard Berkowitz. Vol. 1. Academic Press, pp. 191–229. DOI: https://doi.org/10.1016/S0065-2601(08)60052-0.

Minnite, Lorraine C (2011). *The Myth of Voter Fraud*. Cornell University Press.

Norris, Pippa (2014). *Why Electoral Integrity Matters*. Cambridge University Press.

Nyhan, Brendan (2021). "Why the backfire effect does not explain the durability of political misperceptions". In: *Proceedings of the National Academy of Sciences* 118.15, e1912440117.

Nyhan, Brendan and Jason Reifler (2010). "When corrections fail: The persistence of political misperceptions". In: *Political Behavior* 32.2, pp. 303–330.

Papageorgis, Demetrios and William J McGuire (1961). "The generality of immunity to persuasion produced by pre-exposure to weakened counterarguments." In: *The Journal of Abnormal and Social Psychology* 62.3, p. 475.

Pereira, Frederico Batista et al. (2024). "Inoculation reduces misinformation: Experimental evidence from multidimensional interventions in Brazil". In: *Journal of Experimental Political Science* 11.3, pp. 239–250.

Roozenbeek, Jon, Cecilie S Traberg, and Sander van der Linden (2022). "Technique-based inoculation against real-world misinformation". In: *Royal Society open science* 9.5, p. 211719.

Roozenbeek, Jon and Sander van der Linden (2024). *The psychology of misinformation*. Cambridge University Press.

Sances, Michael W and Charles Stewart III (2015). "Partisanship and confidence in the vote count: Evidence from US national elections since 2000". In: *Electoral Studies* 40, pp. 176–188.

Shino, Enrijeta and Daniel A. Smith (2025). "Vote Method and Confidence in Elections". In: *Political Research Quarterly* 78.2, pp. 568–584. DOI: 10.1177/10659129241309658.

Simchon, Almog, Matthew Edwards, and Stephan Lewandowsky (2024). "The persuasive effects of political microtargeting in the age of generative artificial intelligence". In: *PNAS Nexus* 3 (2). DOI: https://doi.org/10.1093/pnasnexus/pgae035.

Sinclair, Betsy, Steven S Smith, and Patrick D Tucker (2018). ""It's largely a rigged system": voter confidence and the winner effect in 2016". In: *Political Research Quarterly* 71.4, pp. 854–868.

Spearing, Emily R. et al. (June 2025). "Countering AI-generated misinformation with preemptive source discreditation and debunking". In: *Royal Society Open Science* 12.6, p. 242148. ISSN: 2054-5703. DOI: 10.1098/rsos.242148.

Suttmann-Lea, Mara and Thessalia Merivaki (2023). "The Impact of Voter Education on Voter Confidence: Evidence from the 2020 U.S. Presidential Election". In: *Election Law Journal: Rules, Politics, and Policy* 22.2, pp. 145–165. DOI: 10.1089/elj.2022.0055.

Traberg, Charlotte S., Jon Roozenbeek, and Sander van der Linden (2022). "Psychological inoculation against misinformation: Current evidence and future directions". In: *The ANNALS of the American Academy of Political and Social Science* 700.1, pp. 136–151. DOI: 10.1177/00027162221097037.

Uscinski, Joseph E and Joseph M Parent (2014). *American conspiracy theories*. Oxford University Press.

van der Linden, S (2023). *Foolproof: Why Misinformation Infects Our Minds and How to Build Immunity*. WW Norton.

van der Linden, Sander (2022). "Misinformation: Susceptibility, spread, and interventions to immunize the public". In: *Nature Medicine* 28, pp. 460–467. DOI: https://doi.org/10.1038/s41591-022-01713-6.

— (2024). "Countering misinformation through psychological inoculation". In: *Advances in Experimental Social Psychology* 69, pp. 1–58.

van der Linden, Sander, Almog Simchon, and Stephan Lewandowsky (2026). "New Research Shows Cognitive Inoculation Improves Discernment of Misinformation". In: *Skeptical Inquirer*, p. 55.

Velez, Yamil R., Donald P. Green, and Semra Sevi (2025). "Chatbot Voting Advice Applications inform but seldom sway young unaligned voters". In: *Proceedings of the National Academy of Sciences* 122.50, e2515516122. DOI: 10.1073/pnas.2515516122.

Vonnahme, Greg and Beth Miller (2013). "Candidate Cues and Voter Confidence in American Elections". In: *Journal of Elections, Public Opinion and Parties* 23.2, pp. 223–239. DOI: 10.1080/17457289.2012.743466.

Wellman, Elizabeth Iams, Susan D. Hyde, and Thad E. Hall (2018). "Does fraud trump partisanship? The impact of contentious elections on voter confidence". In: *Journal of Elections, Public Opinion and Parties* 28.3, pp. 330–348. DOI: 10.1080/17457289.2017.1394865.

8 Survey, Code, and Data Availability

The exact wording of the questions for our outcome variables is contained in the Supplementary Materials. For convenience, A.6 summarizes the core outcome wording and scale construction used in the main analysis. Additionally, the full text of every question asked, along with the internal survey-chatbot interaction and routing logic, is included in the online supplementary material.

9 Acknowledgements

Funding RMA and ML's work on this project, and the data collection, was supported by a grant to the California Institute of Technology by the John Randolph Haynes and Dora Haynes Foundation.

Author Contributions ML, BS, SvdL and RMA conceived the research strategy and methodology. ML conducted the analysis. RMA oversaw funding and project management. ML and RMA drafted the paper. All authors contributed to writing and editing of the paper.

Competing Interests The authors declare no competing interests.

Ethical Considerations The data collection and analyses in this paper were reviewed by the Institutional Review Board at the California Institute of Technology (IR24-1456). The protocol required respondents to evaluate election-related claims as part of outcome measurement. Respondents read real, published articles containing those claims, but the study did not present those claims as true or rely on a deceptive cover story. We also did not add a separate, study-wide corrective module beyond the assigned intervention materials. The follow-up occurred after the election, and by the end of that wave we saw little evidence of broad increases in rumor confidence, so we avoided re-presenting the false claims to every respondent.

Preregistration This study was preregistered on OSF: `https://osf.io/jhezb`.

There we specify four hypotheses: that inoculation decreases rumor confidence, increases election confidence (we specify three levels, but focus on one in the text for space constraints), that chatbot-based inoculations have a stronger treatment effect than article-based, and that effects persist for at least one week for the chatbot, but not necessarily the article. Finally, we note that while we adhered to the timing outlined in the preregistration, we did not specifically preregister that the follow-up survey would happen after the election.

A Supplementary Material

Here we collect supplementary analyses, robustness checks, measurement details, and implementation materials that elaborate on the main text and document the broader set of model variants. The appendix begins with the primary specifications, sample accounting, missing-data checks, placebo-order diagnostics, and transparency materials cited most directly in the paper, and then turns to additional specification checks, expanded heterogeneity tables, and supplementary figures for readers who want the fuller set of supporting analyses. All code and data necessary to replicate our analysis is available.

A.1 Ethical Protocol

The study protocol was reviewed and approved by the California Institute of Technology IRB (IR24-1456). Respondents received a pre-bunking or placebo intervention before reading two real, published articles containing election-rumor claims, which allowed us to measure rumor confidence and election confidence in a realistic information environment. The study did not present those claims as true, did not rely on a deceptive cover story, and did not add a separate study-wide corrective module after exposure. This choice reflects both practical and theoretical considerations: by the end of the post-election follow-up wave we saw little sign of broad increases in rumor confidence, and prior research has documented that repeating false claims — even in the context of corrections — can sometimes reinforce belief in the original misinformation (Nyhan and Reifler, 2010; Lewandowsky, Ecker, et al., 2012), though more recent work has questioned the prevalence of such backfire effects (Nyhan, 2021). In light of this evidence, repeating the false claims to all respondents in a post-hoc corrective would have created unnecessary reactivation risk with uncertain benefit.

A.2 Alternative Model Specifications

This section reports the full regression tables underlying the main-text results. All models combine article and chatbot delivery into a single treatment arm and estimate treatment-minus-placebo effects using survey-weighted OLS with the preregistered covariate set.

Table A.1 reports the four primary models in the full analytic sample. Table A.2 compares those estimates with the engaged subsample — which excludes chatbot respondents with low engagement (fewer than three messages sent) and article respondents who spent fewer than five seconds on the intervention pages — overall and within each party group; Table A.3 reports the corresponding full models for the engaged subsample. Figure A.1 summarizes the over-time trajectory pattern at the outcome-family level.

Tables A.4–A.6 extend the same specification to ballot-confidence outcomes and non-assigned rumor confidence. Reduced-control variants appear in Tables A.20–A.22 and minimal-control variants in Tables A.23–A.25.

Table A.1: Assigned-rumor and election confidence models (full sample)

	Post-Treatment Rumor Confidence	Post-Election Rumor Confidence	Post-Treatment Election Confidence	Post-Election Election Confidence
Pre Confidence In Rumor	0.644*** (0.017)	0.459*** (0.021)		
Pre Confidence In Election			0.821*** (0.011)	0.445*** (0.020)
Treatment (vs. Placebo)	−0.320*** (0.054)	−0.200* (0.079)	0.163*** (0.039)	0.077 (0.070)
Topic: Voter Rolls	−0.043 (0.054)	0.174* (0.078)	−0.092* (0.039)	0.135 (0.069)
Age Group: 30-44	0.023 (0.092)	0.002 (0.143)	0.034 (0.063)	−0.260* (0.115)
Age Group: 45-64	0.121 (0.087)	0.134 (0.137)	−0.038 (0.060)	−0.184 (0.106)
Age Group: 65+	0.306** (0.094)	0.154 (0.141)	−0.104 (0.063)	−0.213* (0.108)
Gender: Female	0.187*** (0.055)	0.164* (0.081)	−0.095* (0.040)	−0.113 (0.071)
Race Ethnicity: Black	−0.236* (0.112)	0.237 (0.146)	0.086 (0.068)	−0.389** (0.131)
Race Ethnicity: Hispanic	−0.032 (0.097)	−0.001 (0.131)	0.163** (0.062)	−0.058 (0.117)
Race Ethnicity: Other	−0.239* (0.108)	0.380* (0.171)	0.006 (0.071)	−0.241 (0.171)
Education Level: Some college	−0.161* (0.076)	−0.036 (0.107)	−0.007 (0.055)	0.006 (0.096)
Education Level: College grad	−0.027 (0.075)	−0.169 (0.115)	0.007 (0.054)	0.191 (0.102)
Education Level: Postgrad	−0.241** (0.084)	−0.270* (0.122)	0.044 (0.058)	0.259* (0.109)
Party Identification: Independent	0.168* (0.080)	−0.188 (0.109)	−0.188*** (0.052)	0.390*** (0.111)
Party Identification: Republican	0.312** (0.106)	−0.171 (0.147)	−0.176* (0.074)	0.821*** (0.118)
Ideology: Moderate	0.071 (0.083)	0.176 (0.112)	−0.039 (0.052)	0.229* (0.110)
Ideology: Conservative	0.470*** (0.106)	0.389* (0.152)	−0.236** (0.075)	0.502*** (0.123)
Region: Midwest	0.013 (0.085)	−0.088 (0.128)	−0.097 (0.059)	0.006 (0.114)
Region: South	0.064 (0.081)	−0.138 (0.121)	−0.112* (0.055)	−0.123 (0.106)
Region: West	0.041 (0.085)	0.020 (0.128)	−0.078 (0.058)	−0.102 (0.113)
Urban Rural: Suburb	0.104 (0.068)	−0.075 (0.102)	−0.033 (0.048)	0.055 (0.088)
Urban Rural: Town	0.073 (0.088)	−0.187 (0.128)	−0.096 (0.065)	0.109 (0.112)
Urban Rural: Rural area	0.050 (0.087)	−0.045 (0.130)	−0.051 (0.060)	−0.021 (0.121)
Political Interest: Some of the time	0.054 (0.064)	0.014 (0.091)	−0.018 (0.046)	0.114 (0.079)
Political Interest: Only now and then	0.010 (0.105)	0.191 (0.161)	0.002 (0.072)	−0.006 (0.138)
Political Interest: Hardly at all	0.315 (0.163)	0.111 (0.212)	−0.200 (0.121)	0.106 (0.197)
Conspiracy Score	−0.105*** (0.007)	−0.117*** (0.010)	0.034*** (0.004)	−0.004 (0.006)
Constant	3.411*** (0.255)	4.385*** (0.330)	0.781*** (0.140)	3.828*** (0.249)
Observations	5,191	4,203	5,190	4,203

Notes: Entries are weighted OLS coefficients with design-based standard errors from `svyglm`. The four columns correspond to the primary models summarized in Table 1. All outcomes are measured on 0–10 scales. These ITT models retain all model-specific complete cases with observed outcomes, so assignment is the treatment regressor in the full analytic sample. Post-election models use the recontact survey weights. Model-specific N varies with item nonresponse and follow-up response. Appendix Table A.3 reports the corresponding engaged-subsample models. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table A.2: Treatment estimates by party: full sample and engaged subsample

Group	Specification	Post-treatment rumor confidence	Post-election rumor confidence	Post-treatment election confidence	Post-election election confidence
	Full sample				
Overall	(ITT)	-0.320*** (0.054)	-0.200* (0.079)	0.163*** (0.039)	0.077 (0.070)
	Engaged subsample	-0.463*** (0.064)	-0.217* (0.093)	0.177*** (0.045)	0.087 (0.081)
	Estimation sample N (full / engaged)	5,191 / 3,782	4,203 / 3,083	5,190 / 3,782	4,203 / 3,083
	Full sample				
Democrats	(ITT)	-0.291** (0.089)	-0.066 (0.123)	0.111 (0.057)	0.094 (0.123)
	Engaged subsample	-0.460*** (0.101)	-0.044 (0.138)	0.147* (0.065)	0.120 (0.136)
	Estimation sample N (full / engaged)	2,032 / 1,474	1,614 / 1,173	2,031 / 1,474	1,613 / 1,173
	Full sample				
Independents	(ITT)	-0.296*** (0.089)	-0.350* (0.146)	0.135 (0.070)	0.070 (0.135)
	Engaged subsample	-0.406*** (0.106)	-0.333 (0.175)	0.139 (0.083)	0.061 (0.164)
	Estimation sample N (full / engaged)	1,428 / 1,069	1,205 / 908	1,428 / 1,069	1,205 / 908
	Full sample				
Republicans	(ITT)	-0.313** (0.095)	-0.192 (0.135)	0.245*** (0.071)	0.056 (0.097)
	Engaged subsample	-0.460*** (0.118)	-0.267 (0.162)	0.236** (0.085)	0.057 (0.112)
	Estimation sample N (full / engaged)	1,731 / 1,239	1,384 / 1,002	1,731 / 1,239	1,385 / 1,002

Notes: All entries are weighted OLS estimates. Within each group block, the full-sample row reports the ITT estimates for that sample using the primary covariates and functional form; the party-specific blocks omit party identification because it is constant within subgroup. The engaged-subsample row uses that same OLS specification but additionally excludes chatbot-assigned respondents with low engagement (fewer than three messages sent) and article-assigned respondents who spent fewer than five seconds on either assigned intervention page; for article-assigned respondents, the timing rule is applied to the assigned treatment pages in the treatment arm and to the assigned placebo pages in the control arm. The engaged-subsample estimates are generally stronger in the expected direction, consistent with attenuation in the ITT estimates from respondents with little or no meaningful exposure to assigned content. We therefore treat ITT as primary and the engaged-subsample rows as a robustness check rather than as a competing estimand. Post-election models use the recontact survey weights. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table A.3: Assigned-rumor and election confidence models (engaged subsample)

	Post-Treatment Rumor Confidence	Post-Election Rumor Confidence	Post-Treatment Election Confidence	Post-Election Election Confidence
Pre Confidence In Rumor	0.641*** (0.019)	0.456*** (0.025)		
Pre Confidence In Election			0.814*** (0.013)	0.431*** (0.024)
Treatment (vs. Placebo)	−0.463*** (0.064)	−0.217* (0.093)	0.177*** (0.045)	0.087 (0.081)
Topic: Voter Rolls	−0.088 (0.064)	0.144 (0.091)	−0.043 (0.046)	0.138 (0.081)
Age Group: 30-44	0.025 (0.114)	−0.004 (0.178)	0.028 (0.075)	−0.239 (0.134)
Age Group: 45-64	0.080 (0.109)	0.199 (0.172)	−0.020 (0.072)	−0.179 (0.128)
Age Group: 65+	0.319** (0.117)	0.197 (0.174)	−0.111 (0.077)	−0.244 (0.128)
Gender: Female	0.217*** (0.065)	0.107 (0.096)	−0.068 (0.048)	−0.129 (0.084)
Race Ethnicity: Black	−0.283* (0.133)	0.282 (0.176)	0.110 (0.081)	−0.512*** (0.153)
Race Ethnicity: Hispanic	0.019 (0.118)	−0.013 (0.152)	0.191** (0.070)	−0.036 (0.140)
Race Ethnicity: Other	−0.224 (0.130)	0.492* (0.221)	−0.033 (0.089)	−0.378 (0.210)
Education Level: Some college	−0.137 (0.092)	0.020 (0.128)	−0.063 (0.066)	−0.011 (0.116)
Education Level: College grad	−0.041 (0.091)	−0.128 (0.139)	0.011 (0.063)	0.184 (0.119)
Education Level: Postgrad	−0.257* (0.102)	−0.189 (0.143)	0.0004 (0.068)	0.360** (0.125)
Party Identification: Independent	0.118 (0.094)	−0.142 (0.125)	−0.128* (0.058)	0.375** (0.133)
Party Identification: Republican	0.340** (0.131)	−0.184 (0.179)	−0.079 (0.088)	0.748*** (0.146)
Ideology: Moderate	0.139 (0.097)	0.161 (0.130)	−0.076 (0.059)	0.178 (0.129)
Ideology: Conservative	0.479*** (0.132)	0.379* (0.181)	−0.276** (0.087)	0.527*** (0.148)
Region: Midwest	−0.022 (0.104)	−0.115 (0.152)	−0.071 (0.070)	0.044 (0.131)
Region: South	0.054 (0.098)	−0.114 (0.144)	−0.117 (0.065)	−0.093 (0.125)
Region: West	0.037 (0.104)	0.068 (0.153)	−0.063 (0.068)	−0.159 (0.134)
Urban Rural: Suburb	0.084 (0.081)	−0.065 (0.120)	−0.031 (0.057)	0.054 (0.100)
Urban Rural: Town	0.056 (0.108)	−0.136 (0.153)	−0.124 (0.078)	0.145 (0.126)
Urban Rural: Rural area	0.106 (0.102)	−0.064 (0.159)	−0.020 (0.070)	−0.077 (0.145)
Political Interest: Some of the time	0.070 (0.076)	0.090 (0.109)	0.007 (0.054)	0.109 (0.094)
Political Interest: Only now and then	−0.056 (0.131)	0.095 (0.211)	−0.011 (0.083)	0.061 (0.175)
Political Interest: Hardly at all	0.230 (0.206)	0.259 (0.274)	−0.196 (0.147)	0.223 (0.207)
Conspiracy Score	−0.102*** (0.009)	−0.120*** (0.012)	0.034*** (0.005)	−0.002 (0.008)
Constant	3.410*** (0.305)	4.361*** (0.395)	0.773*** (0.168)	3.949*** (0.291)
Observations	3,782	3,083	3,782	3,083

Notes: Entries are weighted OLS coefficients with design-based standard errors from `svyglm`. These columns use the same covariates and functional form as Table A.1 but additionally exclude chatbot-assigned respondents with low engagement (fewer than three messages sent) and article-assigned respondents who spent fewer than five seconds on either assigned intervention page. Post-election models use the recontact survey weights. Model-specific N varies with item nonresponse and follow-up response. These estimates are directionally stronger in the expected way because they omit cases with little or no meaningful exposure to assigned content. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

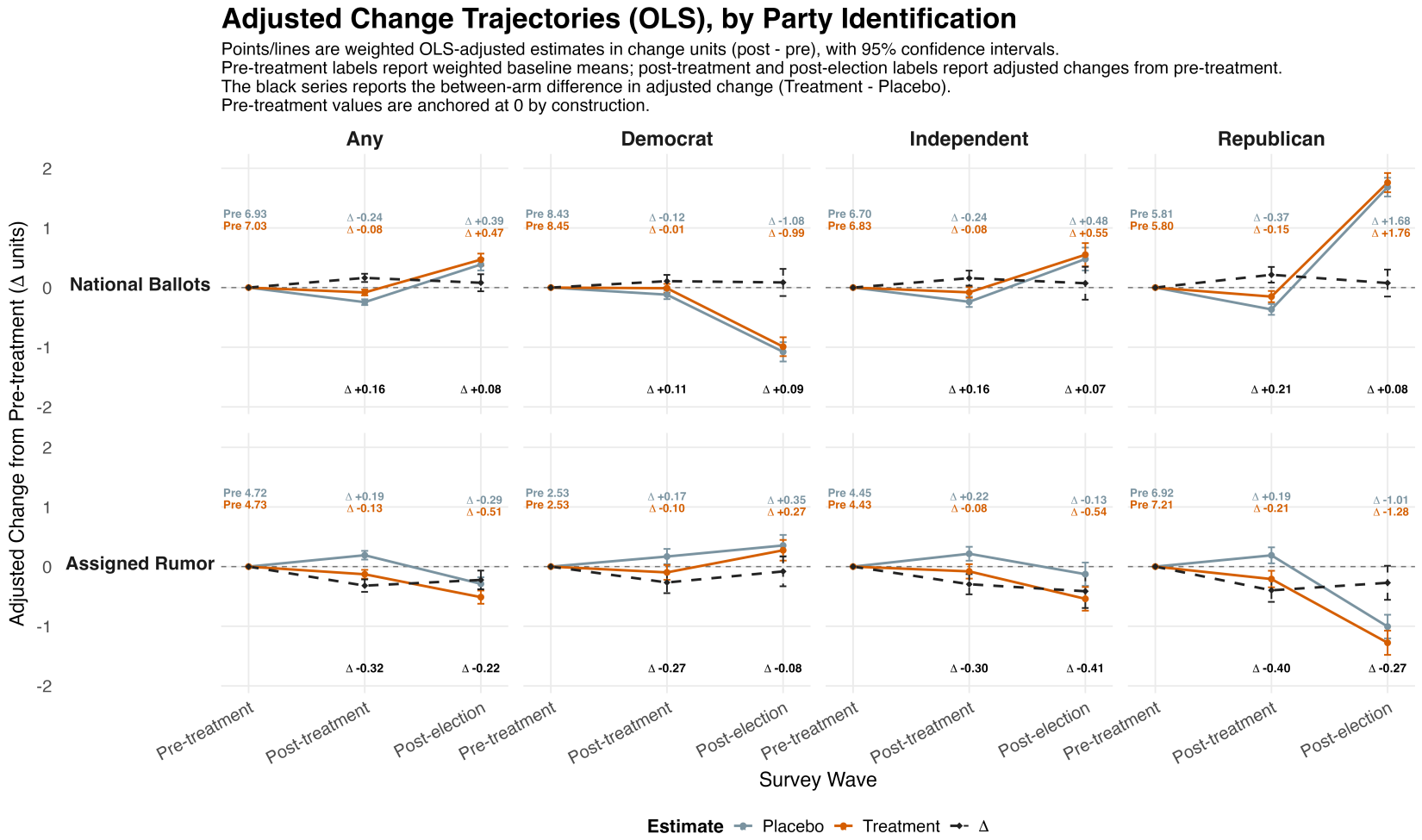


Figure A.1: **Adjusted Change Trajectories by Party and Outcome Family.** Weighted regression-based changes are shown for rumor confidence and national-ballot confidence across waves. Placebo and treatment series are plotted alongside mean differences (Δ = Treatment − Placebo) with 95% confidence intervals. The figure summarizes the same over-time pattern highlighted in the main text: post-election divergence is sharper in election confidence than in rumor confidence.

Table A.4: Ballot-confidence models (full sample)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.834*** (0.014)		
Pre confidence: County ballots		0.805*** (0.013)	
Pre confidence: National ballots			0.820*** (0.011)
Treatment (vs. Placebo)	0.065 (0.038)	0.044 (0.036)	0.172*** (0.038)
Topic: Voter Rolls	−0.154*** (0.038)	−0.086* (0.037)	−0.084* (0.038)
Age group: 45-64	−0.021 (0.056)	−0.046 (0.051)	−0.048 (0.055)
Age group: 65+	−0.009 (0.056)	−0.036 (0.050)	−0.111 (0.058)
Age group: Under 30	−0.070 (0.060)	−0.013 (0.057)	−0.014 (0.062)
Gender: Male	0.072 (0.040)	0.072 (0.038)	0.097* (0.039)
Race/ethnicity: Hispanic	−0.106 (0.074)	−0.121 (0.079)	0.125 (0.080)
Race/ethnicity: Other	−0.167 (0.100)	−0.091 (0.095)	−0.050 (0.087)
Race/ethnicity: White	−0.080 (0.057)	−0.145* (0.060)	−0.067 (0.066)
Education: HS or less	−0.048 (0.054)	−0.047 (0.052)	−0.011 (0.054)
Education: Postgrad	0.066 (0.055)	0.013 (0.054)	0.020 (0.054)
Education: Some college	−0.007 (0.050)	−0.026 (0.048)	−0.026 (0.052)
Party ID: Independent	−0.095* (0.046)	−0.095* (0.046)	−0.191*** (0.051)
Party ID: Republican	−0.156* (0.064)	−0.148* (0.066)	−0.226** (0.072)
Ideology: : Conservative	−0.006 (0.142)	0.218 (0.120)	0.115 (0.142)
Ideology: : Liberal	0.143 (0.141)	0.374** (0.116)	0.316* (0.136)
Ideology: : Moderate	0.088 (0.137)	0.284* (0.114)	0.283* (0.134)
Region: : Northeast	−0.097 (0.062)	−0.019 (0.057)	0.093 (0.059)
Region: : South	−0.051 (0.052)	0.023 (0.049)	−0.011 (0.054)
Region: : West	−0.030 (0.056)	0.006 (0.051)	0.033 (0.056)
Urbanicity: City	0.122 (0.300)	0.190 (0.304)	−0.128 (0.210)
Urbanicity: Rural area	0.034 (0.302)	0.260 (0.306)	−0.200 (0.210)
Urbanicity: Suburb	0.030 (0.299)	0.214 (0.303)	−0.172 (0.208)
Urbanicity: Town	0.021 (0.301)	0.248 (0.305)	−0.242 (0.212)
Political interest: Hardly at all	0.069 (0.251)	0.112 (0.213)	−0.219 (0.214)
Political interest: Most of the time	0.119 (0.210)	0.250 (0.188)	−0.012 (0.188)
Political interest: Only now and then	0.126 (0.213)	0.175 (0.193)	−0.006 (0.195)
Political interest: Some of the time	0.083 (0.208)	0.199 (0.188)	−0.027 (0.188)
Conspiracy score	0.032*** (0.004)	0.031*** (0.003)	0.034*** (0.004)
Constant	0.489 (0.386)	0.294 (0.390)	0.522 (0.315)
Observations	5,430	5,431	5,430
Log Likelihood	−9,054.273	−9,027.454	−9,112.280
Akaike Inf. Crit.	18,170.550	18,116.910	18,286.560

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.5: Ballot-confidence models at recontact (full sample)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.443*** (0.021)		
Pre confidence: County ballots		0.440*** (0.021)	
Pre confidence: National ballots			0.436*** (0.020)
Treatment (vs. Placebo)	0.007 (0.060)	−0.067 (0.062)	0.093 (0.069)
Topic: Voter Rolls	0.074 (0.060)	0.097 (0.062)	0.142* (0.069)
Age group: 45-64	0.110 (0.092)	0.124 (0.095)	0.067 (0.106)
Age group: 65+	0.244** (0.094)	0.212* (0.094)	0.057 (0.107)
Age group: Under 30	−0.019 (0.106)	0.092 (0.112)	0.229* (0.116)
Gender: Male	0.125* (0.061)	0.178** (0.062)	0.142* (0.071)
Race/ethnicity: Hispanic	0.394** (0.142)	0.338* (0.154)	0.479** (0.162)
Race/ethnicity: Other	0.260 (0.177)	0.168 (0.183)	0.174 (0.207)
Race/ethnicity: White	0.456*** (0.114)	0.417*** (0.118)	0.500*** (0.130)
Education: HS or less	−0.213* (0.085)	−0.209* (0.089)	−0.206* (0.101)
Education: Postgrad	0.011 (0.087)	0.093 (0.090)	0.046 (0.103)
Education: Some college	−0.155 (0.081)	−0.158 (0.082)	−0.199* (0.097)
Party ID: Independent	0.258** (0.093)	0.324*** (0.094)	0.423*** (0.112)
Party ID: Republican	0.548*** (0.099)	0.572*** (0.105)	0.832*** (0.120)
Ideology: : Conservative	0.733** (0.231)	0.695* (0.275)	0.629** (0.234)
Ideology: : Liberal	0.441 (0.242)	0.416 (0.273)	0.173 (0.250)
Ideology: : Moderate	0.465* (0.232)	0.466 (0.266)	0.383 (0.233)
Region: : Northeast	0.029 (0.103)	−0.031 (0.107)	0.001 (0.113)
Region: : South	−0.003 (0.081)	−0.049 (0.084)	−0.161 (0.093)
Region: : West	−0.083 (0.087)	−0.170 (0.089)	−0.130 (0.099)
Urbanicity: City	1.483 (1.178)	1.381 (1.202)	1.626 (1.105)
Urbanicity: Rural area	1.491 (1.179)	1.467 (1.202)	1.607 (1.105)
Urbanicity: Suburb	1.504 (1.179)	1.452 (1.203)	1.651 (1.105)
Urbanicity: Town	1.594 (1.178)	1.606 (1.202)	1.697 (1.105)
Political interest: Hardly at all	0.273 (0.399)	0.010 (0.404)	0.273 (0.395)
Political interest: Most of the time	0.468 (0.355)	0.199 (0.370)	0.299 (0.356)
Political interest: Only now and then	0.312 (0.370)	0.068 (0.386)	0.298 (0.371)
Political interest: Some of the time	0.408 (0.355)	0.155 (0.373)	0.417 (0.356)
Conspiracy score	0.006 (0.005)	0.010 (0.006)	−0.004 (0.006)
Constant	1.447 (1.226)	1.646 (1.269)	1.178 (1.163)
Observations	4,384	4,383	4,384
Log Likelihood	−8,917.705	−9,034.738	−9,612.392
Akaike Inf. Crit.	17,897.410	18,131.470	19,286.780

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.6: Non-assigned rumor confidence models (full sample)

	Post Other Rumor	Recontact Other Rumor
Pre confidence: Other rumor	0.729*** (0.013)	0.485*** (0.018)
Treatment (vs. Placebo)	−0.480*** (0.088)	−0.436** (0.139)
Topic: Voter Rolls	0.209* (0.088)	0.174 (0.138)
Age group: 45-64	0.122 (0.129)	0.228 (0.196)
Age group: 65+	0.349* (0.141)	0.431* (0.203)
Age group: Under 30	−0.022 (0.149)	−0.197 (0.262)
Gender: Male	−0.400*** (0.091)	−0.253 (0.142)
Race/ethnicity: Hispanic	0.006 (0.228)	−0.357 (0.330)
Race/ethnicity: Other	0.001 (0.235)	0.002 (0.393)
Race/ethnicity: White	0.208 (0.181)	−0.604* (0.276)
Education: HS or less	0.150 (0.130)	0.336 (0.206)
Education: Postgrad	−0.242 (0.137)	−0.216 (0.200)
Education: Some college	−0.017 (0.116)	0.220 (0.194)
Party ID: Independent	0.246 (0.133)	−0.272 (0.193)
Party ID: Republican	0.374* (0.161)	−0.314 (0.248)
Ideology: : Conservative	0.283 (0.273)	0.738 (0.470)
Ideology: : Liberal	−0.707* (0.283)	0.041 (0.463)
Ideology: : Moderate	−0.674* (0.262)	0.315 (0.450)
Region: : Northeast	0.340* (0.139)	0.357 (0.225)
Region: : South	0.319** (0.113)	0.160 (0.189)
Region: : West	0.224 (0.121)	0.280 (0.206)
Urbanicity: City	−0.004 (0.458)	−0.675 (1.269)
Urbanicity: Rural area	−0.003 (0.461)	−0.726 (1.270)
Urbanicity: Suburb	0.088 (0.454)	−0.695 (1.269)
Urbanicity: Town	0.087 (0.465)	−0.997 (1.274)
Political interest: Hardly at all	0.744 (0.542)	0.578 (0.827)
Political interest: Most of the time	0.708 (0.488)	0.356 (0.764)
Political interest: Only now and then	0.667 (0.496)	0.370 (0.789)
Political interest: Some of the time	0.712 (0.486)	0.417 (0.766)
Conspiracy score	−0.150*** (0.012)	−0.210*** (0.017)
Constant	4.714*** (0.782)	8.746*** (1.557)
Observations	5,427	4,380
Log Likelihood	−13,833.460	−12,582.010
Akaike Inf. Crit.	27,728.920	25,226.030
Note:		*p<0.05; **p<0.01; ***p<0.001

A.3 Wave 1 Dropoff and Disposition Concerns

These tables document where the Wave 1 sample narrows from survey starts to observed outcomes, and where treatment delivery is imperfect due to the external chatbot handoff. Overall Wave 1 attrition is mostly ordinary breakoff, while chatbot engagement differs modestly by assigned arm; we therefore report engagement rates and related robustness checks alongside the primary ITT estimates and the engaged-subsample checks.

Wave 1 accounting. Tables A.7–A.10 summarize the Wave 1 accounting: 8,428 starts, 8,093 randomized respondents, 2,662 randomized respondents who do not appear in the final-complete count, and 5,431 final completes. Broad disposition categories are similar across assigned arms, though the "Respondent did not complete" category is slightly larger in treatment (30.1% vs. 28.0% of randomized cases). The formal randomized-arm completion and compliance tests are summarized in Table A.14. The analytic Wave 1 sample contains 5,432 respondents because it retains one additional respondent with missing post-treatment outcomes whom YouGov does not count as a final complete.

Stage	N
Total survey starts	8,428
Randomized (Control: 4,066; Treatment: 4,027)	8,093
Not treated (pre-randomization)	335
Removed by YouGov (among randomized)	2,662
Breakoffs	2,352
Removals	310
Final delivered sample	5,431

Table A.7: Sample Reconciliation, Wave 1

Disposition	N	Percent
Final sample	5431	64.4%
Breakoffs	2680	31.8%
Removals	310	3.7%
Screenouts	7	0.1%

Table A.8: Disposition Summary, Wave 1

Disposition	Control	Not treated	Treatment	Total
Breakoffs	1139	328	1213	2680
Final sample	2762	0	2669	5431
Removals	165	0	145	310
Screenouts	0	7	0	7

Table A.9: Disposition by Assigned Arm, Wave 1

Disposition	Control	Treatment	Total
Final sample	2762 (67.9%)	2669 (66.3%)	5431
Breakoffs: Respondent did not complete	1139 (28.0%)	1213 (30.1%)	2352
Removals: Response quality control failure	52 (1.3%)	37 (0.9%)	89
Removals: Speeders - Under 5 minutes (median interview 17:13)	113 (2.8%)	108 (2.7%)	221

Table A.10: Detailed disposition by assigned arm, Wave 1. Control and Treatment entries report counts with within-arm percentages in parentheses. The table is restricted to randomized respondents; pre-randomization "Not treated" cases are reported in the reconciliation and flow tables instead. Formal arm-balance tests are summarized in the later attrition/compliance table.

Post-randomization compliance. Tables A.11 and A.12 summarize engagement in the external chatbot interface. Among the 4,808 respondents assigned to chatbot, 1,480 (30.8%) had low engagement (fewer than three messages sent). Engagement is 71% in chatbot placebo and 67% in chatbot treatment, with the same directional pattern across party groups; the companion unadjusted independence test appears in Table A.14. Because these respondents and article-assigned respondents with trivial page time had little or no meaningful exposure to the intervention, the engaged-subsample estimates are directionally stronger. We therefore treat pooled ITT estimates as primary and read the engaged-subsample results as a robustness check rather than as a competing estimand.

Stage	N	Percent
Delivered by YouGov (final cleaned file)	5432	100.0%
Assigned to Chatbot modality	4808	88.5%
Assigned to Article modality	624	11.5%
Chatbot: Engaged ($\geq$3 messages sent)	3328	61.3%
Chatbot: Low engagement ($<$3 messages sent)	1480	27.2%
Received-content sample (excludes low-engagement chatbot respondents)	3952	72.8%

Table A.11: Wave 1 post-randomization engagement (chatbot handoff)

Arm	Party	N	Engaged	Engagement Rate
Placebo	Democrat	947	681	0.72
Placebo	Independent/Other	674	481	0.71
Placebo	Republican	820	581	0.71
Treatment	Democrat	940	628	0.67
Treatment	Independent/Other	660	464	0.70
Treatment	Republican	767	493	0.64

Table A.12: Chatbot engagement by assigned arm and party

Bounds and simple tests. Table A.13 reports the observed unadjusted treatment-control differences together with Lee bounds (Lee, 2009) under the standard monotonicity assumption. The sign of the key contrasts is unchanged within those bounds, which suggests that marginal selection does not overturn the basic direction of the estimates. Table A.14 then reports the companion unadjusted summary independence tests: overall final-completion status is not significantly associated with arm among randomized respondents, but chatbot engagement is. We therefore report both the primary ITT estimates and the engaged-subsample checks, and interpret the bounds as sensitivity checks for treatment delivery and attrition. The expanded-sample reruns and subgroup completion checks appear later in the appendix.

Outcome	Observed unadjusted treatment-control difference	Lee lower bound	Lee upper bound	Treatment completion rate	Placebo completion rate	Trim share	Trimmed arm
Post-treatment election confidence	0.26	0.09	0.34	0.66	0.68	0.02	Placebo
Election-confidence change	0.14	0.03	0.23	0.66	0.68	0.02	Placebo
Rumor-confidence change	-0.31	-0.48	-0.14	0.66	0.68	0.02	Placebo

Table A.13: Lee bounds for Wave 1 differential completion under the standard monotonicity assumption (Lee, 2009). The first three substantive columns report the observed unadjusted treatment-control difference and the corresponding Lee lower and upper bounds. The remaining columns report arm-specific completion rates, the implied trim share, and the arm trimmed to equalize completion.

Null hypothesis	p-value
H_0: final-completion status is independent of assigned arm (Placebo vs. Treatment) among randomized Wave 1 starts	0.1194
H_0: chatbot engagement (interacted with chatbot) is independent of assigned arm (Placebo vs. Treatment) among chatbot-assigned respondents	0.0009

Table A.14: Unadjusted Pearson χ^2 tests of independence. The first row compares final completion status across Placebo and Treatment among randomized Wave 1 starts. The second compares chatbot engagement (interacted with chatbot) across Placebo and Treatment among chatbot-assigned respondents. Smaller p-values indicate more evidence that completion or engagement differs by assigned arm.

A.4 Multiple-Imputation Sensitivity for Recontact Missingness

Missingness is concentrated in the follow-up outcomes rather than in baseline or immediate post-treatment measures, so we estimate a chained-equations multiple-imputation sensitivity model for recontact attrition. The imputation model uses treatment assignment, AI modality, assigned topic, baseline measures, immediate post-treatment outcomes, and the covariates from the main specifications; treatment assignment, AI modality, and topic are treated as fixed design variables and are not imputed.

The imputed estimates remain close to the complete-case results and do not alter the qualitative interpretation. This table focuses only on recontact outcomes, where effects are weaker than in the immediate post-treatment models emphasized in the main text and where roughly one-fifth of cases are missing. Rumor-confidence reductions remain the more durable pattern, while follow-up election-confidence estimates are weaker and more variable. We therefore retain the preregistered complete-case models as primary and treat imputation as a robustness check under plausible missing-at-random assumptions.

Table A.15: Follow-Up Outcome Missingness

Variable	Missing n	Missing (%)
Post-election confidence: Own ballot	1047	19.27
Post-election confidence: County ballots	1049	19.31
Post-election confidence: National election	1047	19.27
Post-election confidence: Assigned rumor	1049	19.31
Post-election confidence: Other rumor	1051	19.35
Post-election confidence change: Assigned rumor	1049	19.31
Post-election confidence change: Other rumor	1052	19.37

Notes: Missingness is concentrated in post-election (recontact) outcomes; pre-treatment and immediate post-treatment outcomes are nearly complete.

Table A.16: Recontact Estimates: Complete Case vs. Multiple Imputation

Specification	Outcome	Complete-case treatment coefficient	MI treatment coefficient	Coefficient shift	Missing (%)
Minimal specification					
Minimal	Assigned rumor (post-election level)	-0.20 [-0.38, -0.02]	-0.22 [-0.40, -0.03]	-0.02	19.31
Minimal	Other rumor (post-election level)	-0.37 [-0.69, -0.04]	-0.37 [-0.69, -0.06]	-0.01	19.35
Minimal	Own ballot (post-election level)	-0.03 [-0.15, +0.09]	-0.02 [-0.15, +0.11]	+0.01	19.27
Minimal	County ballots (post-election level)	-0.11 [-0.23, +0.02]	-0.08 [-0.21, +0.05]	+0.03	19.31
Minimal	National ballots (post-election level)	+0.06 [-0.08, +0.21]	+0.08 [-0.07, +0.23]	+0.02	19.27
Main specification					
Main	Assigned rumor (post-election delta)	-0.47 [-0.98, +0.05]	-0.39 [-0.89, +0.10]	+0.07	19.31
Main	Other rumor (post-election delta)	-0.68 [-1.58, +0.22]	-0.61 [-1.49, +0.26]	+0.07	19.37
Main	Own ballot (post-election level)	-0.33 [-0.66, -0.00]	-0.25 [-0.62, +0.12]	+0.08	19.27
Main	County ballots (post-election level)	-0.31 [-0.64, +0.03]	-0.21 [-0.59, +0.16]	+0.09	19.31
Main	National ballots (post-election level)	-0.07 [-0.47, +0.33]	-0.02 [-0.42, +0.39]	+0.05	19.27

Notes: MI uses chained equations (MICE) with 30 imputations and 20 iterations, with predictive mean matching for continuous variables and logistic/polytomous models for factors. Entries report the Treatment coefficient with 95% confidence intervals. MI pooled intervals use Rubin's rules. Complete-case estimates use the same weighted model forms as the main specifications.

A.5 Active Placebo Interpretation and Order Diagnostics

Our design compares prebunking treatment to an active placebo, so the reported comparison is the incremental change in treatment relative to placebo rather than a contrast against a null arm. Let $\Delta_{g,w} = E[Y_w - Y_{pre} \mid g]$ denote average change from baseline at wave w for group $g \in \{T, P\}$, where T is treatment and P is placebo. The quantity plotted in Figures 3 and A.1 is therefore the between-arm delta, $\delta_w = \Delta_{T,w} - \Delta_{P,w}$. If placebo content shifts outcomes, that delta remains informative about the active-control comparison but should not be read as a contrast against an inert control.

We use randomized content order as a diagnostic for differential non-inertness. Table A.17 starts with the simplest treatment-arm check, which includes only baseline outcomes, AI modality, and the B-A order indicator. Tables A.18 and A.19 then show the fuller pooled and placebo-only diagnostics. Across those core checks, the "Felon first" coefficients remain small and imprecise. These diagnostics cannot rule out a common placebo effect shared across both placebo components, but they provide little evidence of a strong first-position placebo channel that would spuriously inflate the treatment-minus-placebo contrast. Party-specific and higher-dimensional order interactions appear later in the appendix.

Table A.17: Treatment-only election-confidence models with order control

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.910*** (0.017)		
Pre confidence: County ballots		0.887*** (0.015)	
Pre confidence: National ballots			0.905*** (0.011)
AI type: Chatbot	0.082 (0.097)	0.046 (0.101)	0.076 (0.086)
Topic: Voter Rolls	−0.102 (0.055)	−0.090 (0.051)	−0.016 (0.053)
Order: B-A (felon first)	−0.050 (0.056)	−0.049 (0.052)	−0.042 (0.053)
Constant	0.539** (0.178)	0.784*** (0.161)	0.549*** (0.112)
Observations	2,668	2,668	2,667
Log Likelihood	−4,549.107	−4,467.305	−4,490.755
Akaike Inf. Crit.	9,108.214	8,944.611	8,991.510
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.18: Election-confidence models with order control

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.906*** (0.012)		
Pre confidence: County ballots		0.878*** (0.011)	
Pre confidence: National ballots			0.908*** (0.008)
Treatment (vs. Placebo)	−0.049 (0.126)	−0.065 (0.120)	0.020 (0.110)
AI type: Chatbot	−0.060 (0.090)	−0.083 (0.075)	−0.110 (0.081)
Topic: Voter Rolls	−0.137*** (0.039)	−0.076* (0.037)	−0.071 (0.039)
Order: B-A (felon first)	−0.032 (0.039)	−0.031 (0.037)	−0.007 (0.039)
Treatment (vs. Placebo):AI type: Chatbot	0.140 (0.132)	0.130 (0.126)	0.183 (0.118)
Constant	0.633*** (0.145)	0.910*** (0.115)	0.523*** (0.098)
Observations	5,430	5,431	5,430
Log Likelihood	−9,208.226	−9,172.616	−9,286.005
Akaike Inf. Crit.	18,430.450	18,359.230	18,586.010
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.19: Placebo-only election-confidence models with order control

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.903*** (0.017)		
Pre confidence: County ballots		0.869*** (0.015)	
Pre confidence: National ballots			0.912*** (0.010)
AI type: Chatbot	−0.060 (0.090)	−0.083 (0.075)	−0.111 (0.082)
Topic: Voter Rolls	−0.171** (0.056)	−0.060 (0.054)	−0.126* (0.058)
Order: B-A (felon first)	−0.015 (0.055)	−0.016 (0.054)	0.027 (0.058)
Constant	0.668*** (0.182)	0.962*** (0.144)	0.508*** (0.116)
Observations	2,762	2,763	2,763
Log Likelihood	−4,658.759	−4,704.476	−4,791.326
Akaike Inf. Crit.	9,327.517	9,418.951	9,592.653

Note: *p<0.05; **p<0.01; ***p<0.001

A.6 Outcome Wording and Scale Construction

For readability in the main text, we summarize the core outcome wording and scale construction used in the analysis. The full questionnaire text, routing logic, and chatbot workflow documentation remain in the supplementary materials.

Election-confidence outcomes (0–10) Participants were asked how confident they were that (i) their own ballot, (ii) county ballots, and (iii) ballots throughout the U.S. would be counted accurately, on a 0 (no confidence) to 10 (total confidence) scale.

Rumor-confidence outcomes (0–10) Participants rated agreement with election-rumor statements on a 0 (completely disagree) to 10 (completely agree) scale.

Assigned-rumor and other-rumor outcomes Assigned-rumor outcomes use the rumor item from the participant's randomized topic. Other-rumor outcomes use the analogous rumor item from the non-assigned topic, so they serve as a spillover check rather than the directly targeted rumor. In the regression tables, columns labeled "Post Other Rumor" and "Recontact Other Rumor" refer to this non-assigned-topic measure.

Conspiracy-score construction The conspiracy score is constructed from six 1–5 items. Items are first direction-harmonized (higher = more conspiratorial response) and then summed:

$$\text{Conspiracy Score} = \sum_{j=1}^{6} \text{Conspiracy Item}_j,$$

yielding a theoretical range of 6 to 30.

A.7 Chatbot Materials, Handoff, and Preregistration Notes

This subsection collects the minimum transparency material needed to understand how the chatbot intervention was implemented and how that implementation maps onto the manuscript's estimands.

Script excerpt The archived chatbot materials are structured JSON scripts with numbered steps, question text, user-answer requirements, flag logic, and tool calls that inject the relevant prebunking articles. For example, the treatment script in `data/election2024_v4_fixed.json` follows a fixed sequence of introduction, first-article discussion, second-article discussion, and reflection rather than an unconstrained open-domain conversation:

> **Step 0: Initial Engagement.** "Would you like a refresher about claims of election fraud?"
> **Step 2: First Article Discussion.** "I want to share an article that addresses some of these claims ..." (`ARTICLE1`)

Step 3: Second Article Discussion. "Here's another article that provides additional context on these issues." (`ARTICLE2`)
Step 4: Reflection and Conclusion. "Based on our conversation ... has anything changed in how you think about these claims?"

External handoff and compliance Most Study 1 respondents assigned to chatbot were routed out of the survey instrument to the external chatbot interface and then returned to the survey after the conversation. The key implementation consequence is the chatbot engagement variable summarized in Tables A.11 and A.12. The main-text models use the full assignment-based ITT estimates, while the appendix reports the companion engaged-subsample checks and the engagement diagnostics so readers can compare the full-sample and engaged-subsample estimates directly.

Preregistration notes The OSF preregistration specified rumor confidence, election confidence across three ballot-confidence items, modality heterogeneity, and one-week persistence as core expectations. The submitted paper narrows the main-text presentation in three ways. First, for readability, the main text centers the assigned-rumor and national-election outcomes, while the appendix reports the other ballot-confidence items and ANES outcomes. Second, the article-versus-chatbot persistence comparison is treated as secondary because the design targeted an approximately 90/10 modality split and realized a Wave 1 split of 4,808 chatbot versus 624 article respondents (88.5% versus 11.5%), which leaves limited precision for small modality differences. Third, the follow-up survey occurred after the election, which matched field timing but was not itself preregistered as a post-election wave. These changes affect emphasis rather than the underlying design, and the appendix preserves the fuller set of specifications.

Additional Supplementary Material

The remaining tables and figures collect additional model variants, extended robustness checks, and supplementary visuals for readers who want to inspect the broader set of analyses beyond the sections cited most directly in the main text.

A.8 Additional Model Variants

These tables retain design-consistent alternatives to the ITT models reported earlier in the appendix. They are useful for triangulation, but the paper emphasizes the pooled treatment-versus-placebo estimates reported above.

Reduced-control tables. Tables A.20–A.22 report reduced-control versions of the ITT models. They retain the same combined treatment-versus-placebo contrast while trimming back the broader covariate set.

Table A.20: Ballot-confidence models (reduced controls)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.835*** (0.014)		
Pre confidence: County ballots		0.808*** (0.013)	
Pre confidence: National ballots			0.824*** (0.011)
Treatment (vs. Placebo)	0.067 (0.038)	0.047 (0.036)	0.176*** (0.038)
Topic: Voter Rolls	−0.153*** (0.038)	−0.087* (0.037)	−0.081* (0.038)
Age group: 45-64	−0.016 (0.055)	−0.030 (0.051)	−0.042 (0.055)
Age group: 65+	0.003 (0.055)	−0.008 (0.050)	−0.095 (0.058)
Age group: Under 30	−0.073 (0.060)	−0.013 (0.057)	−0.014 (0.063)
Gender: Male	0.071 (0.038)	0.078* (0.037)	0.092* (0.039)
Race/ethnicity: Hispanic	−0.128 (0.073)	−0.141 (0.079)	0.094 (0.080)
Race/ethnicity: Other	−0.190 (0.100)	−0.108 (0.095)	−0.092 (0.088)
Race/ethnicity: White	−0.112* (0.056)	−0.170** (0.059)	−0.116 (0.065)
Education: HS or less	−0.054 (0.054)	−0.064 (0.052)	−0.022 (0.053)
Education: Postgrad	0.070 (0.055)	0.020 (0.054)	0.026 (0.054)
Education: Some college	−0.007 (0.050)	−0.029 (0.048)	−0.028 (0.052)
Ideology: : Conservative	−0.027 (0.135)	0.231* (0.116)	0.106 (0.136)
Ideology: : Liberal	0.207 (0.133)	0.469*** (0.114)	0.424** (0.135)
Ideology: : Moderate	0.103 (0.133)	0.320** (0.114)	0.303* (0.135)
Region: : Northeast	−0.095 (0.062)	−0.014 (0.057)	0.096 (0.059)
Region: : South	−0.051 (0.053)	0.024 (0.049)	−0.010 (0.054)
Region: : West	−0.027 (0.057)	0.009 (0.051)	0.034 (0.056)
Urbanicity: City	0.138 (0.297)	0.209 (0.303)	−0.111 (0.208)
Urbanicity: Rural area	0.042 (0.299)	0.268 (0.305)	−0.199 (0.208)
Urbanicity: Suburb	0.040 (0.296)	0.225 (0.302)	−0.169 (0.206)
Urbanicity: Town	0.031 (0.298)	0.261 (0.304)	−0.237 (0.210)
Conspiracy score	0.034*** (0.003)	0.032*** (0.003)	0.035*** (0.004)
Constant	0.464 (0.343)	0.328 (0.341)	0.285 (0.260)
Observations	5,430	5,431	5,430
Log Likelihood	−9,059.008	−9,034.176	−9,125.219
Akaike Inf. Crit.	18,168.010	18,118.350	18,300.440
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.21: Ballot-confidence models at recontact (reduced controls)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.443*** (0.021)		
Pre confidence: County ballots		0.436*** (0.021)	
Pre confidence: National ballots			0.427*** (0.020)
Treatment (vs. Placebo)	0.010 (0.061)	−0.065 (0.062)	0.093 (0.070)
Topic: Voter Rolls	0.058 (0.061)	0.082 (0.062)	0.130 (0.070)
Age group: 45-64	0.138 (0.095)	0.142 (0.098)	0.062 (0.108)
Age group: 65+	0.273** (0.097)	0.228* (0.097)	0.030 (0.110)
Age group: Under 30	0.008 (0.107)	0.111 (0.113)	0.260* (0.116)
Gender: Male	0.149* (0.060)	0.205*** (0.061)	0.154* (0.069)
Race/ethnicity: Hispanic	0.468** (0.142)	0.416** (0.154)	0.591*** (0.162)
Race/ethnicity: Other	0.339 (0.178)	0.257 (0.185)	0.281 (0.206)
Race/ethnicity: White	0.578*** (0.113)	0.544*** (0.115)	0.667*** (0.126)
Education: HS or less	−0.223** (0.083)	−0.216* (0.087)	−0.182 (0.098)
Education: Postgrad	0.017 (0.088)	0.096 (0.092)	0.039 (0.104)
Education: Some college	−0.160* (0.081)	−0.162 (0.083)	−0.200* (0.097)
Ideology: : Conservative	0.965*** (0.234)	0.887*** (0.261)	0.890*** (0.232)
Ideology: : Liberal	0.367 (0.245)	0.290 (0.269)	−0.030 (0.247)
Ideology: : Moderate	0.510* (0.238)	0.486 (0.265)	0.409 (0.238)
Region: : Northeast	0.017 (0.102)	−0.043 (0.106)	−0.020 (0.114)
Region: : South	−0.009 (0.081)	−0.054 (0.084)	−0.170 (0.093)
Region: : West	−0.100 (0.088)	−0.188* (0.089)	−0.161 (0.099)
Urbanicity: City	1.462 (1.166)	1.347 (1.183)	1.540 (1.069)
Urbanicity: Rural area	1.494 (1.166)	1.465 (1.183)	1.550 (1.069)
Urbanicity: Suburb	1.500 (1.167)	1.441 (1.183)	1.586 (1.069)
Urbanicity: Town	1.608 (1.166)	1.611 (1.183)	1.651 (1.069)
Conspiracy score	0.002 (0.005)	0.007 (0.006)	−0.009 (0.006)
Constant	2.028 (1.167)	2.071 (1.194)	1.997 (1.074)
Observations	4,384	4,383	4,384
Log Likelihood	−8,938.398	−9,054.218	−9,643.620
Akaike Inf. Crit.	17,926.790	18,158.440	19,337.240
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.22: Non-assigned rumor confidence models (reduced controls)

	Post Other Rumor	Recontact Other Rumor
Pre confidence: Other rumor	0.732*** (0.013)	0.482*** (0.018)
Treatment (vs. Placebo)	−0.481*** (0.088)	−0.436** (0.139)
Topic: Voter Rolls	0.202* (0.088)	0.181 (0.138)
Age group: 45-64	0.138 (0.126)	0.234 (0.194)
Age group: 65+	0.359** (0.137)	0.438* (0.201)
Age group: Under 30	−0.005 (0.149)	−0.195 (0.261)
Gender: Male	−0.382*** (0.089)	−0.276* (0.140)
Race/ethnicity: Hispanic	0.057 (0.230)	−0.400 (0.329)
Race/ethnicity: Other	0.070 (0.233)	−0.057 (0.391)
Race/ethnicity: White	0.295 (0.178)	−0.672* (0.271)
Education: HS or less	0.144 (0.127)	0.344 (0.203)
Education: Postgrad	−0.247 (0.137)	−0.215 (0.202)
Education: Some college	−0.019 (0.116)	0.224 (0.194)
Ideology: : Conservative	0.427 (0.259)	0.710 (0.457)
Ideology: : Liberal	−0.754** (0.268)	0.180 (0.457)
Ideology: : Moderate	−0.615* (0.256)	0.346 (0.452)
Region: : Northeast	0.338* (0.140)	0.367 (0.225)
Region: : South	0.313** (0.113)	0.166 (0.190)
Region: : West	0.219 (0.122)	0.291 (0.207)
Urbanicity: City	−0.036 (0.462)	−0.636 (1.248)
Urbanicity: Rural area	−0.024 (0.464)	−0.723 (1.248)
Urbanicity: Suburb	0.066 (0.458)	−0.681 (1.246)
Urbanicity: Town	0.072 (0.468)	−0.990 (1.251)
Conspiracy score	−0.151*** (0.012)	−0.209*** (0.017)
Constant	5.491*** (0.654)	8.921*** (1.380)
Observations	5,427	4,380
Log Likelihood	−13,839.560	−12,583.950
Akaike Inf. Crit.	27,729.120	25,217.900
Note:		*p<0.05; **p<0.01; ***p<0.001

Minimal-control tables. Tables A.23–A.25 provide the corresponding stripped-down models. Because treatment is randomized, the core comparison of interest is the pooled treated-versus-placebo contrast without relying on a large covariate set for identification. For election-confidence outcomes, the minimal specification still includes the corresponding pre-treatment confidence measure, which is the natural ANCOVA version for a randomized post-treatment outcome. For rumor outcomes, by contrast, the dependent variables are already expressed as post-minus-pre changes, so the minimal model reduces to the treatment indicator plus topic assignment.

Table A.23: Ballot-confidence models (minimal controls)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.906*** (0.012)		
Pre confidence: County ballots		0.878*** (0.011)	
Pre confidence: National ballots			0.908*** (0.008)
Treatment (vs. Placebo)	0.075 (0.039)	0.049 (0.037)	0.182*** (0.039)
Topic: Voter Rolls	−0.138*** (0.039)	−0.076* (0.037)	−0.071 (0.039)
Constant	0.565*** (0.104)	0.822*** (0.094)	0.423*** (0.062)
Observations	5,430	5,431	5,430
Log Likelihood	−9,209.534	−9,173.868	−9,287.543
Akaike Inf. Crit.	18,427.070	18,355.740	18,583.090
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.24: Ballot-confidence models at recontact (minimal controls)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.422*** (0.019)		
Pre confidence: County ballots		0.420*** (0.018)	
Pre confidence: National ballots			0.345*** (0.016)
Treatment (vs. Placebo)	−0.032 (0.063)	−0.107 (0.065)	0.062 (0.072)
Topic: Voter Rolls	0.055 (0.063)	0.079 (0.065)	0.123 (0.072)
Constant	4.921*** (0.175)	4.853*** (0.166)	4.885*** (0.128)
Observations	4,384	4,383	4,384
Log Likelihood	−9,071.178	−9,179.290	−9,762.352
Akaike Inf. Crit.	18,150.360	18,366.580	19,532.700
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.25: Non-assigned rumor confidence models (minimal controls)

	Post Other Rumor	Recontact Other Rumor
Pre confidence: Other rumor	0.893*** (0.006)	0.663*** (0.010)
Treatment (vs. Placebo)	−0.505*** (0.093)	−0.390** (0.143)
Topic: Voter Rolls	0.171 (0.094)	0.193 (0.143)
Constant	1.126*** (0.101)	2.531*** (0.155)
Observations	5,427	4,380
Log Likelihood	−14,113.640	−12,756.410
Akaike Inf. Crit.	28,235.270	25,520.820
Note:		*p<0.05; **p<0.01; ***p<0.001

Minimal interaction benchmarks. These tables retain the stripped-down interaction specifications that keep Treatment × Modality and topic assignment but omit the broader

covariate set. They are useful as secondary heterogeneity checks, but the appendix emphasizes the primary specifications reported earlier.

Table A.26: Election confidence models with treatment × modality interaction (minimal controls)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre-treatment own-ballot confidence	0.906*** (0.012)		
Pre-treatment county-ballot confidence		0.878*** (0.011)	
Pre-treatment national-ballot confidence			0.908*** (0.008)
Treatment	−0.048 (0.126)	−0.064 (0.120)	0.021 (0.110)
Chatbot (vs. article)	−0.060 (0.090)	−0.083 (0.075)	−0.110 (0.081)
Topic: Voter Rolls	−0.138*** (0.039)	−0.076* (0.037)	−0.071 (0.039)
Treatment × Chatbot (vs. article)	0.139 (0.132)	0.129 (0.126)	0.182 (0.118)
Constant	0.617*** (0.144)	0.895*** (0.114)	0.520*** (0.095)
Observations	5,430	5,431	5,430
Log Likelihood	−9,208.671	−9,173.041	−9,286.027
Akaike Inf. Crit.	18,429.340	18,358.080	18,584.060
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.27: Election confidence models at recontact with treatment × modality interaction (minimal controls)

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre-treatment own-ballot confidence	0.422*** (0.019)		
Pre-treatment county-ballot confidence		0.420*** (0.018)	
Pre-treatment national-ballot confidence			0.345*** (0.016)
Treatment	−0.347 (0.186)	−0.313 (0.183)	−0.096 (0.218)
Chatbot (vs. article)	−0.165 (0.123)	−0.167 (0.118)	−0.093 (0.162)
Topic: Voter Rolls	0.055 (0.063)	0.078 (0.065)	0.123 (0.072)
Treatment × Chatbot (vs. article)	0.355 (0.197)	0.232 (0.195)	0.178 (0.231)
Constant	5.068*** (0.200)	5.000*** (0.189)	4.968*** (0.188)
Observations	4,384	4,383	4,384
Log Likelihood	−9,069.083	−9,178.264	−9,761.966
Akaike Inf. Crit.	18,150.170	18,368.530	19,535.930
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.28: Rumor confidence models with treatment × modality interaction (minimal controls)

	Post Assigned Rumor	Recontact Assigned Rumor	Post Other Rumor	Recontact Other Rumor
Treatment	−0.659*** (0.178)	−0.477 (0.268)	−0.747** (0.274)	−0.687 (0.487)
Chatbot (vs. article)	−0.129 (0.135)	−0.014 (0.217)	−0.111 (0.181)	0.211 (0.372)
Topic: Voter Rolls	−0.160** (0.059)	0.033 (0.092)	0.210* (0.097)	0.262 (0.165)
Treatment × Chatbot (vs. article)	0.369 (0.189)	0.314 (0.286)	0.286 (0.293)	0.357 (0.517)
Constant	0.391** (0.135)	−0.299 (0.209)	0.166 (0.173)	−0.941* (0.368)
Observations	5,430	4,383	5,427	4,380
Log Likelihood	−11,602.660	−10,789.510	−14,275.540	−13,377.240
Akaike Inf. Crit.	23,215.320	21,589.030	28,561.080	26,764.480
Note:				*p<0.05; **p<0.01; ***p<0.001

Five-point ideology alternatives. These reruns replace party identification and the legacy three-category ideology covariate with the updated five-point ideology item. In the interacted table, "Moderate" is the omitted category and "Not sure" or missing ideology responses are excluded. We retain these models here as secondary specification checks; the earlier appendix sections keep the ideology material that matters directly for the missing-data discussion.

Table A.29: Assigned-rumor and election confidence models with five-point ideology

	Post-treatment	Post-election	Post-treatment	Post-election
	Assigned-rumor confidence	Assigned-rumor confidence	Election confidence	Election confidence
Pre-treatment assigned-rumor confidence	0.650*** (0.016)	0.454*** (0.020)		
Pre-treatment national-ballot confidence			0.825*** (0.011)	0.435*** (0.020)
Treatment	−0.310*** (0.054)	−0.198* (0.077)	0.166*** (0.039)	0.098 (0.070)
Topic: Voter Rolls	−0.042 (0.054)	0.182* (0.077)	−0.080* (0.038)	0.131 (0.070)
Age 45-64	0.114 (0.076)	0.129 (0.112)	−0.055 (0.056)	0.068 (0.108)
Age 65+	0.293*** (0.082)	0.148 (0.116)	−0.121* (0.059)	0.027 (0.110)
Age Under 30	0.004 (0.091)	−0.013 (0.144)	−0.019 (0.063)	0.262* (0.115)
Male (vs. female)	−0.190*** (0.055)	−0.179* (0.079)	0.082* (0.039)	0.141* (0.071)
Race: Hispanic	0.247 (0.134)	−0.237 (0.173)	0.048 (0.080)	0.448** (0.164)
Race: Other	0.030 (0.144)	0.109 (0.212)	−0.115 (0.089)	0.231 (0.205)
Race: White	0.289** (0.108)	−0.266 (0.139)	−0.126 (0.066)	0.535*** (0.129)
Education: HS or less	0.016 (0.075)	0.144 (0.112)	−0.003 (0.054)	−0.194 (0.103)
Education: Postgrad	−0.223** (0.077)	−0.115 (0.113)	0.040 (0.054)	0.048 (0.105)
Education: Some college	−0.137 (0.072)	0.110 (0.107)	−0.016 (0.052)	−0.215* (0.098)
Ideology: : (5-point): Very liberal	−0.119 (0.091)	−0.150 (0.128)	0.144* (0.056)	−0.480*** (0.137)
Ideology: : (5-point): Liberal	−0.166 (0.086)	−0.100 (0.117)	0.099 (0.057)	−0.417*** (0.115)
Ideology: : (5-point): Conservative	0.475*** (0.081)	0.231* (0.114)	−0.201*** (0.058)	0.490*** (0.092)
Ideology: : (5-point): Very conservative	0.486*** (0.097)	0.245 (0.145)	−0.202** (0.070)	0.594*** (0.114)
Region: : : Northeast	−0.020 (0.084)	0.073 (0.126)	0.109 (0.058)	−0.046 (0.116)
Region: : : South	0.038 (0.072)	−0.041 (0.101)	−0.008 (0.054)	−0.158 (0.094)
Region: : : West	0.029 (0.077)	0.112 (0.111)	0.021 (0.057)	−0.138 (0.101)
Urbanicity: City	0.031 (0.206)	0.482 (0.617)	−0.083 (0.220)	1.848 (1.158)
Urbanicity: Rural area	0.084 (0.210)	0.402 (0.614)	−0.158 (0.220)	1.895 (1.157)
Urbanicity: Suburb	0.154 (0.203)	0.398 (0.613)	−0.130 (0.219)	1.932 (1.157)
Urbanicity: Town	0.133 (0.211)	0.281 (0.617)	−0.198 (0.223)	2.012 (1.157)
Political interest: Hardly at all	0.273 (0.421)	−0.370 (0.400)	−0.428* (0.218)	0.186 (0.389)
Political interest: Most of the time	−0.051 (0.393)	−0.470 (0.354)	−0.227 (0.189)	0.053 (0.349)
Political interest: Only now and then	−0.030 (0.402)	−0.288 (0.376)	−0.225 (0.197)	0.082 (0.365)
Political interest: Some of the time	0.009 (0.393)	−0.471 (0.354)	−0.246 (0.189)	0.178 (0.348)
Conspiracy score	−0.106*** (0.007)	−0.117*** (0.010)	0.035*** (0.004)	−0.008 (0.007)
Constant	3.547*** (0.503)	4.633*** (0.762)	0.816** (0.306)	1.998 (1.208)
Observations	5,254	4,251	5,253	4,251
Log Likelihood	−10,636.400	−9,660.203	−8,771.285	−9,313.404
Akaike Inf. Crit.	21,332.800	19,380.410	17,602.570	18,686.810

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.30: Assigned-rumor and election confidence models with treatment × five-point ideology

	Post-treatment	Post-election	Post-treatment	Post-election
	Assigned-rumor confidence	Assigned-rumor confidence	Election confidence	Election confidence
Pre-treatment assigned-rumor confidence	0.651*** (0.016)	0.454*** (0.020)		
Pre-treatment national-ballot confidence			0.825*** (0.011)	0.434*** (0.020)
Treatment	−0.193 (0.099)	−0.096 (0.145)	0.096 (0.069)	0.065 (0.128)
Ideology: (5-point): Very liberal	−0.077 (0.128)	−0.263 (0.157)	0.140 (0.084)	−0.578** (0.184)
Ideology: (5-point): Liberal	−0.067 (0.121)	−0.057 (0.165)	0.062 (0.085)	−0.445** (0.164)
Ideology: (5-point): Conservative	0.642*** (0.110)	0.344* (0.151)	−0.332*** (0.083)	0.533*** (0.129)
Ideology: (5-point): Very conservative	0.467*** (0.121)	0.431* (0.189)	−0.188* (0.095)	0.516** (0.158)
Topic: Voter Rolls	−0.043 (0.054)	0.185* (0.077)	−0.080* (0.038)	0.129 (0.070)
Age 45-64	0.114 (0.076)	0.130 (0.111)	−0.055 (0.055)	0.070 (0.108)
Age 65+	0.292*** (0.082)	0.153 (0.116)	−0.120* (0.059)	0.029 (0.110)
Age Under 30	0.005 (0.091)	−0.016 (0.144)	−0.019 (0.063)	0.261* (0.116)
Male (vs. female)	−0.190*** (0.055)	−0.178* (0.079)	0.082* (0.039)	0.143* (0.071)
Race: Hispanic	0.247 (0.133)	−0.228 (0.174)	0.046 (0.080)	0.449** (0.164)
Race: Other	0.024 (0.144)	0.099 (0.212)	−0.111 (0.089)	0.226 (0.206)
Race: White	0.285** (0.108)	−0.264 (0.139)	−0.123 (0.066)	0.532*** (0.129)
Education: HS or less	0.012 (0.075)	0.137 (0.112)	−0.001 (0.054)	−0.194 (0.103)
Education: Postgrad	−0.224** (0.077)	−0.128 (0.113)	0.041 (0.054)	0.047 (0.105)
Education: Some college	−0.142* (0.072)	0.098 (0.107)	−0.011 (0.052)	−0.218* (0.098)
Region: : Northeast	−0.025 (0.084)	0.071 (0.126)	0.112 (0.059)	−0.051 (0.116)
Region: : South	0.032 (0.072)	−0.041 (0.101)	−0.003 (0.054)	−0.163 (0.094)
Region: : West	0.026 (0.077)	0.114 (0.111)	0.023 (0.057)	−0.140 (0.101)
Urbanicity: City	0.034 (0.203)	0.486 (0.616)	−0.088 (0.216)	1.884 (1.161)
Urbanicity: Rural area	0.091 (0.208)	0.409 (0.613)	−0.167 (0.216)	1.933 (1.160)
Urbanicity: Suburb	0.156 (0.201)	0.402 (0.612)	−0.136 (0.215)	1.967 (1.160)
Urbanicity: Town	0.144 (0.208)	0.294 (0.615)	−0.208 (0.219)	2.046 (1.161)
Political interest: Hardly at all	0.276 (0.418)	−0.351 (0.398)	−0.432* (0.217)	0.181 (0.390)
Political interest: Most of the time	−0.052 (0.390)	−0.456 (0.353)	−0.231 (0.187)	0.053 (0.350)
Political interest: Only now and then	−0.031 (0.399)	−0.276 (0.374)	−0.230 (0.195)	0.084 (0.366)
Political interest: Some of the time	0.010 (0.391)	−0.454 (0.353)	−0.252 (0.187)	0.178 (0.350)
Conspiracy score	−0.105*** (0.007)	−0.117*** (0.010)	0.035*** (0.004)	−0.009 (0.007)
Treatment × Ideology: (5-point): Very liberal	−0.090 (0.172)	0.213 (0.240)	0.014 (0.106)	0.190 (0.264)
Treatment × Ideology: (5-point): Liberal	−0.201 (0.165)	−0.083 (0.221)	0.075 (0.108)	0.057 (0.217)
Treatment × Ideology: (5-point): Conservative	−0.338* (0.145)	−0.233 (0.208)	0.264* (0.109)	−0.085 (0.180)
Treatment × Ideology: (5-point): Very conservative	0.050 (0.171)	−0.421 (0.260)	−0.039 (0.129)	0.172 (0.212)
Constant	3.489*** (0.501)	4.563*** (0.764)	0.854** (0.305)	1.985 (1.215)
Observations	5,254	4,251	5,253	4,251
Log Likelihood	−10,631.920	−9,656.893	−8,765.993	−9,312.311
Akaike Inf. Crit.	21,331.850	19,381.780	17,599.990	18,692.620

Note: *p<0.05; **p<0.01; ***p<0.001

A.9 Expanded Attrition and Sample-Scope Checks

These tables preserve the fuller Wave 1 attrition diagnostics and expanded-sample reruns referenced earlier in the appendix. They ask whether the main conclusions change when additional non-final cases with observed outcomes are reincorporated and when completion is examined within observed party and ideology groups.

Expanded-sample robustness. This comparison asks whether the main Wave 1 conclusions depend heavily on restricting attention to final completes. To assess that possibility, we re-estimate the immediate post-treatment results in an expanded Wave 1 sample that retains non-final respondents whenever their outcome data are observed. The starting point is the same Wave 1 accounting described above: 8,428 starts, 8,093 randomized respondents, and 335 who never reached treatment. The disposition report records 5,431 final completes. The main Wave 1 analysis sample contains 5,432 respondents because it also includes one respondent with missing post-treatment outcomes, and the expanded Wave 1 sample contains 8,429 respondents because it also retains one respondent with a missing disposition code. Relative to the $n = 5{,}431$ final completes, this expanded sample adds 2,998 non-final respondents, mostly ordinary breakoffs (2,628), plus smaller numbers of speeders (221), response-quality removals (89), failed-attention breakoffs (52), and screenouts (7).

Two questions matter here. One is whether the missing cases are concentrated enough in particular arms or party groups to threaten the randomized comparison. The other is whether the outcome differences change materially once those additional cases are brought back in. Table A.31 addresses the party-specific question as far as the observed data allow. Within the subset where party identification is observed, treatment-minus-placebo completion differences run from -1.4 to $+1.4$ percentage points, and none of the no-controls tests are statistically significant. This remains only a partial check because party identification is missing for most randomized non-final respondents. The broader arm-level evidence comes from Table A.14 and Table A.10: overall Wave 1 completion does not differ significantly by arm ($p = 0.1194$), but the detailed disposition breakdown shows a modest imbalance in the "Respondent did not complete" category ($p = 0.037$). The expanded-sample comparison then shows how much the estimates move once the final-complete restriction is relaxed as far as the observed Wave 1 outcomes allow.

Party	Placebo completed/N	Treatment completed/N	Placebo completion rate	Treatment completion rate	Completion-rate difference	Unadjusted Fisher exact p-value
Democrat	1048/1546	1064/1593	67.8%	66.8%	-1.0 pp	0.5427
Independent	782/1141	748/1117	68.5%	67.0%	-1.6 pp	0.4440
Republican	932/1375	857/1316	67.8%	65.1%	-2.7 pp	0.1528
Overall observed-party subset	2762/4062	2669/4026	68.0%	66.3%	-1.7 pp	0.1023

Table A.31: Wave 1 completion by assigned arm within the subset of randomized respondents with observed party identification. Entries report completed cases over total cases, completion rates, treatment-minus-placebo completion-rate differences, and unadjusted Fisher exact p-values. Party identification is observed for 2,656 of the 2,662 randomized non-final respondents (99.8%), so this is close to a full party-specific attrition check.

Table A.32 compares simple immediate post-treatment means in the main Wave 1 analysis

sample and in the expanded sample. The election-confidence gaps decrease by 0.037 points for own-ballot confidence, 0.032 points for county-ballot confidence, and 0.044 points for national-ballot confidence. The misinformation outcomes change much less. Assigned-rumor confidence shifts from −0.304 to −0.285, assigned-rumor confidence change shifts from −0.308 to −0.304, and other-rumor confidence change is nearly identical at −0.471 and −0.472.

Outcome	Main: Placebo	Main: Treatment	Main: T−C	Expanded: Placebo	Expanded: Treatment	Expanded: T−C	Change in T−C
Own ballot confidence	7.966	8.071	+0.105	7.981	8.050	+0.069	-0.037
County ballots confidence	7.862	7.994	+0.132	7.871	7.971	+0.101	-0.032
National ballots confidence	6.829	7.085	+0.257	6.871	7.084	+0.213	-0.044
Assigned-rumor confidence	4.733	4.429	-0.304	4.769	4.485	-0.285	+0.019
Assigned-rumor confidence change	0.181	-0.127	-0.308	0.172	-0.132	-0.304	+0.004
Other-rumor confidence change	0.166	-0.305	-0.471	0.150	-0.322	-0.472	-0.001

Table A.32: Immediate post-treatment means in the main Wave 1 analysis sample and in an expanded Wave 1 sample that retains non-final respondents whenever immediate post-treatment outcomes are observed. Columns report placebo means, treatment means, treatment-minus-control gaps, and the change in that gap under the expanded sample definition.

The same pattern appears when the two samples are put on the same unweighted reduced-control specification in Table A.33. The chatbot estimate for national-ballots confidence changes from 0.162 (SE 0.036) in the main sample to 0.148 (SE 0.035) in the expanded sample. For assigned-rumor confidence change, the chatbot estimate changes from −0.268 to −0.271, and for other-rumor confidence change it changes from −0.422 to −0.438. The article estimates are also directionally similar across the two samples: the national-ballots estimate changes from 0.000 to 0.012, the assigned-rumor estimate from −0.610 to −0.498, and the other-rumor estimate from −0.642 to −0.547. Their standard errors remain much larger than the corresponding chatbot estimates.

Outcome	Delivery	Main sample	Expanded sample	Coefficient shift
National ballots confidence	AI Chatbot	0.162 (0.036)	0.148 (0.035)	-0.015
National ballots confidence	AI Article	0.000 (0.100)	0.012 (0.090)	+0.011
Assigned-rumor confidence change	AI Chatbot	-0.268 (0.057)	-0.271 (0.056)	-0.004
Assigned-rumor confidence change	AI Article	-0.610 (0.160)	-0.498 (0.142)	+0.113
Other-rumor confidence change	AI Chatbot	-0.422 (0.095)	-0.438 (0.092)	-0.016
Other-rumor confidence change	AI Article	-0.642 (0.264)	-0.547 (0.233)	+0.095

Table A.33: Harmonized unweighted OLS comparisons for the main immediate post-treatment outcomes in the main Wave 1 analysis sample and the expanded Wave 1 sample. Entries are treatment coefficients by delivery mode with standard errors in parentheses. The election-confidence model adds the lagged pre-treatment outcome; the change-score models omit it by construction.

These tables point to a fairly clear conclusion. Wave 1 attrition matters enough to require close inspection, which is why the sample-flow tables, chatbot-compliance diagnostics, and expanded-sample comparison are reported here. The comparison indicates that the core misinformation results do not appear to be driven primarily by the final-complete restriction: those estimates change little across the two samples. The election-confidence estimates are more sensitive. In the expanded sample, the national-ballots treatment-control gap declines from +0.257 to +0.213, and the chatbot estimate in the harmonized regression declines from

0.162 to 0.148. These checks therefore suggest that attrition bears more on election-confidence outcomes than on rumor-confidence outcomes.

A.10 Expanded Order Diagnostics

These tables extend the shorter core placebo-order discussion with party-specific and higher-dimensional variants. Across these follow-up checks, the same broad conclusion remains: order coefficients are generally small relative to their uncertainty.

Table A.34: Election-confidence models with order × party interactions

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.882*** (0.012)		
Pre confidence: County ballots		0.854*** (0.011)	
Pre confidence: National ballots			0.875*** (0.009)
Treatment (vs. Placebo)	−0.079 (0.126)	−0.094 (0.119)	−0.019 (0.111)
AI type: Chatbot	−0.078 (0.091)	−0.104 (0.075)	−0.131 (0.081)
Topic: Voter Rolls	−0.143*** (0.039)	−0.083* (0.037)	−0.076 (0.039)
Order: B-A (felon first)	−0.007 (0.048)	−0.038 (0.049)	0.038 (0.058)
Party ID: Independent	−0.196*** (0.058)	−0.203*** (0.056)	−0.277*** (0.066)
Party ID: Republican	−0.454*** (0.067)	−0.472*** (0.067)	−0.483*** (0.072)
Treatment (vs. Placebo):AI type: Chatbot	0.164 (0.133)	0.155 (0.126)	0.219 (0.118)
Order: B-A (felon first) × Party ID: Independent	−0.057 (0.084)	−0.054 (0.081)	−0.054 (0.093)
Order: B-A (felon first) × Party ID: Republican	−0.022 (0.092)	0.068 (0.089)	−0.085 (0.091)
Constant	1.075*** (0.159)	1.356*** (0.134)	1.036*** (0.117)
Observations	5,430	5,431	5,430
Log Likelihood	−9,150.202	−9,120.693	−9,219.912
Akaike Inf. Crit.	18,322.400	18,263.390	18,461.820

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.35: Placebo-only election-confidence models with order × party interactions

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.879*** (0.017)		
Pre confidence: County ballots		0.843*** (0.016)	
Pre confidence: National ballots			0.876*** (0.012)
AI type: Chatbot	−0.082 (0.091)	−0.107 (0.075)	−0.137 (0.081)
Topic: Voter Rolls	−0.186*** (0.055)	−0.080 (0.054)	−0.139* (0.058)
Order: B-A (felon first)	−0.025 (0.065)	−0.057 (0.071)	0.044 (0.086)
Party ID: Independent	−0.224** (0.080)	−0.216* (0.086)	−0.313** (0.104)
Party ID: Republican	−0.524*** (0.103)	−0.597*** (0.098)	−0.593*** (0.107)
Order: B-A (felon first) × Party ID: Independent	−0.058 (0.111)	−0.065 (0.121)	−0.081 (0.141)
Order: B-A (felon first) × Party ID: Republican	0.093 (0.132)	0.187 (0.124)	0.029 (0.130)
Constant	1.147*** (0.203)	1.485*** (0.176)	1.096*** (0.151)
Observations	2,762	2,763	2,763
Log Likelihood	−4,627.010	−4,669.203	−4,752.820
Akaike Inf. Crit.	9,272.020	9,356.406	9,523.640

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.36: Assigned- and non-assigned rumor confidence models with order × party interactions

	Post Assigned Rumor	Post-election Assigned Rumor	Post Other Rumor	Post-election Other Rumor
Pre confidence: Assigned rumor	0.803*** (0.010)	0.632*** (0.014)		
Pre confidence: Other rumor			0.860*** (0.008)	0.660*** (0.012)
Treatment (vs. Placebo)	−0.539** (0.165)	−0.294 (0.230)	−0.682** (0.263)	−0.673 (0.414)
AI type: Chatbot	−0.110 (0.126)	−0.007 (0.187)	−0.116 (0.179)	0.018 (0.325)
Topic: Voter Rolls	−0.117* (0.056)	0.129 (0.080)	0.179 (0.093)	0.198 (0.144)
Order: B-A (felon first)	−0.031 (0.096)	0.047 (0.128)	−0.127 (0.163)	0.135 (0.238)
Party ID: Independent	0.389*** (0.101)	−0.004 (0.136)	0.498** (0.180)	−0.086 (0.246)
Party ID: Republican	0.869*** (0.119)	0.260 (0.155)	1.185*** (0.185)	0.208 (0.273)
Treatment (vs. Placebo):AI type: Chatbot	0.252 (0.176)	0.113 (0.246)	0.219 (0.281)	0.321 (0.442)
Order: B-A (felon first) × Party ID: Independent	0.032 (0.135)	0.105 (0.195)	−0.032 (0.236)	0.072 (0.352)
Order: B-A (felon first) × Party ID: Republican	−0.048 (0.138)	−0.085 (0.188)	−0.125 (0.220)	−0.172 (0.342)
Constant	0.867*** (0.141)	1.258*** (0.200)	1.054*** (0.206)	2.457*** (0.379)
Observations	5,430	4,383	5,427	4,380
Log Likelihood	−11,293.210	−10,178.850	−14,066.680	−12,754.510
Akaike Inf. Crit.	22,608.420	20,379.710	28,155.360	25,531.030

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.37: Election-confidence models with treatment × order interactions

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.906*** (0.012)		
Pre confidence: County ballots		0.878*** (0.011)	
Pre confidence: National ballots			0.908*** (0.008)
Treatment (vs. Placebo)	−0.095 (0.166)	−0.077 (0.154)	0.029 (0.153)
AI type: Chatbot	−0.154 (0.134)	−0.089 (0.087)	−0.117 (0.122)
Order: B-A (felon first)	−0.177 (0.171)	−0.026 (0.136)	0.014 (0.151)
Topic: Voter Rolls	−0.138*** (0.039)	−0.076* (0.037)	−0.071 (0.039)
Treatment (vs. Placebo):AI type: Chatbot	0.211 (0.176)	0.162 (0.163)	0.212 (0.165)
Treatment (vs. Placebo):Order: B-A (felon first)	0.085 (0.253)	0.025 (0.239)	−0.016 (0.221)
AI type: Chatbot × Order: B-A (felon first)	0.184 (0.180)	0.012 (0.148)	0.014 (0.163)
Treatment (vs. Placebo):AI type: Chatbot × Order: B-A (felon first)	−0.136 (0.266)	−0.065 (0.252)	−0.059 (0.237)
Constant	0.707*** (0.176)	0.906*** (0.116)	0.511*** (0.126)
Observations	5,430	5,431	5,430
Log Likelihood	−9,207.268	−9,172.441	−9,285.472
Akaike Inf. Crit.	18,434.540	18,364.880	18,590.940

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.38: Assigned- and non-assigned rumor confidence models with order control

	Post Assigned Rumor	Post-election Assigned Rumor	Post Other Rumor	Post-election Other Rumor
Pre confidence: Assigned rumor	0.849*** (0.007)	0.644*** (0.011)		
Pre confidence: Other rumor			0.893*** (0.006)	0.664*** (0.010)
Treatment (vs. Placebo)	−0.607*** (0.168)	−0.311 (0.230)	−0.749** (0.264)	−0.678 (0.413)
AI type: Chatbot	−0.142 (0.128)	−0.016 (0.186)	−0.149 (0.178)	0.016 (0.323)
Topic: Voter Rolls	−0.136* (0.056)	0.120 (0.080)	0.174 (0.093)	0.193 (0.143)
Order: B-A (felon first)	−0.039 (0.056)	0.056 (0.079)	−0.187* (0.093)	0.103 (0.143)
Treatment (vs. Placebo):AI type: Chatbot	0.311 (0.179)	0.128 (0.245)	0.275 (0.282)	0.325 (0.440)
Constant	1.125*** (0.137)	1.301*** (0.192)	1.351*** (0.193)	2.463*** (0.347)
Observations	5,430	4,383	5,427	4,380
Log Likelihood	−11,363.970	−10,182.080	−14,110.650	−12,755.380
Akaike Inf. Crit.	22,741.940	20,378.170	28,235.290	25,524.760

Note: *p<0.05; **p<0.01; ***p<0.001

Table A.39: Election-confidence models with treatment × order × party interactions

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.882*** (0.012)		
Pre confidence: County ballots		0.854*** (0.011)	
Pre confidence: National ballots			0.875*** (0.009)
Treatment (vs. Placebo)	0.032 (0.061)	−0.009 (0.069)	0.115 (0.084)
Order: B-A (felon first)	−0.026 (0.065)	−0.057 (0.071)	0.045 (0.086)
Party ID: Independent	−0.217** (0.078)	−0.201* (0.085)	−0.311** (0.103)
Party ID: Republican	−0.512*** (0.103)	−0.575*** (0.095)	−0.586*** (0.103)
AI type: Chatbot	−0.001 (0.066)	−0.031 (0.062)	−0.029 (0.059)
Topic: Voter Rolls	−0.145*** (0.039)	−0.085* (0.037)	−0.079* (0.039)
Treatment (vs. Placebo):Order: B-A (felon first)	0.036 (0.095)	0.037 (0.099)	−0.012 (0.116)
Treatment (vs. Placebo):Party ID: Independent	0.043 (0.114)	−0.004 (0.110)	0.068 (0.130)
Treatment (vs. Placebo):Party ID: Republican	0.119 (0.130)	0.210 (0.125)	0.210 (0.134)
Order: B-A (felon first) × Party ID: Independent	−0.054 (0.111)	−0.062 (0.121)	−0.077 (0.141)
Order: B-A (felon first) × Party ID: Republican	0.088 (0.132)	0.185 (0.124)	0.022 (0.131)
Treatment (vs. Placebo):Order: B-A (felon first) × Party ID: Independent	−0.0002 (0.167)	0.021 (0.161)	0.051 (0.184)
Treatment (vs. Placebo):Order: B-A (felon first) × Party ID: Republican	−0.225 (0.184)	−0.237 (0.179)	−0.219 (0.183)
Constant	1.023*** (0.145)	1.318*** (0.133)	0.976*** (0.116)
Observations	5,430	5,431	5,430
Log Likelihood	−9,149.813	−9,119.335	−9,219.124
Akaike Inf. Crit.	18,329.630	18,268.670	18,468.250

Note: *p<0.05; **p<0.01; ***p<0.001

A.11 Attention Check Sensitivity Analysis

These tables regroup the attention-check reruns here because they are post hoc sensitivity analyses rather than preferred specifications. We did not preregister exclusions based on these checks, and the sample shrinks substantially once they are imposed. Passing Attention Check 1 retains about 65% of the main Wave 1 sample, passing Attention Check 2 retains about 53%, and requiring respondents to pass both checks retains about 37%. We therefore report these reruns as a secondary robustness check, but continue to emphasize the full-sample main results in the paper.

Table A.40: Assigned-rumor and election confidence models: attention check 1 subsample

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.832*** (0.018)		
Pre confidence: County ballots		0.814*** (0.016)	
Pre confidence: National ballots			0.809*** (0.013)
Treatment (vs. Placebo)	0.044 (0.048)	0.010 (0.045)	0.106* (0.048)
Topic: Voter Rolls	−0.185*** (0.048)	−0.134** (0.045)	−0.082 (0.047)
Age group: 45-64	−0.030 (0.068)	−0.051 (0.061)	−0.111 (0.066)
Age group: 65+	−0.037 (0.072)	−0.033 (0.063)	−0.090 (0.073)
Age group: Under 30	−0.088 (0.073)	−0.036 (0.067)	−0.058 (0.072)
Gender: Male	0.083 (0.048)	0.080 (0.046)	0.110* (0.048)
Race/ethnicity: Hispanic	−0.206* (0.088)	−0.262** (0.099)	−0.016 (0.103)
Race/ethnicity: Other	−0.172 (0.118)	−0.090 (0.115)	−0.129 (0.109)
Race/ethnicity: White	−0.128 (0.070)	−0.185* (0.073)	−0.174* (0.083)
Education: HS or less	−0.077 (0.068)	−0.055 (0.062)	−0.021 (0.064)
Education: Postgrad	0.026 (0.071)	−0.015 (0.064)	0.054 (0.065)
Education: Some college	−0.023 (0.064)	−0.020 (0.058)	−0.010 (0.064)
Ideology: : Conservative	0.041 (0.160)	0.276* (0.129)	0.227 (0.163)
Ideology: : Liberal	0.250 (0.160)	0.528*** (0.130)	0.507** (0.164)
Ideology: : Moderate	0.149 (0.158)	0.388** (0.128)	0.376* (0.161)
Region: : Northeast	−0.143 (0.077)	−0.003 (0.073)	0.101 (0.074)
Region: : South	−0.064 (0.068)	−0.008 (0.060)	0.006 (0.067)
Region: : West	−0.053 (0.073)	−0.020 (0.064)	0.032 (0.070)
Urbanicity: City	0.004 (0.449)	−0.193 (0.405)	−0.098 (0.221)
Urbanicity: Rural area	−0.054 (0.451)	−0.179 (0.407)	−0.194 (0.221)
Urbanicity: Suburb	−0.097 (0.447)	−0.185 (0.403)	−0.163 (0.218)
Urbanicity: Town	−0.151 (0.450)	−0.191 (0.405)	−0.292 (0.224)
Conspiracy score	0.039*** (0.004)	0.032*** (0.004)	0.042*** (0.005)
Constant	0.533 (0.495)	0.711 (0.439)	0.218 (0.283)
Observations	3,532	3,533	3,532
Log Likelihood	−5,919.904	−5,853.811	−5,950.869
Akaike Inf. Crit.	11,889.810	11,757.620	11,951.740
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.41: Assigned-rumor and election confidence models: attention check 2 subsample

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.849*** (0.019)		
Pre confidence: County ballots		0.814*** (0.016)	
Pre confidence: National ballots			0.834*** (0.015)
Treatment (vs. Placebo)	−0.034 (0.051)	−0.030 (0.050)	0.145** (0.052)
Topic: Voter Rolls	−0.151** (0.053)	−0.127* (0.051)	−0.062 (0.053)
Age group: 45-64	−0.024 (0.079)	−0.063 (0.069)	−0.123 (0.072)
Age group: 65+	−0.026 (0.079)	−0.077 (0.070)	−0.151 (0.078)
Age group: Under 30	−0.066 (0.084)	−0.036 (0.075)	−0.031 (0.079)
Gender: Male	0.068 (0.051)	0.149** (0.051)	0.060 (0.052)
Race/ethnicity: Hispanic	−0.082 (0.107)	−0.062 (0.115)	0.152 (0.117)
Race/ethnicity: Other	−0.049 (0.151)	0.104 (0.145)	−0.081 (0.136)
Race/ethnicity: White	−0.019 (0.088)	−0.068 (0.097)	−0.094 (0.100)
Education: HS or less	−0.025 (0.071)	0.057 (0.070)	0.102 (0.070)
Education: Postgrad	−0.012 (0.083)	0.017 (0.081)	−0.006 (0.071)
Education: Some college	−0.029 (0.070)	−0.008 (0.069)	0.018 (0.072)
Ideology: : Conservative	−0.127 (0.195)	0.203 (0.161)	0.094 (0.204)
Ideology: : Liberal	0.040 (0.197)	0.438** (0.166)	0.387 (0.205)
Ideology: : Moderate	−0.060 (0.194)	0.222 (0.162)	0.189 (0.204)
Region: : Northeast	−0.138 (0.076)	−0.015 (0.077)	0.100 (0.076)
Region: : South	−0.076 (0.073)	0.070 (0.068)	−0.024 (0.078)
Region: : West	−0.089 (0.084)	−0.052 (0.074)	0.002 (0.080)
Urbanicity: City	0.355 (0.449)	0.253 (0.458)	−0.295 (0.301)
Urbanicity: Rural area	0.231 (0.451)	0.323 (0.459)	−0.344 (0.300)
Urbanicity: Suburb	0.270 (0.448)	0.318 (0.457)	−0.329 (0.300)
Urbanicity: Town	0.311 (0.449)	0.374 (0.458)	−0.413 (0.302)
Conspiracy score	0.031*** (0.004)	0.034*** (0.004)	0.033*** (0.005)
Constant	0.350 (0.507)	0.133 (0.510)	0.481 (0.373)
Observations	2,855	2,855	2,855
Log Likelihood	−4,696.639	−4,743.025	−4,735.035
Akaike Inf. Crit.	9,443.277	9,536.049	9,520.069
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.42: Assigned-rumor and election confidence models: both attention checks subsample

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.847*** (0.023)		
Pre confidence: County ballots		0.821*** (0.019)	
Pre confidence: National ballots			0.823*** (0.018)
Treatment (vs. Placebo)	−0.065 (0.062)	−0.068 (0.057)	0.046 (0.061)
Topic: Voter Rolls	−0.178** (0.063)	−0.189** (0.059)	−0.087 (0.062)
Age group: 45-64	0.025 (0.094)	−0.063 (0.079)	−0.105 (0.086)
Age group: 65+	−0.022 (0.097)	−0.089 (0.082)	−0.093 (0.098)
Age group: Under 30	−0.067 (0.097)	−0.069 (0.083)	−0.037 (0.089)
Gender: Male	0.096 (0.060)	0.125* (0.059)	0.083 (0.062)
Race/ethnicity: Hispanic	−0.030 (0.121)	−0.074 (0.139)	0.131 (0.142)
Race/ethnicity: Other	0.063 (0.173)	0.213 (0.169)	0.019 (0.153)
Race/ethnicity: White	0.021 (0.103)	−0.027 (0.116)	−0.069 (0.120)
Education: HS or less	0.017 (0.084)	0.044 (0.077)	0.133 (0.082)
Education: Postgrad	−0.058 (0.103)	−0.042 (0.090)	0.038 (0.085)
Education: Some college	−0.064 (0.084)	−0.032 (0.079)	0.054 (0.083)
Ideology: : Conservative	−0.070 (0.241)	0.152 (0.177)	0.144 (0.251)
Ideology: : Liberal	0.065 (0.245)	0.402* (0.184)	0.408 (0.255)
Ideology: : Moderate	−0.017 (0.241)	0.214 (0.177)	0.214 (0.251)
Region: : Northeast	−0.157 (0.094)	0.061 (0.092)	0.140 (0.090)
Region: : South	−0.025 (0.089)	0.101 (0.078)	0.027 (0.092)
Region: : West	−0.106 (0.103)	−0.066 (0.087)	−0.013 (0.093)
Urbanicity: City	0.265 (0.820)	−0.282 (0.745)	−0.396 (0.320)
Urbanicity: Rural area	0.136 (0.822)	−0.386 (0.746)	−0.567 (0.319)
Urbanicity: Suburb	0.189 (0.819)	−0.244 (0.743)	−0.446 (0.317)
Urbanicity: Town	0.140 (0.820)	−0.245 (0.744)	−0.648* (0.324)
Conspiracy score	0.034*** (0.005)	0.031*** (0.005)	0.036*** (0.006)
Constant	0.320 (0.877)	0.793 (0.789)	0.538 (0.415)
Observations	2,017	2,017	2,017
Log Likelihood	−3,280.274	−3,280.699	−3,342.117
Akaike Inf. Crit.	6,610.547	6,611.398	6,734.234
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.43: Non-assigned rumor confidence models: both attention checks subsample

	Post Assigned Rumor	Recontact Assigned Rumor	Post Other Rumor	Recontact Other Rumor
Pre confidence: Assigned rumor	0.687*** (0.025)	0.450*** (0.032)		
Pre confidence: Other rumor			0.751*** (0.020)	0.470*** (0.030)
Treatment (vs. Placebo)	−0.316*** (0.083)	−0.332** (0.119)	−0.468*** (0.138)	−0.565** (0.216)
Topic: Voter Rolls	−0.012 (0.084)	0.150 (0.116)	0.207 (0.139)	0.206 (0.212)
Age group: 45-64	0.134 (0.115)	0.293 (0.171)	0.179 (0.206)	1.003** (0.333)
Age group: 65+	0.448*** (0.131)	0.366* (0.186)	0.597** (0.224)	1.205*** (0.343)
Age group: Under 30	−0.146 (0.130)	−0.088 (0.206)	−0.023 (0.223)	−0.191 (0.384)
Gender: Male	−0.184* (0.083)	−0.402*** (0.119)	−0.258 (0.137)	−0.619** (0.220)
Race/ethnicity: Hispanic	0.254 (0.229)	−0.559* (0.269)	0.379 (0.403)	−0.488 (0.544)
Race/ethnicity: Other	0.181 (0.252)	0.129 (0.341)	0.372 (0.399)	0.176 (0.616)
Race/ethnicity: White	0.374* (0.180)	−0.415 (0.222)	0.670* (0.306)	−0.730 (0.460)
Education: HS or less	0.067 (0.115)	0.311 (0.171)	0.100 (0.193)	0.482 (0.330)
Education: Postgrad	−0.267* (0.124)	−0.220 (0.172)	−0.533* (0.223)	−0.430 (0.319)
Education: Some college	−0.140 (0.112)	0.096 (0.156)	0.022 (0.185)	0.150 (0.289)
Ideology: : Conservative	−0.119 (0.240)	−0.088 (0.442)	−0.253 (0.418)	−1.172 (0.777)
Ideology: : Liberal	−0.457 (0.254)	−0.312 (0.444)	−1.005* (0.437)	−1.485 (0.787)
Ideology: : Moderate	−0.341 (0.247)	−0.231 (0.437)	−0.854* (0.429)	−1.409 (0.772)
Region: : Northeast	0.086 (0.125)	0.128 (0.173)	0.504* (0.214)	0.651 (0.342)
Region: : South	0.108 (0.110)	0.034 (0.164)	0.236 (0.191)	0.307 (0.326)
Region: : West	0.089 (0.116)	0.224 (0.171)	0.136 (0.191)	0.254 (0.328)
Urbanicity: City	0.121 (0.418)	0.164 (0.957)	0.422 (0.302)	−0.560 (1.213)
Urbanicity: Rural area	0.242 (0.425)	−0.005 (0.957)	0.738* (0.338)	−0.683 (1.205)
Urbanicity: Suburb	0.180 (0.414)	0.040 (0.954)	0.476 (0.294)	−0.542 (1.202)
Urbanicity: Town	0.281 (0.421)	−0.042 (0.959)	0.419 (0.334)	−1.180 (1.216)
Conspiracy score	−0.096*** (0.012)	−0.130*** (0.015)	−0.142*** (0.019)	−0.227*** (0.026)
Constant	3.370*** (0.610)	5.186*** (1.143)	4.680*** (0.785)	10.783*** (1.658)
Observations	2,017	1,629	2,017	1,629
Log Likelihood	−4,000.895	−3,632.593	−5,020.928	−4,607.823
Akaike Inf. Crit.	8,051.791	7,315.186	10,091.860	9,265.645

Note: *p<0.05; **p<0.01; ***p<0.001

A.12 Topic-Specific Specifications

Participants were randomized to either voter-rolls rumors or non-citizen-voting rumors. The topic-specific estimates are generally imprecise and do not materially qualify the main interpretation, but we retain them here for completeness.

Table A.44: Election-confidence models by rumor topic

	Voter Rolls	Non-Citizen Voting
Pre confidence: National ballots	0.908*** (0.011)	0.909*** (0.010)
Treatment (vs. Placebo)	0.235*** (0.057)	0.127* (0.055)
Constant	0.325*** (0.094)	0.443*** (0.077)
Observations	2,731	2,699
Log Likelihood	−4,745.003	−4,537.805
Akaike Inf. Crit.	9,496.005	9,081.610
Note:		*p<0.05; **p<0.01; ***p<0.001

Table A.45: Election-confidence models: voter-rolls topic

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.896*** (0.020)		
Pre confidence: County ballots		0.860*** (0.017)	
Pre confidence: National ballots			0.908*** (0.011)
Treatment (vs. Placebo)	0.108 (0.061)	0.034 (0.056)	0.235*** (0.057)
Constant	0.487** (0.180)	0.899*** (0.152)	0.325*** (0.094)
Observations	2,732	2,732	2,731
Log Likelihood	−4,914.348	−4,809.789	−4,745.003
Akaike Inf. Crit.	9,834.696	9,625.578	9,496.005
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.46: Election-confidence models: non-citizen-voting topic

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.915*** (0.013)		
Pre confidence: County ballots		0.895*** (0.013)	
Pre confidence: National ballots			0.909*** (0.010)
Treatment (vs. Placebo)	0.039 (0.049)	0.062 (0.049)	0.127* (0.055)
Constant	0.505*** (0.116)	0.676*** (0.114)	0.443*** (0.077)
Observations	2,698	2,699	2,699
Log Likelihood	−4,219.993	−4,328.227	−4,537.805
Akaike Inf. Crit.	8,445.986	8,662.455	9,081.610
Note:			*p<0.05; **p<0.01; ***p<0.001

A.13 Topic-Specific Party Interactions

These tables extend the topic-specific specifications with treatment-by-party interactions. The interaction estimates are not stable or precise enough to support a strong heterogeneity claim, so we preserve them as secondary detail rather than core support.

Table A.47: Election-confidence models with treatment × party interactions, by rumor topic

	Voter Rolls	Non-Citizen Voting
Pre confidence: National ballots	0.872*** (0.013)	0.880*** (0.012)
Treatment (vs. Placebo)	0.158* (0.074)	0.060 (0.091)
Party ID: Independent	−0.363*** (0.101)	−0.335** (0.104)
Party ID: Republican	−0.642*** (0.106)	−0.512*** (0.098)
AI type: Chatbot	−0.032 (0.077)	−0.026 (0.091)
Treatment (vs. Placebo):Party ID: Independent	0.089 (0.128)	0.093 (0.135)
Treatment (vs. Placebo):Party ID: Republican	0.129 (0.132)	0.076 (0.127)
Constant	0.948*** (0.141)	0.965*** (0.140)
Observations	2,731	2,699
Log Likelihood	−4,707.010	−4,509.426
Akaike Inf. Crit.	9,430.020	9,034.853
Note:		*p<0.05; **p<0.01; ***p<0.001

Table A.48: Election-confidence models with treatment × party interactions: voter-rolls topic

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.866*** (0.021)		
Pre confidence: County ballots		0.830*** (0.018)	
Pre confidence: National ballots			0.872*** (0.013)
Treatment (vs. Placebo)	0.072 (0.066)	0.020 (0.071)	0.158* (0.074)
Party ID: Independent	−0.338*** (0.088)	−0.243** (0.086)	−0.363*** (0.101)
Party ID: Republican	−0.588*** (0.110)	−0.561*** (0.100)	−0.642*** (0.106)
AI type: Chatbot	−0.041 (0.106)	−0.007 (0.092)	−0.032 (0.077)
Treatment (vs. Placebo):Party ID: Independent	0.048 (0.125)	−0.077 (0.116)	0.089 (0.128)
Treatment (vs. Placebo):Party ID: Republican	0.046 (0.145)	0.113 (0.138)	0.129 (0.132)
Constant	1.086*** (0.245)	1.408*** (0.191)	0.948*** (0.141)
Observations	2,732	2,732	2,731
Log Likelihood	−4,879.401	−4,779.261	−4,707.010
Akaike Inf. Crit.	9,774.801	9,574.521	9,430.020
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.49: Election-confidence models with treatment × party interactions: non-citizen-voting topic

	Own Ballot Confidence	County Ballots Confidence	National Ballots Confidence
Pre confidence: Own ballot	0.898*** (0.014)		
Pre confidence: County ballots		0.876*** (0.014)	
Pre confidence: National ballots			0.880*** (0.012)
Treatment (vs. Placebo)	0.033 (0.068)	0.002 (0.069)	0.060 (0.091)
Party ID: Independent	−0.140 (0.072)	−0.210* (0.086)	−0.335** (0.104)
Party ID: Republican	−0.346*** (0.081)	−0.404*** (0.088)	−0.512*** (0.098)
AI type: Chatbot	0.044 (0.077)	−0.053 (0.083)	−0.026 (0.091)
Treatment (vs. Placebo):Party ID: Independent	0.027 (0.112)	0.077 (0.113)	0.093 (0.135)
Treatment (vs. Placebo):Party ID: Republican	−0.036 (0.113)	0.071 (0.114)	0.076 (0.127)
Constant	0.783*** (0.151)	1.095*** (0.160)	0.965*** (0.140)
Observations	2,698	2,699	2,699
Log Likelihood	−4,195.682	−4,306.200	−4,509.426
Akaike Inf. Crit.	8,407.363	8,628.400	9,034.853
Note:			*p<0.05; **p<0.01; ***p<0.001

Table A.50: Assigned-rumor confidence models with treatment × party interactions, by rumor topic

	Voter Rolls	Non-Citizen Voting
Pre confidence: Assigned rumor	0.827*** (0.013)	0.776*** (0.015)
Treatment (vs. Placebo)	−0.278* (0.124)	−0.280 (0.148)
Party ID: Independent	0.125 (0.133)	0.728*** (0.156)
Party ID: Republican	0.709*** (0.150)	1.067*** (0.175)
AI type: Chatbot	−0.023 (0.105)	0.029 (0.143)
Treatment (vs. Placebo):Party ID: Independent	0.312 (0.178)	−0.403* (0.201)
Treatment (vs. Placebo):Party ID: Republican	−0.095 (0.181)	−0.061 (0.207)
Constant	0.623*** (0.128)	0.724*** (0.169)
Observations	2,731	2,699
Log Likelihood	−5,517.428	−5,747.668
Akaike Inf. Crit.	11,050.850	11,511.330
Note:		*p<0.05; **p<0.01; ***p<0.001

A.14 ANES Representation Outcomes

These tables retain the pooled, topic-specific, and interacted ANES models. They do not alter the basic conclusion that ANES-style efficacy outcomes are smaller and less precisely estimated than the main rumor-confidence effects.

Table A.51: ANES representation models (full sample)

	Officials Don't Care	No Say in Gov't
Treatment (vs. Placebo)	0.201 (0.135)	0.049 (0.140)
Article (vs. Chatbot)	0.322 (0.275)	0.249 (0.299)
Party Identification: Independent	-1.129^{***} (0.143)	-1.030^{***} (0.167)
Party Identification: Republican	-0.582^{***} (0.158)	-0.468^{**} (0.181)
Topic: Voter Rolls	0.036 (0.070)	−0.056 (0.080)
Age Group: 30-44	-0.313^{*} (0.127)	-0.444^{**} (0.139)
Age Group: 45-64	-0.436^{***} (0.117)	-0.584^{***} (0.129)
Age Group: 65+	-0.261^{*} (0.123)	-0.511^{***} (0.136)
Gender: Female	-0.140^{*} (0.070)	-0.200^{*} (0.080)
Race Ethnicity: Black	0.362^{**} (0.124)	0.624^{***} (0.140)
Race Ethnicity: Hispanic	0.089 (0.112)	0.266^{*} (0.134)
Race Ethnicity: Other	−0.124 (0.150)	0.014 (0.181)
Education Level: Some college	−0.105 (0.089)	−0.038 (0.100)
Education Level: College grad	−0.007 (0.097)	0.139 (0.115)
Education Level: Postgrad	0.449^{***} (0.125)	0.471^{***} (0.133)
Ideology: : : Moderate	0.327^{**} (0.101)	0.244^{*} (0.115)
Ideology: : : Conservative	0.079 (0.128)	0.104 (0.148)
Region: : : Midwest	−0.028 (0.104)	−0.055 (0.122)
Region: : : South	0.067 (0.099)	0.107 (0.115)
Region: : : West	0.026 (0.111)	0.070 (0.127)
Urban Rural: Suburb	-0.247^{**} (0.092)	−0.133 (0.102)
Urban Rural: Town	-0.396^{***} (0.114)	-0.404^{**} (0.128)
Urban Rural: Rural area	−0.176 (0.109)	−0.113 (0.126)
Political interest: Some of the time	-0.169^{*} (0.083)	-0.352^{***} (0.095)
Political interest: Only now and then	−0.129 (0.126)	-0.434^{**} (0.141)
Political interest: Hardly at all	-0.459^{**} (0.172)	-0.876^{***} (0.183)
Conspiracy Score	0.112^{***} (0.006)	0.122^{***} (0.007)
Treatment (vs. Placebo) × Article (vs. Chatbot)	-0.921^{*} (0.365)	−0.715 (0.377)
Treatment (vs. Placebo) × Party Identification: Independent	−0.201 (0.194)	−0.010 (0.220)
Treatment (vs. Placebo) × Party Identification: Republican	−0.138 (0.180)	0.010 (0.201)
Article (vs. Chatbot) × Party Identification: Independent	−0.390 (0.360)	−0.412 (0.400)
Article (vs. Chatbot) × Party Identification: Republican	−0.275 (0.334)	−0.508 (0.368)
Treatment (vs. Placebo) × Article (vs. Chatbot) × Party Identification: Independent	0.755 (0.506)	0.827 (0.554)
Treatment (vs. Placebo) × Article (vs. Chatbot) × Party Identification: Republican	0.646 (0.462)	0.680 (0.497)
Constant	-2.613^{***} (0.221)	-2.153^{***} (0.260)
Observations	3,794	3,789
Log Likelihood	−8,061.769	−8,522.802
Akaike Inf. Crit.	16,193.540	17,115.600
Note:		*p<0.05; **p<0.01; ***p<0.001

Table A.52: ANES representation models by rumor topic

	Voter Rolls	Voter Rolls	Non-Citizen Voting	Non-Citizen Voting
	Officials Don't Care	No Say in Gov't	Officials Don't Care	No Say in Gov't
Treatment (vs. Placebo)	0.059 (0.105)	−0.056 (0.118)	0.128 (0.107)	0.174 (0.127)
Article (vs. Chatbot)	0.011 (0.189)	−0.302 (0.219)	0.160 (0.173)	0.124 (0.195)
Age Group: 30-44	−0.470** (0.168)	−0.690*** (0.183)	−0.121 (0.192)	−0.187 (0.212)
Age Group: 45-64	−0.540*** (0.160)	−0.675*** (0.175)	−0.325 (0.168)	−0.494** (0.189)
Age Group: 65+	−0.308 (0.171)	−0.664*** (0.190)	−0.193 (0.175)	−0.347 (0.197)
Gender: Female	−0.170 (0.099)	−0.283** (0.109)	−0.112 (0.099)	−0.129 (0.115)
Race Ethnicity: Black	0.303 (0.164)	0.441* (0.174)	0.450* (0.189)	0.846*** (0.224)
Race Ethnicity: Hispanic	0.079 (0.157)	0.262 (0.191)	0.087 (0.158)	0.265 (0.188)
Race Ethnicity: Other	−0.316 (0.211)	−0.073 (0.241)	0.108 (0.202)	0.113 (0.263)
Education Level: Some college	−0.072 (0.123)	−0.00001 (0.137)	−0.143 (0.127)	−0.092 (0.145)
Education Level: College grad	−0.098 (0.132)	−0.009 (0.154)	0.080 (0.142)	0.298 (0.170)
Education Level: Postgrad	0.446* (0.174)	0.362 (0.185)	0.464** (0.175)	0.609** (0.190)
Party Identification: Independent	−1.118*** (0.145)	−1.003*** (0.165)	−1.370*** (0.140)	−1.070*** (0.169)
Party Identification: Republican	−0.529** (0.175)	−0.456* (0.201)	−0.783*** (0.162)	−0.562** (0.199)
Ideology: : : Moderate	0.359* (0.146)	0.364* (0.165)	0.300* (0.135)	0.128 (0.156)
Ideology: : : Conservative	−0.013 (0.183)	0.165 (0.212)	0.200 (0.174)	0.076 (0.203)
Region: : : Midwest	−0.091 (0.143)	−0.098 (0.161)	0.046 (0.154)	0.004 (0.185)
Region: : : South	0.125 (0.136)	0.152 (0.153)	−0.020 (0.147)	0.060 (0.172)
Region: : : West	−0.104 (0.160)	−0.075 (0.178)	0.144 (0.155)	0.210 (0.180)
Urban Rural: Suburb	−0.381** (0.128)	−0.265 (0.140)	−0.085 (0.132)	0.034 (0.147)
Urban Rural: Town	−0.417* (0.163)	−0.506** (0.185)	−0.404* (0.160)	−0.330 (0.182)
Urban Rural: Rural area	−0.327* (0.155)	−0.291 (0.169)	−0.025 (0.153)	0.085 (0.187)
Political interest: Some of the time	−0.131 (0.117)	−0.428** (0.131)	−0.194 (0.118)	−0.273 (0.140)
Political interest: Only now and then	−0.081 (0.160)	−0.508** (0.182)	−0.177 (0.196)	−0.360 (0.217)
Political interest: Hardly at all	−0.535* (0.230)	−0.977*** (0.249)	−0.331 (0.252)	−0.753** (0.268)
Conspiracy Score	0.119*** (0.008)	0.130*** (0.009)	0.106*** (0.008)	0.111*** (0.010)
Treatment (vs. Placebo) × Article (vs. Chatbot)	−0.364 (0.262)	0.126 (0.300)	−0.483 (0.270)	−0.479 (0.294)
Constant	−2.471*** (0.300)	−2.006*** (0.337)	−2.636*** (0.296)	−2.345*** (0.358)
Observations	1,944	1,939	1,850	1,850
Log Likelihood	−4,135.052	−4,357.996	−3,916.488	−4,153.421
Akaike Inf. Crit.	8,326.104	8,771.993	7,888.976	8,362.841
Note:				*p<0.05; **p<0.01; ***p<0.001

Table A.53: ANES representation models with treatment × topic interactions

	Voter Rolls	Voter Rolls	Non-Citizen Voting	Non-Citizen Voting
	Officials Don't Care	No Say in Gov't	Officials Don't Care	No Say in Gov't
Treatment (vs. Placebo)	0.141 (0.184)	−0.041 (0.185)	0.272 (0.197)	0.169 (0.209)
Article (vs. Chatbot)	0.206 (0.379)	0.045 (0.434)	0.522 (0.370)	0.578 (0.364)
Party Identification: Independent	−1.098*** (0.200)	−1.022*** (0.221)	−1.212*** (0.202)	−1.047*** (0.248)
Party Identification: Republican	−0.480* (0.227)	−0.413 (0.254)	−0.691** (0.216)	−0.523* (0.257)
Age Group: 30-44	−0.474** (0.167)	−0.695*** (0.182)	−0.120 (0.192)	−0.183 (0.213)
Age Group: 45-64	−0.546*** (0.159)	−0.684*** (0.175)	−0.316 (0.168)	−0.480* (0.190)
Age Group: 65+	−0.319 (0.171)	−0.677*** (0.190)	−0.197 (0.175)	−0.344 (0.197)
Gender: Female	−0.170 (0.099)	−0.281* (0.109)	−0.107 (0.099)	−0.123 (0.115)
Race Ethnicity: Black	0.293 (0.164)	0.430* (0.173)	0.455* (0.189)	0.861*** (0.224)
Race Ethnicity: Hispanic	0.070 (0.157)	0.245 (0.192)	0.087 (0.158)	0.267 (0.188)
Race Ethnicity: Other	−0.331 (0.211)	−0.085 (0.243)	0.110 (0.200)	0.125 (0.263)
Education Level: Some college	−0.071 (0.122)	0.003 (0.136)	−0.142 (0.127)	−0.089 (0.146)
Education Level: College grad	−0.099 (0.132)	−0.016 (0.154)	0.079 (0.142)	0.294 (0.170)
Education Level: Postgrad	0.448* (0.174)	0.370* (0.186)	0.473** (0.174)	0.619** (0.190)
Ideology: : : Moderate	0.372* (0.147)	0.378* (0.166)	0.298* (0.136)	0.120 (0.157)
Ideology: : : Conservative	−0.013 (0.183)	0.167 (0.212)	0.200 (0.174)	0.069 (0.203)
Region: : : Midwest	−0.082 (0.143)	−0.090 (0.161)	0.048 (0.154)	0.015 (0.186)
Region: : : South	0.126 (0.135)	0.149 (0.153)	−0.015 (0.146)	0.061 (0.171)
Region: : : West	−0.113 (0.160)	−0.093 (0.178)	0.151 (0.155)	0.220 (0.181)
Urban Rural: Suburb	−0.382** (0.128)	−0.274* (0.139)	−0.093 (0.132)	0.026 (0.147)
Urban Rural: Town	−0.415* (0.163)	−0.498** (0.184)	−0.410* (0.161)	−0.340 (0.182)
Urban Rural: Rural area	−0.321* (0.154)	−0.289 (0.168)	−0.034 (0.152)	0.080 (0.186)
Political interest: Some of the time	−0.135 (0.117)	−0.429*** (0.130)	−0.188 (0.118)	−0.265 (0.141)
Political interest: Only now and then	−0.080 (0.160)	−0.503** (0.182)	−0.175 (0.196)	−0.347 (0.218)
Political interest: Hardly at all	−0.521* (0.227)	−0.971*** (0.248)	−0.319 (0.252)	−0.733** (0.270)
Conspiracy Score	0.119*** (0.008)	0.131*** (0.009)	0.106*** (0.008)	0.112*** (0.010)
Treatment (vs. Placebo) × Article (vs. Chatbot)	−0.922 (0.501)	−0.489 (0.533)	−1.003* (0.510)	−1.059* (0.502)
Treatment (vs. Placebo) × Party Identification: Independent	−0.090 (0.265)	0.004 (0.293)	−0.280 (0.281)	−0.005 (0.328)
Treatment (vs. Placebo) × Party Identification: Republican	−0.158 (0.254)	−0.044 (0.276)	−0.150 (0.256)	0.018 (0.292)
Article (vs. Chatbot) × Party Identification: Independent	−0.356 (0.511)	−0.534 (0.569)	−0.466 (0.474)	−0.389 (0.530)
Article (vs. Chatbot) × Party Identification: Republican	−0.157 (0.468)	−0.411 (0.555)	−0.481 (0.453)	−0.750 (0.463)
Treatment (vs. Placebo) × Article (vs. Chatbot) × Party Identification: Independent	1.077 (0.684)	1.395 (0.782)	0.629 (0.710)	0.478 (0.756)
Treatment (vs. Placebo) × Article (vs. Chatbot) × Party Identification: Republican	0.607 (0.635)	0.507 (0.715)	0.824 (0.648)	1.083 (0.675)
Constant	−2.495*** (0.308)	−2.014*** (0.346)	−2.734*** (0.309)	−2.389*** (0.381)
Observations	1,944	1,939	1,850	1,850
Log Likelihood	−4,133.074	−4,355.123	−3,915.082	−4,151.788
Akaike Inf. Crit.	8,334.148	8,778.246	7,898.163	8,371.577
Note:				*p<0.05; **p<0.01; ***p<0.001

A.15 AI-Type and Party Supplementary Figures

These figures preserve modality- and party-based visual diagnostics used during manuscript development. They are directionally consistent with the tabular results, but they are not needed to read the main paper or the earlier appendix sections.

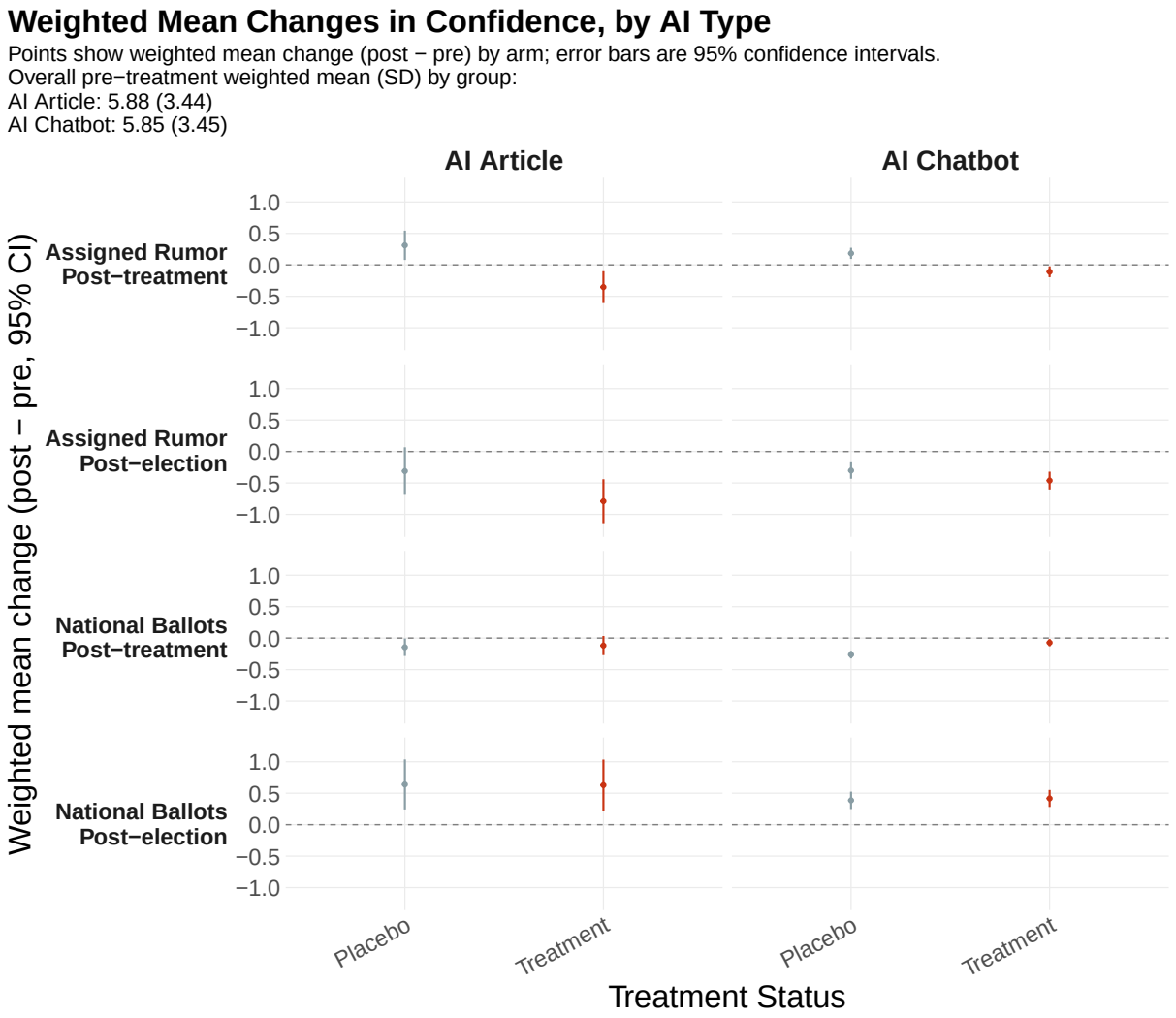


Figure A.2: **AI Type: Weighted Mean Changes.** Weighted post-minus-pre mean changes by arm and AI type. The continued page reports the corresponding regression-based treatment effects for the same subgroup splits.

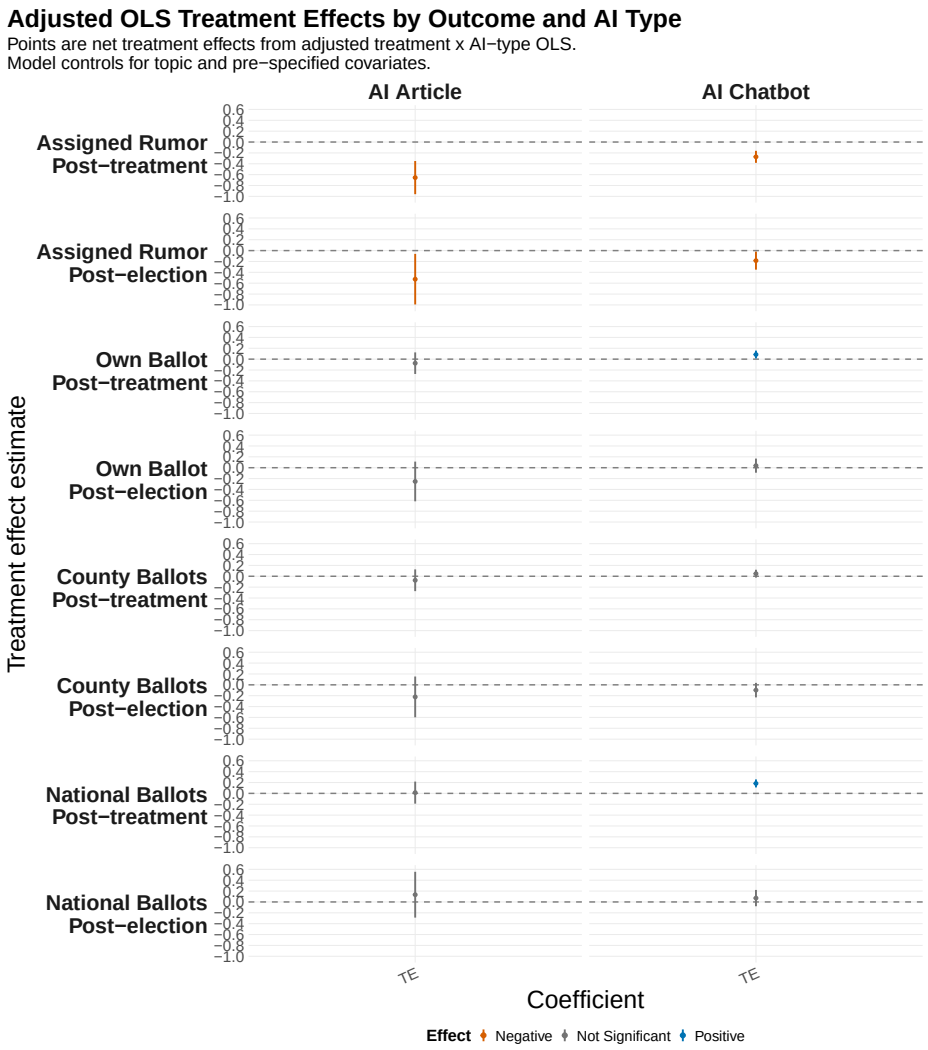


Figure A.2: **AI Type: Regression-Based Treatment Effects.** Regression-based treatment effects for the same AI-type partitions shown in Figure A.2.

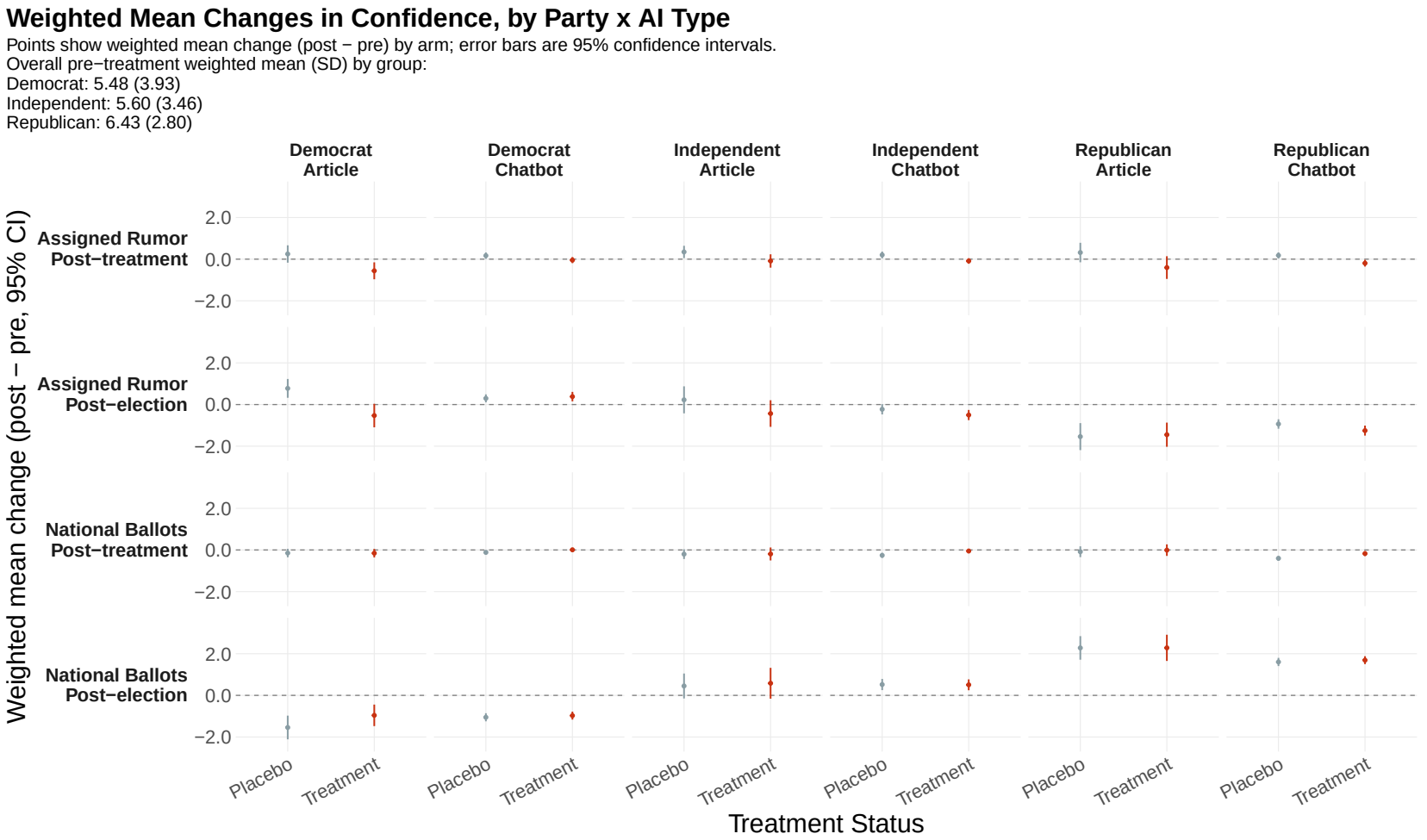


Figure A.3: **Party by AI Type: Weighted Mean Changes.** Weighted post-minus-pre mean changes by arm and party-by-AI-type subgroup. The continued page reports the corresponding regression-based treatment effects for the same subgroup splits.

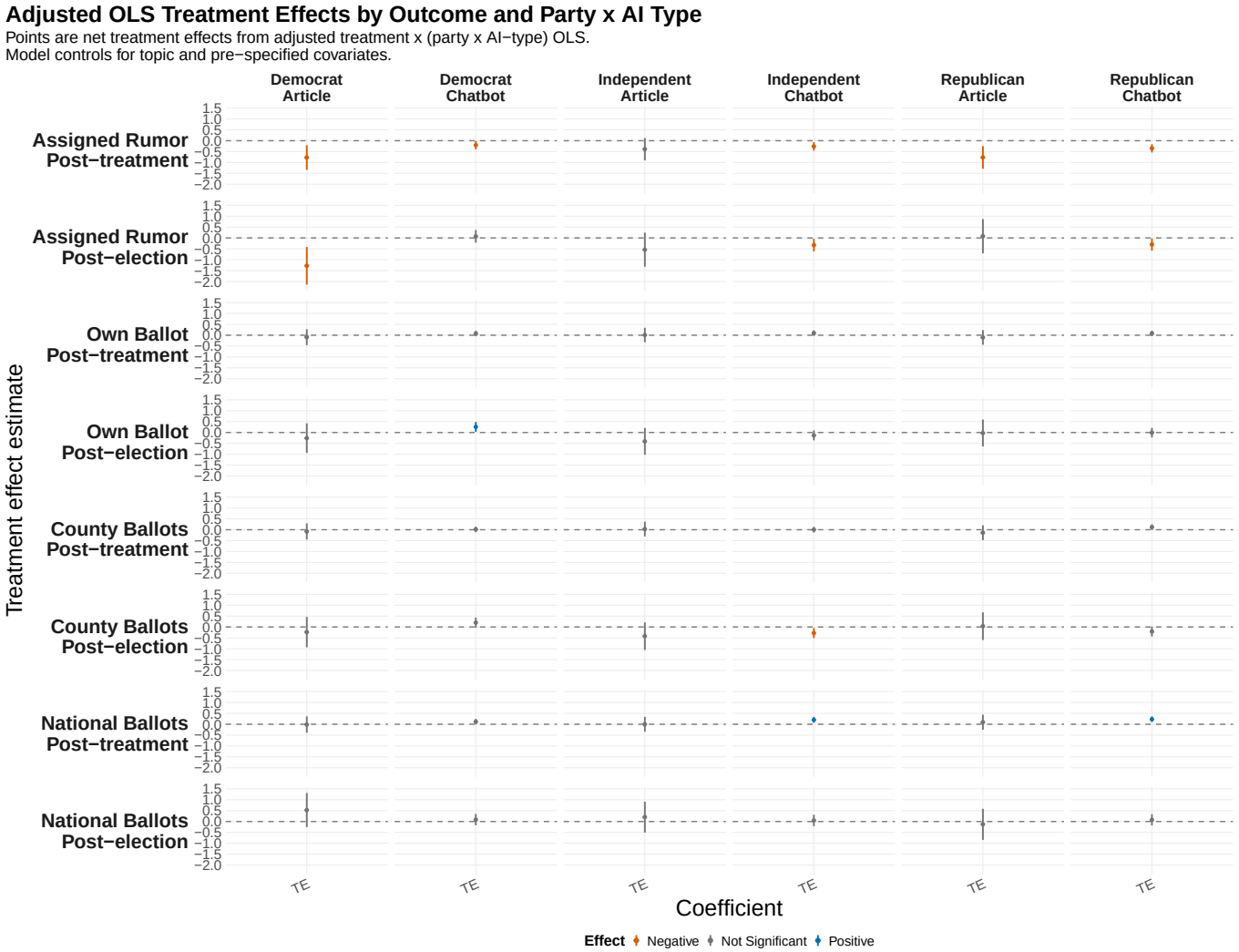


Figure A.3: **Party by AI Type: Regression-Based Treatment Effects.** Regression-based treatment effects for the same party-by-AI-type partitions shown in Figure A.3.

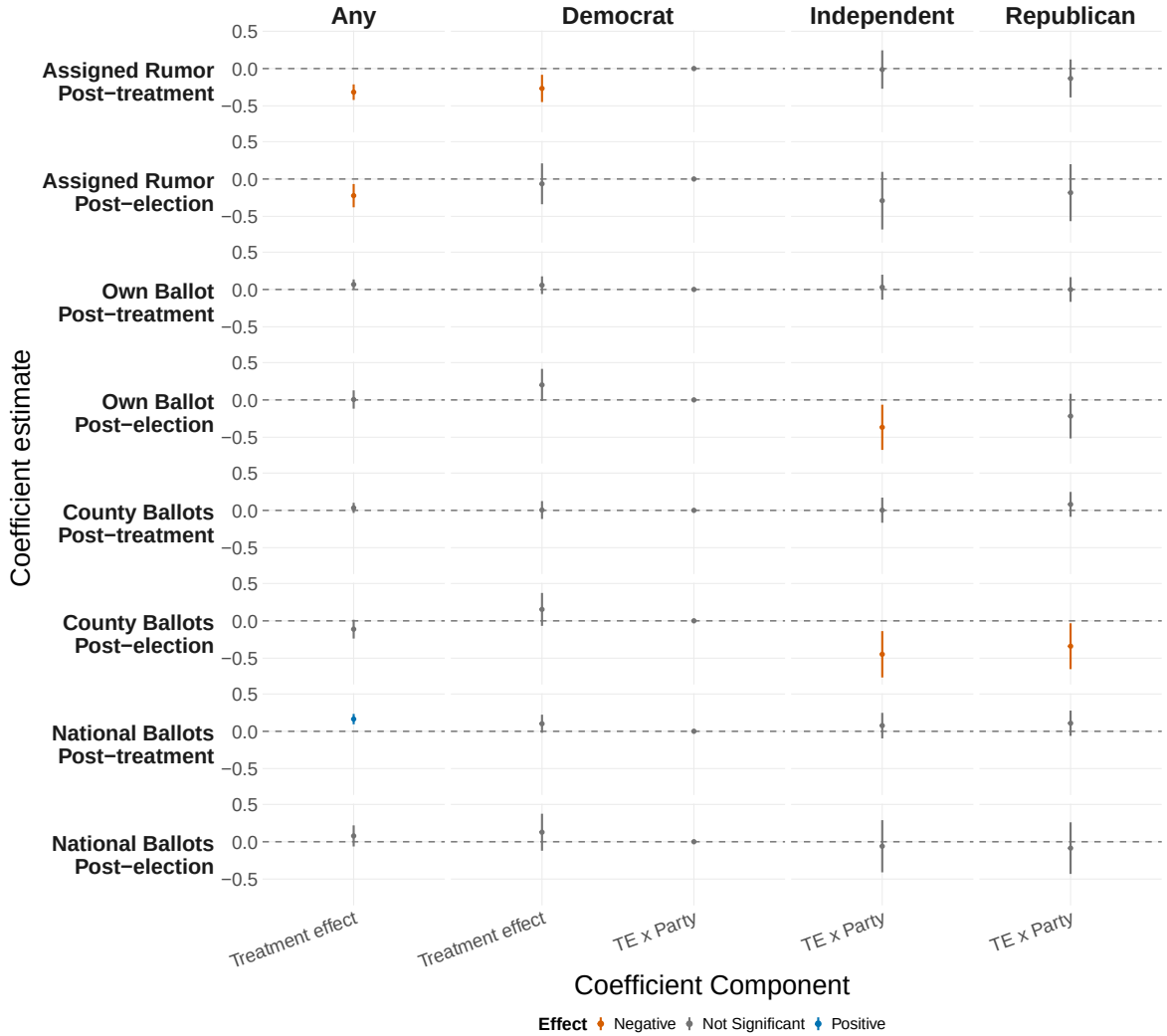


Figure A.4: **Party Decomposition of Treatment Effects by Outcome.** Orange estimates show adjusted treatment effects and gray estimates show party-differential components from interaction models with Democrats as the reference group.

Table A.54: Treatment effects and party interaction coefficients by outcome

Coefficient	Assigned rumor Post-treatment	Assigned rumor Post-election	Own ballot Post-treatment	Own ballot Post-election	County ballots Post-treatment	County ballots Post-election	National ballots Post-treatment	National ballots Post-election
Any: treatment effect	-0.345*** (0.054)	-0.200* (0.080)	0.074* (0.035)	-0.012 (0.063)	0.026 (0.035)	-0.117 (0.064)	0.151*** (0.036)	0.056 (0.073)
Democrat: treatment effect	-0.286** (0.094)	-0.061 (0.140)	0.055 (0.061)	0.191 (0.109)	0.015 (0.062)	0.135 (0.111)	0.089 (0.062)	0.099 (0.126)
Independent: differential vs. Democrat	-0.020 (0.134)	-0.225 (0.199)	0.025 (0.086)	-0.380* (0.155)	-0.044 (0.088)	-0.436** (0.158)	0.041 (0.089)	-0.063 (0.180)
Republican: differential vs. Democrat	-0.149 (0.131)	-0.190 (0.196)	0.030 (0.085)	-0.232 (0.152)	0.071 (0.086)	-0.322* (0.155)	0.139 (0.087)	-0.066 (0.176)
Estimation sample N	5,191	4,203	5,190	4,203	5,191	4,202	5,190	4,203

Notes: Entries are weighted OLS coefficients on post-minus-pre change outcomes and reproduce the coefficient components plotted in Figure A.4. The "Any" row reports the pooled treatment effect from models without party interactions. The "Democrat" row reports the treatment effect from the corresponding treatment × party model with Democrats as the reference group. The Independent and Republican rows report the party-differential interaction components relative to Democrats from that same interaction model. All models adjust for randomized rumor topic and the preregistered demographic and attitudinal covariates. Post-election outcomes use the recontact survey weights. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.