

**AI-Supported Information Seeking: The Role of Sycophancy and Biased
Information on Attitudes on Public Policy**

Jennifer Stromer-Galley, Syracuse University, jstromer@syr.edu

Sarah Hoover, Syracuse University, shoove@syr.edu

Abstract

As AI-supported search becomes an increasingly common mode of information seeking, questions arise about how AI shapes public opinion on policies. Drawing on the Search Engine Manipulation Effect (SEME; Epstein & Robertson, 2015) and the HAIL-TIME framework (Sundar, 2020), this paper reports findings from a pre-registered between-subjects experiment examining how AI information bias and sycophancy—as a flattering tone—influence attitudes toward a lesser known policy of ranked-choice voting (RCV). Participants were randomly assigned to one of four conditions varying the bias of the information provided in search returns (pro vs. anti-RCV), and message tone (neutral vs. high sycophancy). Results indicate that information bias had a strong effect on RCV attitudes. Sycophancy produced a smaller but significant independent effect. Prior trust in AI weakened susceptibility to biased content, while machine heuristic beliefs amplified susceptibility to biased content; prior exposure to RCV buffered against both effects. Additionally, those with lower levels of political attention were more likely to align their beliefs with the AI system. These findings raise important concerns about the role of AI-supported search in shaping public opinion on low-salience policy topics.

Keywords: Artificial Intelligence, search, sycophancy, search engine manipulation effect, interactivity

Introduction

AI capabilities are evolving at a remarkable rate. Generative large language models (LLMs) such as ChatGPT can search the Web in real time. Google's search engine embeds Gemini, their generative LLM, directly in search results. Thus, the information seeking environment is now substantially changed, raising questions about the effects that conversational information search will have on people's attitudes around current events and public policy. This, in turn, affects policy-making and governance in representative democracies.

Information seeking on the internet as an interactive process is novel. Google's older, keyword based, and non-interactive search returns information tailored to the ideological predispositions of the searcher (Hindman, 2009), which shapes public opinion on policy topics (Epstein & Li, 2024). With AI functioning as a search partner, the process and outcomes may change. Research on interactive search suggests that it can exacerbate echo chambers of ideologically-aligned (that is, *biased*) information seeking (N. Sharma et al., 2024).

Search engines are part of the information eco-system that supports public opinion. Quality public opinion is a cornerstone for effective representative democracies (Delli Carpini & Keeter, 1996; Habermas, 1981), as the public needs to be informed and hold well-founded opinions that frame their own worldviews while also understanding that of others in a diverse society. Search engines provide quick access to high-quality information, such as news stories, government websites, advocacy and public interest websites. Prior research on keyword search returns from traditional search engines finds that they often provide a narrow range of ideologically-aligned information that can shape the public's attitudes in problematic ways (Hindman, 2009). Indeed, the Search Engine Manipulation Effect (SEME) finds that people's attitudes on current events and policy topics are shaped by search engine returns (Epstein & Li, 2024), including on current events and elections (Epstein & Robertson, 2015).

Sycophancy, understood as a flattering tone by AI (Cheng, Yu, et al., 2025) is an emerging topic in AI. Research suggests that sycophantic AI may be perceived as more authentic (Sun & Wang, 2025) but has not explored if it changes attitudes. Perceived trust in AI and the machine heuristic—where people judge machines more capable and trustworthy (Sundar & Kim, 2019)—may moderate the effects of bias and sycophancy on attitudes.

To test these relationships, we conducted a preregistered experiment in which we manipulated sycophancy and information bias—creating an AI search tool that was either high or neutral in sycophancy and either provided source materials and summaries in favor of or against a lesser-known policy topic. Specifically, we focused on ranked-choice voting (RCV)—a preferential voting system as compared with the more common plurality voting—that is an emerging policy topic in the United States (Kimball et al., 2021). Given that the public knows less about it, their opinions are potentially more malleable; thus, it is a useful policy topic to test AI effects on attitudes.

Participants recruited via Prolific’s general population sample (N=339) were tasked with searching for information on RCV. The participants searched using a simple chat interface with a Claude-driven chatbot. The chatbot’s responses were focused using Retrieval Augmented Generation (RAG) on a pre-selected set of news stories and Wikipedia pages on RCV. The pro-RCV condition included links to supportive advocacy websites; the anti-RCV condition linked to antagonistic advocacy websites. To create the high and neutral sycophancy conditions, the AI received instructions that dictated its tone.

The findings from this study suggest that AI bias and sycophancy have significant effects on participants’ attitudes towards RCV. Furthermore, the effect of information bias was higher for participants with lower exposure to RCV, while demographic variables were not found to

modify experimental effects. Pre-existing trust towards AI dampened the effects of information bias, while the machine heuristic—a cognitive tendency to believe AI is more secure and trustworthy than humans (Sundar & Kim, 2019)—amplified the effects of information bias. Although users in the anti-RCV condition reported lower levels of post-interaction trust in AI compared to the pro-RCV condition, post-test trust in AI did not mediate the effects of sycophancy or bias. These results raise concerns about generative AI shaping public opinion on policy topics, particularly for users who are unfamiliar with the topic.

Literature Review

Search engines have been part of Web-based information seeking processes since the 1990s, starting with indexes (i.e., Yahoo!), and then the dominant crawler-based, link-tracking, and algorithmic approach developed by Google (Halavais, 2013). In 2025, AI companies, such as Anthropic’s Claude and OpenAI’s ChatGPT implemented real-time search into their free and premium generative AI tools. Google’s AI, Gemini, was integrated into Google search for an optional “AI Mode” search experience in the same year. By 2026, Google shifted their Google search experience from a key-word search to an AI-mode where search returns activated interactive chat with Gemini for further inquiry on that search.

Researchers have explored the extent to which searches with Google’s keyword-based engine shapes knowledge and opinions on policy topics. Research shows that these older, key-word approaches filter users to a narrow subset of total information available (Hindman, 2009), and that users’ prior search histories and preferences are used in Google’s algorithm to shape returns to likely preferred information (Hindman, 2009; Tripodi, 2022). This effect has been found on a variety of topics, including news and current events (Trielli & Diakopoulos, 2019). Epstein and Robertson (2015) refer to this as the *Search Engine Manipulation Effect* (SEME). In

a series of experiments, they find that bias in the rankings of keyword searches about elections can significantly shift undecided voters' attitudes (Epstein & Robertson, 2015). They also find biased returns can significantly shift attitudes on policy topics on which the searchers have weak prior attitudes (Epstein & Li, 2024).

AI-supported search raises new questions about potential effects. Searching for information with AI entails a user experience in which the searcher is conversing in complete sentences, rather than keywords, to initiate a query in an interactive mode. An AI-search experience also entails that the search chatbot is responding with not only search results but also short summaries that synthesize those results. The experience is thus a conversational, interactive experience, and implicates the scholarship on interactivity in computer-mediated communication (CMC).

A large body of scholarship has explored the importance of interactivity in CMC and in human-machine communication (HMC). Conceptually, interactivity has two core dimensions (Stromer-Galley, 2004; Sundar, 2007): (1) interaction between humans that is mediated by digital technology, such as through a chat interface, a threaded message board, or a social media posting and commenting; and, (2) interaction between humans and a digital interface, such as reactions to social media posts, click-polls, and hyperlinks.

Communicating with a chatbot potentially blurs the boundaries between those two dimensions, since a chatbot is a technology that has a user interface, but that communicates in ways that feel distinctly human through social and emotional cues (Araujo, 2018; Pezenka et al., 2024). Research on chatbots suggests that the more human-communication features they possess the greater the perceptions of anthropomorphism—that is, the feeling that humans are chatting with other humans (Schuetzler et al., 2020). The Computers as Social Actors (CASA) paradigm

postulates that humans treat computers as social actors in a variety of contexts and even when they know they are communicating with machines (Reeves & Nass, 2003). As such, human-AI communication may be more akin to interaction between humans in terms of the cognitive processes that are invoked. Interactivity between humans is understood as a sequence of turns between interlocutors such that subsequent messages are shaped by prior messages in a mutual feedback loop (Rafaeli, 1988). This is conceptualized as *message interactivity* and operationalized as *contingency*, meaning that message input shapes message output (Sundar et al., 2016). Heightened contingency in interaction is theorized to increase cognitive engagement, which influences communication outcomes (Sundar, 2007). Indeed, research suggests that interactivity is a dimension of persuasion in digital media because interactivity increases cognitive absorption of information (Oh & Sundar, 2015). For example, a study of humans interacting with a chatbot as compared with a website providing product recommendations found that chatbots affected attitudes more strongly than a website (Ischen et al., 2020).

One dimension on which human-AI interaction has distinct effects on attitudes that are different from human-human interaction is in perceptions of trust. Sundar and Kim (2019) find that people trust machines to handle personal information more securely and objectively than humans. In a study in which participants were told they were interacting with a human agent or an AI agent to help them book a flight, participants reported higher trust in the AI agent to secure their personal information. Sundar and Kim label these perceptions the *machine heuristic* and describe it as similar to other cognitive shortcuts that humans engage in when processing information and making decisions, such as confirmation bias and anchoring bias. While trust is defined in different ways across disciplines, generally it shares a common grounding in the idea

that people are willing to be vulnerable while expecting others to be honest and cooperative (Rousseau et al., 1998; Van Lange et al., 1998).

The interactivity cues in digital systems enable functionalities that have psychological effects. This has been conceptualized into the Theory of Interactive Media Effects (TIME; (Sundar et al., 2015). In the model, interactivity affordances provide cues and actions that trigger heuristics, perceptions, and levels of engagement, which in turn affect several variables including attitudes, knowledge, and behaviors. Yet, research has only begun to understand how humans interacting with AI (or human-AI interaction, HAI) affect a variety of outcomes. The Human-AI Interaction based on the Theory of Interactive Media Effects (HAI-TIME Model; Sundar, 2020) was developed to help focus researchers on key predictors, mediating variables, and potential outcomes when humans interact with AI. The model highlights that AI communication features and their underlying algorithms affect the cues that humans attend to when interacting with AI. Thus, AI messaging is mediated by the cognitive processing of that messaging, in turn shaping perceptions and outcomes of the interaction.

In the HAI-TIME model, message features can have noteworthy effects. The current generative models, such as OpenAI's ChatGPT, have developed a distinctive quality of sycophancy (Cheng, Yu, et al., 2025; M. Sharma et al., 2025). Popular news stories have highlighted this trend and both the upsides—increased engagement and feelings of companionship in users (Bryan, 2026)—and downsides—increased dependency and even engaging in delusional spirals where users have a cognitive break with reality through their interactions with generative AI (Hill & Freedman, 2025). Conceptually, AI scholars are not in agreement on what exactly sycophancy is, focusing on the alignment between the human and the AI response by tone, such as high flattery (Cheng et al., 2025), or by preferred response (Perez et

al., 2022), generally understood as attitudinal alignment in the content between the AI and the user (Sharma et al., 2025). Social scientists conceptualize sycophancy (outside of the context of AI) as having three dimensions: (1) content that conforms favorably between interlocutors with (2) an affirming tone, typically used in a (3) power-down relationship, often in political or workplace settings (Baturu et al., 2024). Research of sycophancy between humans finds that ingratiation by subordinates to leaders improves promotion opportunities and positive work experiences (Bolino et al., 2016; Kim et al., 2018), which suggests positive evaluations of those who engage in flattery by the targets of it. Preliminary research on sycophancy in AI suggests it has noteworthy effects on users: it may be perceived as more authentic (Sun & Wang, 2024), and it can increase feelings of trust in the AI (Cheng, Lee, et al., 2025). Unknown is how sycophancy interacts with user's prior attitudes on policy topics in the context of search.

One effect of working with interactive media is changes in attitudes. SEME highlights that opinion shifts on topics when held attitudes are weak. One public policy topic on which the public has not formed strong attitudes is ranked choice voting (RCV). This approach provides an alternative to the dominant standard plurality ballots in the United States where voters chose only one candidate from a slate. Instead, in RCV voters rank their preferences for all of the candidates on the slate; if their first choice does not get a majority of votes, then it transfers to their second choice, and so on. Research has found that while several municipalities and states are exploring or have adopted RCV (including New York City, which led to the mayoral win by Zohran Mamdani in 2025), the public has little knowledge of nor strong attitudes about this alternative to plurality voting (Kimball et al., 2021).

Given the theory of SEME in the context of RCV, we examine the effects of AI-supported search and hypothesize that:

H1: Participants exposed to pro-RCV information will report more favorable RCV attitudes than those exposed to anti-RCV information.

We further postulate that sycophancy in the AI's communication may reduce resistance to perspectives that run counter to their held beliefs, called reactance. In the context of persuasion, the Elaboration Likelihood Model (ELM) is especially relevant. ELM is a dual-processing model that theorizes attitude change in the context of persuasion can be explained not just on the quality or validity of the message or the credibility of the source but also upon the routes through which the message is processed (Petty et al., 1981; Petty & Cacioppo, 1984). Central route processing is a deliberative route wherein the information is evaluated relative to prior information the individual possesses. Peripheral route processing devotes minimal effort to scrutinizing the message's arguments or source credibility and instead relies on preexisting attitudes and the presence or absence of heuristic cues, such as an attractive source, message length, or number of arguments to evaluate the message. Individuals relying on the peripheral route are more likely to be persuaded by a message when it aligns with their preexisting attitudes or when the source has likability characteristics.

The degree of motivation to process a message is affected by several variables. Of particular interest here is a cognitive heuristic that likely emerges around information seeking: the confirmation heuristic (Nickerson, 1998)—where people unconsciously attend to preferential information that aligns with preexisting beliefs and ignore disconfirming information. Research, for example, finds that when people hold pre-existing beliefs and are confronted with contrary information, they engage in motivated reasoning to counter-argue the information (Lodge & Taber, 2013). This is also called *reactance*, and is theorized as a cognitive way to reduce the feeling that one's freedom is under threat and to regain control (Brehm & Brehm, 1981). We

postulate that sycophancy, however, may reduce reactance by reducing feelings of threat thereby increasing the uptake of biased information. We thus hypothesize:

H2: The effect of information bias on attitudes will be stronger when the AI uses heightened sycophancy compared to more neutral communication.

Given prior research on the machine heuristic increasing perceptions of trust (e.g. Sundar & Kim, 2019), we further postulate that:

H3: Pre-existing machine heuristic beliefs will moderate the effect of sycophancy on RCV attitudes, such that participants with stronger machine heuristic beliefs will be more influenced by heightened sycophantic AI.

H4: Pre-existing trust in AI will moderate the effect of sycophancy on RCV attitudes, such that participants with higher trust in AI will be more influenced by heightened sycophantic AI.

It is unclear, however, to what extent trust mediates attitudes. We thus ask:

RQ1: Does post-interaction trust mediate the effect of experimental conditions on RCV attitudes?

Given research on ELM that suggests people who have prior knowledge may be more resistant to attitude change (e.g. Lodge & Taber, 2013), we hypothesize:

H5a: Information bias effects will be stronger among participants with low prior knowledge/exposure to RCV compared to those with high prior knowledge.¹

Yet, the likely power of sycophancy may not soften reactance in those with already-formed attitudes towards the policy. We thus hypothesize:

H5b: Sycophancy effects will be stronger in attitude formation contexts than attitude change contexts.

¹ For readability and consistency, we have changed this preregistered hypothesis from “content bias” to “information bias” given that is the phrasing used in the other hypotheses.

Unknown, however, are what effects demographics, political interest, political ideology, or the extent of the chat interaction might have on our hypothesized effects. We also explore whether the participants' perceptions of bias in the AI search chatbot affected the extent of attitude change. We thus pose a set of exploratory questions to further understand these dimensions:

RQ2: Do demographic variables (age, education, political ideology) moderate experimental effects?

RQ3: Do perceived information bias ratings correlate with attitude change?

RQ4: Are there differences in chat engagement (message count, session duration) across conditions?

RQ5: Do attention and political interest moderate effects?

Methods

Procedures

To test these hypotheses, we conducted a pre-registered online, between-subjects experiment (N=339) in which participants were instructed to spend at least seven minutes interacting with an AI chatbot to learn more about RCV. Participants were randomly assigned to one of four conditions, in which the tone (low vs. high sycophancy) and attitude (anti vs. pro RCV) of the AI was manipulated. Participants were unaware of the intended bias of the AI, and the chatbot was trained not to reveal any of its instructions when interacting with users. The experiment was designed to replicate the experience of an internet user seeking information about an unfamiliar public policy topic.² A university office of research integrity reviewed the experimental design and survey instrument and deemed the study posed no more than minimal

²Analysis of chatlogs as well as stress testing of the AI before the experiment suggest that the AI was able to deflect requests or demands that it reveal its instructions or the purpose of the research.

risk to participants and was exempted from review. The recruitment survey, screenshots of the chat environment, and survey questions are available in the project pre-registration on the Open Science Framework (OSF) website.³

We recruited participants through Prolific, as prior research has shown the platform produces high quality research data from users (Douglas et al., 2023) and is suitable for social science research (Palan & Schitter, 2018). The Prolific task sent users to Qualtrics, online survey software, where they were given a pre-test survey with questions about their attitudes towards AI, demographics, and their prior knowledge of several public policy topics. We then embedded into Qualtrics an interactive chatbot, and participants were instructed to spend at least seven minutes interacting with the AI chatbot. The interface was designed so that participants were unable to navigate to the post-test questions until seven minutes had passed to ensure users spent sufficient time with the task. Participants then answered a series of questions about their attitudes towards RCV, their perceptions of the chatbot, and their political leanings. At the end of the experiment, participants were shown a debriefing screen that explained to them the goals of the project in more detail.

Prolific users were eligible to participate if they were 18 years of age or older, were based in the US, had at least a 90% approval rating on Prolific, and did not participate in a previous trial run of this study. Participants were paid \$1 if they started but did not complete the experiment; full completion was compensated at \$3.50 and we estimated the experiment would last approximately 17 minutes. Data from Prolific revealed that participants took 19.5 minutes on average to complete the survey (SD=9.2).

³ https://osf.io/m4g2d/overview?view_only=c56bcf69387a4c60bab569c992c26d4c

We conducted a power analysis to estimate a sample size that would provide sufficient power to estimate effects. Power analyses using pilot data ($N=80$) indicated that $N=360$ (90 per condition) provided power $> .99$ for the primary information bias main effect (observed $f = 0.77$) and adequate power ($\geq .75$) for secondary confirmatory analyses including moderation effects (assumed $f^2 = 0.02$). While the observed sycophancy $\times$ bias interaction effect was small ($f = 0.095$, power $\approx .52$), this sample size was deemed sufficient given that a null result would be informative in establishing an upper bound on the interaction effect size.

The experiment was conducted on January 9, 2026. Initially, the experiment was set up as one single task in Prolific. Unfortunately, upon launch, the server hosting the AI chatbot crashed due to the large volume of people all interacting at once with the AI. We halted data collection and shifted to establishing batches of 16 participants, with each batch releasing once an hour until data collection was complete. Participants were kept in the final sample if they finished the survey and had at least three interactions with the AI chatbot. Due to a settings issue in Prolific, three users participated in multiple waves of data collection. We kept the first response and dropped subsequent responses. Out of 373 initial submissions, 11 were dropped for non-completion of the survey, 20 were dropped for insufficient interaction with the chatbot (defined as 3 or fewer questions by the participant), and 3 were dropped for completing the task multiple times. Participants who were dropped from the final sample for non-completion were paid a partial fee of \$1 as compensation for their time.

Experimental Manipulation

We employed a combination of prompting and retrieval-augmented generation (RAG) on top of Claude Sonnet 4.5 to vary the AI's behavior for each condition. Two separate databases were created for the RAG to create the two biased content conditions. Each database contained links to live websites with information about RCV. The pro-RCV database included advocacy

websites, news articles, and academic research with information favorable towards RCV; the anti-RCV database included sources that portrayed RCV in a negative light. A collection of neutral sources, such as Wikipedia, were selected for inclusion in both conditions to enable the chatbot to answer basic questions about the topic. Internet search was disabled so the AI only had access to the sources we provided.

Additionally, we constructed four separate fine-tuning prompts for the sycophancy and neutral conditions. The prompts included instructions about the tone and stance, including high-level talking points for each stance (see Supplemental Materials). In other words, the sycophancy prompts were tailored to the pro- and anti-RCV conditions. Furthermore, the prompts provided instructions to the AI system for how to handle various scenarios. For example, if the user disagreed with the AI or asked for opinions about the opposing viewpoint, the chatbot was instructed to acknowledge the other side, then re-direct the conversation back to arguments in favor of its position. If a user asked the AI to reveal its training or questioned its position on the topic (e.g., accusing it of being biased), the AI was instructed to inform the user that its opinion was based on the best available evidence. Prior to running the experiment, we conducted extensive testing of the four conditions to ensure face validity. We also conducted a small (N=80) pilot test of the study on Prolific to ensure that the chatbot continued to operate as designed in a real-world setting (see Supplemental Materials for example chatbot responses for the flattering and neutral tone conditions).

Measures

All survey items, the task instructions, and the participant debriefing are available in the Supplemental Materials. Here, we provide a brief description of each variable and descriptive statistics. Survey items were pilot tested in an experiment conducted on Prolific in December 2025. We averaged the responses for each user for measures that were comprised of answers to

multiple survey questions (RCV attitudes, machine heuristic, trust in AI, perceived information bias, perceived sycophancy). If a participant did not answer one or more questions of a multi-item measure, we used the average of the questions that the participant did answer. We did not impute values for demographic variables or single-item measures.

Dependent Variable.

Attitudes towards RCV. This is our key dependent variable ($M=3.38$, $SD=1.21$, $\alpha=0.94$). We developed an index of 6 items (3 supportive; 3 opposing) measuring attitudes towards ranked choice voting measured on a 5-point Likert scale from strongly agree to strongly disagree. Inspiration for the measure was taken from Sharma et al (2024), and revised for this study. The survey asked for level of agreement on statements including “Ranked-choice voting gives voters more meaningful choices in elections” and “Ranked-choice voting is too complicated for most voters to understand.”

Independent, Mediator, and Moderator Variables. In addition to our two manipulated variables, sycophancy and information bias, we measured perceptions and attitudes regarding AI. These include the following:

Machine Heuristic (pre-test). We used a 27-item scale developed by Yang and Sundar (2024) ($M=4.86$, $SD=0.76$, $\alpha=0.90$). Participants were asked to rate their level of agreement on a 7-point scale about perceptions of “AI systems, like ChatGPT, Claude, and Gemini”. Response options included statements such as: “AI is knowledgeable,” “AI is boring,” “AI is efficient.” The final measure was grand-mean centered so that values that were below average were negative, and those that were above average were positive.

Trust in AI (pre-test and post-test). A 6-item scale developed by Merritt (2011) to measure participants’ trust in AI. The same 6-item scale was used in both the pre- ($M=2.59$,

SD=1.03, alpha=0.95) and post-test (M=2.61, SD=1.10, alpha=0.95) questionnaires, however the post-test questions referenced the AI system from the experiment, such as “I believe the AI search system is a competent performer.” Both measures were grand-mean centered.

*Perceived bias in AI.*⁴ A 4-item scale administered post-test that measured participants’ perceptions about the degree of bias in the AI system they used (M=3.00, SD=1.48, alpha=0.94). The question asked participants to consider the information they received from the system and to indicate their level of agreement with the statements on a 5-point Likert scale, such as “The AI system presented ranked-choice voting in a mostly positive light”. The measure also included several reverse-coded items, such as “The AI system emphasized drawbacks to ranked-choice voting.” An additional survey item asked participants to rate their agreement with the following statement on a 5-point scale: “The AI search tool seemed to favor one side of the ranked-choice voting debate.” Participants assigned to the pro-RCV (M=3.88, SD=1.28) and anti-RCV conditions (M=3.97, SD=1.23) tended to agree with the statement on average. There was no significant difference between the two groups (p=0.54).

Control Variables.

Prior Knowledge of and Exposure to RCV. To understand users’ pre-existing knowledge and exposure to RCV, participants were asked to rate how much they knew about RCV (M=2.13, SD=1.10) and how much they had heard about it (M=2.44, SD=1.17) on a 5-point Likert scale. These questions were embedded in a series of similar questions about other policy topics, to avoid priming participants about RCV. The responses were assigned a score of 1-5 for analysis, with reverse-coding where appropriate.

⁴ The preregistration erroneously designated the perceived information bias scale as both a manipulation check and a mediator variable. We report it here solely as a mediator, as this is the theoretically appropriate role in this analysis.

Political Interest. One question asked “How interested in politics are you?”. Four response options ranged on a scale from 1 to 4, from “not at all interested” to “very interested” ($M=2.99$, $SD=0.89$).

Political Attention. One question asked “How much attention to you pay to current events and politics?”. Provided response options, ranked from 1 to 6, included “I rarely follow current events”, “Once or twice a month”, “Less than once a week”, “Once or twice a week”, “Two or three times a week”, and “Daily” ($M=4.91$, $SD=1.35$).

Political Partisanship. Two questions measured political partisanship. The first question asked participants: “Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or something else?”. If they selected Independent, Other, or No Preference, they were asked a second question: “Do you lean more towards the Democrats or the Republicans.” Response options included “Lean Republican,” “Lean Democrat,” or “Neither.” 105 participants answered the question about political lean. We then created two dummy variables to indicate those who identified as Republicans and those who identified as Democrats, treating Republicans as the reference category. Participants who did not answer the question about political affiliation were categorized according to their response to the question about political lean. Write-in responses were reviewed manually and assigned to the most similar category.

Demographics. The pre-test survey included four demographics questions. These included self-described gender, highest level of education, race and/or ethnicity, and age. The final sample was 68% White, 10% Black, 3% Hispanic/Latino, 6% Asian, and 13% mixed. 55% had a Bachelor’s degree or higher. 55% of respondents were female, 42% were male, and 1% were non-binary. The most common age category was 26-35. Our sample overrepresented

women, slightly under-represented Blacks and Whites, and significantly under-represented Hispanic/Latinos, while over-representing those who self-reported as mixed-race as compared with the U.S. Census (U.S. Census, 2025). Although Prolific gave a more diverse sample than research pools at universities and some professional panels, we note that our findings may not fully generalize to the population.

Manipulation Check.

Sycophancy. This measure consisted of 5 7-scale items ($\alpha=0.79$). Two-tailed t-tests ($p<0.001$) confirmed that participants assigned to the sycophantic condition ($M=5.34$, $SD=0.94$) were more likely to report that the AI had a flattering tone than participants assigned to the neutral condition ($M=3.13$, $SD=0.99$).

Results

Our findings suggest that those in the pro-RCV conditions (H1) and those in the sycophantic conditions (H2) were more likely to support RCV. We tested these hypotheses with an ANOVA model, treating the two experimental manipulations as independent variables. The neutral tone condition and anti-RCV condition were treated as reference categories.⁵ Among participants assigned to the neutral-tone condition, those in the pro-RCV condition were more likely to indicate their support for RCV ($F(1, 184)=94.1$, $p<.0001$, $\eta^2=.34$). The effect size for bias was similar when all participants were included in a 2x2 ANOVA model ($F(1, 335)=193.75$, $p<0.0001$, $\eta^2=.37$). Results from Tukey's test confirm that the difference in RCV support between those in the pro-RCV and anti-RCV conditions is statistically significant ($M_{\text{pro}} = 4.14$,

⁵ Our pre-registration indicated that H2 will be supported if the interaction between bias and sycophancy is significant. We instead report the main effect for sycophancy in this section, to address the impact of sycophancy on its own, and acknowledge that the interaction is not significant. Furthermore, the pre-registration indicated that we would use effects coding of (-0.5, 0.5) for experimental effects in the ANOVA model: for simplicity, and to align with standard ANOVA procedures, we instead used dummy coding (0,1).

$M_{\text{anti}} = 2.68$, $p < 0.001$). Furthermore, we find significant effects for the sycophantic tone ($F(1, 335) = 5.92$, $p = 0.0155$, $\eta^2 = .02$), which are confirmed by the result of Tukey's test ($M_{\text{sync}} = 3.51$, $M_{\text{neutral}} = 3.27$, $p = 0.015$). The interaction between sycophancy and bias was not significant ($F(1, 335) = 0.017$, $p = 0.895$, $\eta^2 < 0.0001$). When restricting the analysis to those assigned to the pro-RCV condition, we found significant effects for sycophancy ($F(1, 160) = 4.406$, $p = 0.037$, $\eta^2 = .03$), but no such effect when considering those assigned to the anti-RCV condition ($F(1, 175) = 2.211$, $p = 0.139$, $\eta^2 = .01$). Thus, while the interaction effect was non-significant, we conclude that H2 is supported because introducing a sycophantic tone amplified experimental effects for those assigned to the pro-RCV condition.

Our results suggest that H3 and H4 are not supported, while H5 is partially supported. We hypothesized that prior attitudes towards the machine heuristic (H3) and pre-existing trust in AI (H4) would moderate the effect of sycophancy. We conducted separate linear regressions to test these hypotheses; full regression results are available in Tables 1-4. The tables show that the interaction between the machine heuristic and sycophancy and the interaction between trust in AI and sycophancy were not significant, indicating that H3 and H4 are not supported. Additionally, supplemental exploratory analysis revealed significant interaction effects between bias and trust and bias and the machine heuristic. The interaction between bias and the machine heuristic was positive, indicating that the machine heuristic moderates the effect of bias. Contrary to expectations, the interaction between bias and trust was negative, implying that those with higher levels of self-reported trust in AI were less likely to be influenced by AI bias. Furthermore, we identified significant, negative interactions between prior exposure to RCV and both bias (H5a) and sycophancy (H5b), suggesting that experimental effects are stronger in attitude formation contexts as hypothesized. However, only the interaction between sycophancy and prior

knowledge of RCV was significant (H5a); we did not find an interaction effect between bias and prior knowledge (H5b).

<Table 1 Here>

<Table 2 Here>

<Table 3 Here>

<Table 4 Here>

We also investigated whether post-test trust in AI mediates the effects of bias and sycophancy (RQ1). We built two separate models using Hayes (2022) PROCESS model 4, with 5,000 bootstrap samples and 95% confidence intervals. The first model treated sycophancy as the mediated variable and bias as a covariate; the second model treated bias as the mediated variable and sycophancy as a covariate. Post-test trust in AI was highly negatively correlated with bias; that is, those in the pro-RCV condition reported lower levels of trust in AI after interacting with the chatbot. However, AI trust did not mediate the effect of either bias or sycophancy. Full results are shown in Table 5.

<Table 5 Here>

We conducted additional linear regression models with demographic variables as moderators to address whether age, political party/lean, or education moderate experimental effects (RQ2). Full results are presented in Tables 6-8. Our analysis did not reveal any moderating effects for age. Participants were asked to identify their age as belonging to one of seven groups, which we then coded with values ranging from 1-7, with one representing the youngest group. We found that age category was negatively correlated with support for RCV

overall ($r=-0.0065$), however the interactions between age and bias and age and sycophancy were not statistically significant.

While political party alignment was a significant predictor of support for RCV, it did not moderate experimental effects. Among those who answered the questions about political party or political lean, we found that Democrats were more likely to support RCV compared to Republicans (the reference category). There were no significant moderating effects, however.

Furthermore, we found that education did not moderate the effect of bias or sycophancy. Participants were asked to provide their highest level of education, which was then coded on a scale from 1 (less than a high school education) to 7 (completed graduate or professional degree). No significant effects were identified for education level on its own, as a moderator for bias, or as a moderator for sycophancy.

<Table 6 Here>

<Table 7 Here>

<Table 8 Here>

To probe for possible mechanisms behind the observed experimental effects, we also investigated whether information bias ratings correlate with support for RCV (RQ3) and whether chat engagement differed across conditions (RQ4). Independent samples t-tests did not reveal any significant differences between chat duration in minutes ($p=.1064$) and number of AI interactions ($p=.3281$) between the pro-RCV and anti-RCV conditions. Similarly, no significant effect was identified between the neutral and sycophantic conditions for chat duration ($p=.4817$) and number of AI interactions ($p=.2781$). However, we identified a significant, positive correlation between perceived information bias ratings and support for RCV ($r = 0.556$, $p < 0.0001$). We also tested differences in perceived AI bias by sycophancy. Among those in the

anti-RCV condition, there was a significant difference ($p < 0.001$) in perceived bias ratings between the low-sycophancy ($M = 1.61$, $SD = 0.83$) and high-sycophancy ($M = 2.12$, $SD = 1.07$) groups. However, no such difference was observed for those assigned to the pro-RCV group. To probe whether perceived bias mediates the observed experimental effects, we again used Hayes (2022) PROCESS model 4. The findings suggest that perceived bias mediated the effect of bias, but not sycophancy. Full results from the mediation are displayed in Table 9.

<Table 9 here>

Finally, we explored the moderating effects of political attention and political interest (RQ5). We found a significant, negative interaction between bias and political attention, however no moderating effect was found for sycophancy. Participants with higher levels of political interest were more likely to support RCV overall, however political interest did not moderate the effects of sycophancy or bias. Full results are provided in Tables 10 and 11.

<Table 10 here>

<Table 11 here>

Discussion

The experiment, which tested the effects of sycophancy and information bias on attitudes towards a lesser-known public policy in the context of AI-supported search, provided several noteworthy results. Comporting with prior research on the Search Engine Manipulation Effect (SEME: Epstein & Robertson, 2015), we similarly find that participants in the pro-RCV content condition reported more favorable attitudes toward RCV than those in the anti-RCV condition (H1). Results suggest a significant and strong effect, with information bias explaining 37% of the variance in attitudes. Some of the shift we see may be due to an underlying attitude that favors

RCV, but additional research would be needed to better understand to what extent and what drivers may explain this beyond interacting with the chatbot.

The results regarding sycophancy indicate that its effects are more muted in the context of AI search. While we found that sycophancy had a significant independent effect, the effect size was substantially smaller (H2), explaining only 2% of the variance in attitudes. We interpret this to mean that the tone of the AI in the context of information seeking can affect attitudes. This underscores the observation from the HAIL-TIME model of the importance of message features in the communication process with AI (Sundar, 2020).

The moderator analyses predicted that the machine heuristic and AI trust would amplify the sycophancy effect (H3, H4), but the results did not bear that out. Instead, the results suggest that the machine heuristic amplified the information bias effect, while AI trust dampened the information bias effect. That is, people who already felt that AI was more objective and secure were more susceptible to biased search results, while those who trusted AI were less susceptible to biased search results. Both groups, however, were not more susceptible to flattery. Results suggest, though, that those with prior perceived exposure and knowledge of RCV were *less likely* to be persuaded by the information bias from the AI (H5). These findings comport with the Elaboration Likelihood Model (ELM), suggesting that counter-attitudinal information activates reactance (Brehm & Brehm, 1981), thereby reducing attitude shifts along the lines argued by the AI.

Regarding our mediation analysis of perceptions of trust in AI and its effect on bias and sycophancy (RQ1), the results may be counterintuitive. We find that those who were assigned to the pro-RCV condition trusted the AI *less* after the interaction when compared to those assigned to the anti-RCV condition. However, those assigned to the pro-RCV group were also more likely

to report favorable attitudes towards RCV. We might expect that attitude formation with assistance from the AI would *increase* trust. This result may signal that participants were aware of the persuasive manipulation by the AI and thus reported feeling less trust, especially in the pro-RCV condition. Indeed, the finding that people who perceived the chatbot as biased in favor of RCV were more likely to support RCV (RQ3) seems to suggest that was the case. If this finding holds in future studies, it means that even if people have some conscious awareness of the bias, it still shapes their attitudes. Yet, more research is needed to better understand the dynamics of supportive versus oppositional information on public policy matters. Because this study was a between-subjects experiment and did not measure prior attitudes on the policy topic—only perceived awareness and knowledge—the design is unable to determine to what extent prior attitudes further moderate the effects of information bias.

Regarding demographics and partisanship, we found no moderating effects on the experimental manipulations (RQ2). The results suggest that information bias has strong effects regardless of ideological predispositions, age, or educational level. The policy topic we chose of RCV may cut across ideological and demographic lines and may not activate partisan issue ownership heuristics found in other policy issues (Petrocik, 1996). This may further explain the stronger attitude shifts observed in the pro-RCV condition.

Although partisanship did not have effects, political attention did (RQ5). The results suggest that how closely people report following current events affects how much the AI's biased content moved their attitudes. Those who pay more attention are less susceptible to biased search results shaping their attitudes. This aligns with the Elaboration Likelihood Model that theorizes that people with more information are able to counter-argue or resist biased content. Those with high political interest, however, are not protected from such effects. They may have

more defined attitudes but it does not necessarily enable them to resist bias as observed in the SEME.

We also found no meaningful effects on attitudes with longer as compared to shorter interactions with the AI (RQ4). Thus, more interactions did not have an additive effect on attitudes. Here, though, more research is needed to examine the nature of the queries. For example, did participants who pushed back or challenged the AI have different attitudinal effects than those who did not?

The design of the study has limitations. First, the experimental design constrains some of the conclusions. We only compared biased content and its effects on attitudes. The study did not have a neutral condition where the AI search return delivered both favorable and oppositional positions on the policy. The lack of such a condition means that we do not have a control condition to compare the relative effects of biased AI as compared with neutral AI on attitudes. Moreover, we cannot see if the effects on attitudes are greater when the AI is supportive of a policy position versus oppositional on that policy position. The between-subjects design means we cannot explore whether prior attitudes further moderate the effects of biased content from the AI. Second, we note that the small sample size may limit our ability to draw conclusions about moderating effects. While we did not find evidence that the machine heuristic or prior trust in AI moderated the effects of sycophancy (H3 and H4) and demographics did not moderate any experimental effects (RQ2), further research is needed to fully explore the relationship between these measures and attitudinal alignment with AI. Finally, this experiment does not compare the potential effects of non-AI supported search (i.e. traditional search) with AI-supported search. It may be that the effects identified in our experiment are simply an extension of SEME; that is, search results in themselves may have effects on attitudes regardless of the communicative

message features as postulated in the HAIL-TIME model. Subsequent research should compare AI-supported search and traditional search to ascertain what, if any, effects AI interactivity has on attitudes.

Conclusion

The Search Engine Manipulation Effect (SEME) established that biased search returns have effects on attitudes in more traditional key-word searches (Epstein & Li, 2015), raising the concern that people with less knowledge and weak attitudes could be manipulated simply by the search returns. With the rise of interactive, AI-supported search engines, questions emerge regarding the effects of these search chatbots on the attitudes people form on the topics they search. In the context of public policy topics, where the public may not have strongly held attitudes, there is the risk that such AI can meaningfully affect attitudes. The results of our experiment suggest there is continued reason for concern. When the source results are biased for or against a policy position, there is a risk that AI-supported search will return results with summaries that ultimately make persuasive arguments about that policy position that in turn shapes the attitudes of those information seekers. While the message feature of sycophancy also has an effect, the magnitude of that effect is much smaller. Nevertheless, it still highlights the importance of message features and the design choices that AI engineers and the corporations behind them make in the search experience. Ultimately, the public can make good decisions when given quality information. AI has the potential to support that process, but how it is designed and the underlying information that shapes AI answers matters substantially.

References

- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183–189. <https://doi.org/10.1016/j.chb.2018.03.051>
- Bolino, M., Long, D., & Turnley, W. (2016). Impression Management in Organizations: Critical Questions, Answers, and Areas for Future Research. *Annual Review of Organizational Psychology and Organizational Behavior*, 3(1), 377–406. <https://doi.org/10.1146/annurev-orgpsych-041015-062337>
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. Academic Press.
- Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2025). *Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence* (arXiv:2510.01395). arXiv. <https://doi.org/10.48550/arXiv.2510.01395>
- Cheng, M., Yu, S., Lee, C., Khadpe, P., Ibrahim, L., & Jurafsky, D. (2025). *Social Sycophancy: A Broader Understanding of LLM Sycophancy*. arXiv. <https://doi.org/10.48550/ARXIV.2505.13995>
- Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, Cloud Research, Qualtrics, and SONA. *PLOS ONE*, 18(3), e0279720. <https://doi.org/10.1371/journal.pone.0279720>
- Epstein, R., & Li, J. (2024). Can biased search results change people’s opinions about anything at all? A close replication of the Search Engine Manipulation Effect (SEME). *PLOS ONE*, 19(3), e0300727. <https://doi.org/10.1371/journal.pone.0300727>

- Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33). <https://doi.org/10.1073/pnas.1419828112>
- Habermas, J. (1981). *The theory of communicative action: 1. Reason and the rationalization of society* (T. McCarthy, Tran.). New York, NY: Polity Press.
- Halavais, A. (2013). *Search Engine Society*. New York, NY: Wiley.
- Hindman, M. S. (2009). *The myth of digital democracy*. Princeton University Press.
- Ischen, C., Araujo, T., Van Noort, G., Voorveld, H., & Smit, E. (2020). “I Am Here to Assist You Today”: The Role of Entity, Interactivity and Experiential Perceptions in Chatbot Persuasion. *Journal of Broadcasting & Electronic Media*, 64(4), 615–639. <https://doi.org/10.1080/08838151.2020.1834297>
- Kim, J. K., LePine, J. A., & Chun, J. U. (2018). Stuck between a rock and a hard place: Contrasting upward and downward effects of leaders’ ingratiation. *Personnel Psychology*, 71(4), 495–518. <https://doi.org/10.1111/peps.12267>
- Kimball, D., Anthony, J., Scott, J., & Santucci, J. (2021). Support for Ranked Choice Voting and Partisanship of Voters: Results from a National Survey Experiment. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4015904>
- Lodge, M., & Taber, C. S. (2013). *The rationalizing voter*. Cambridge University Press.
- Merritt, S. M. (2011). Affective Processes in Human–Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(4), 356–370. <https://doi.org/10.1177/0018720811411912>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.

- Oh, J., & Sundar, S. S. (2015). How Does Interactivity Persuade? An Experimental Test of Interactivity on Cognitive Absorption, Elaboration, and Attitudes: Persuasive Effects of Interactivity. *Journal of Communication*, 65(2), 213–236.
<https://doi.org/10.1111/jcom.12147>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.
<https://doi.org/10.1016/j.jbef.2017.12.004>
- Perez, E., Ringer, S., Lukošiušė, K., Nguyen, K., Chen, E., Heiner, S., Pettit, C., Olsson, C., Kundu, S., Kadavath, S., Jones, A., Chen, A., Mann, B., Israel, B., Seethor, B., McKinnon, C., Olah, C., Yan, D., Amodei, D., ... Kaplan, J. (2022). *Discovering Language Model Behaviors with Model-Written Evaluations*. arXiv.
<https://doi.org/10.48550/ARXIV.2212.09251>
- Petrocik, J. R. (1996). Issue Ownership in Presidential Elections, with a 1980 Case Study. *American Journal of Political Science*, 40(3), 825. <https://doi.org/10.2307/2111797>
- Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, 46(1), 69–81. <https://doi.org/10.1037/0022-3514.46.1.69>
- Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41(5), 847–855. <https://doi.org/10.1037/0022-3514.41.5.847>
- Pezenka, I., Aunimo, L., Janous, G., & Dobrowsky, D. (2024). Emotionality in Task-Oriented Chatbots – The Effect of Emotion Expression on Chatbot Perception. *Communication Studies*, 75(6), 825–843. <https://doi.org/10.1080/10510974.2024.2363259>

- Rafaeli, S. (1988). Interactivity: From new media to communication. In R. P. Hawkins, J. M. Wiemann, & S. Pingree (Eds.), *Advancing communication science: Merging mass and interpersonal processes* (pp.110-134). Sage Publications, Inc.
- Reeves, B., & Nass, C. I. (2003). *The media equation: How people treat computers, television, and new media like real people and places*. CSLI Publishers.
- Rousseau, D. M., Sitkin, Sim B., Burt, Ronald S., & Camerer, Colin. (1998). Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*, 23(3), 393-404.
- Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37(3), 875–900. <https://doi.org/10.1080/07421222.2020.1790204>
- Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askeel, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2025). *Towards Understanding Sycophancy in Language Models* (arXiv:2310.13548). arXiv. <https://doi.org/10.48550/arXiv.2310.13548>
- Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17. <https://doi.org/10.1145/3613904.3642459>
- Stromer-Galley, J. (2004). Interactivity-as-Product and Interactivity-as-Process. *The Information Society*, 20(5), 391–394. <https://doi.org/10.1080/01972240490508081>
- Sun, Y., & Wang, T. (2025). *Be Friendly, Not Friends: How LLM Sycophancy Shapes User Trust* (arXiv:2502.10844). arXiv. <https://doi.org/10.48550/arXiv.2502.10844>

- Sundar, S. S. (2007). *Social psychology of interactivity in human-website interaction*. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips, (Eds.), *Oxford Handbook of Internet Psychology* (pp. 89-102). New York, NY: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199561803.013.0007>
- Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74–88. <https://doi.org/10.1093/jcmc/zmz026>
- Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H.-S. (2016). Theoretical Importance of Contingency in Human-Computer Interaction: Effects of Message Interactivity on User Engagement. *Communication Research*, 43(5), 595–625.
<https://doi.org/10.1177/0093650214534962>
- Sundar, S. S., Jia, H., Waddel, F. T., & Huang. (2015). Toward a theory of interactive media effects (TIME). In S. S. Sundar (Ed.), *Handbook of the Psychology of Communication Technology* (pp. 47–86). John Wiley & Sons, Inc.
- Sundar, S. S., & Kim, J. (2019, May 2). Machine Heuristic: When We Trust Computers More than Humans with Our Personal Information. CHI '19: CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland UK.
<https://doi.org/10.1145/3290605.3300768>
- Trielli, D., & Diakopoulos, N. (2019). Search as News Curator: The Role of Google in Shaping Attention to News Information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3290605.3300683>
- Tripodi, F. B. (2022). *The propagandists' playbook: How conservative elites manipulate search and threaten democracy*. New Haven, CT: Yale University Press.

Van Lange, P. A. M., Vugt, M. V., Meertens, R. M., & Ruiter, R. A. C. (1998). A Social Dilemma Analysis of Commuting Preferences: The Roles of Social Value Orientation and Trust. *Journal of Applied Social Psychology*, 28(9), 796–820.

<https://doi.org/10.1111/j.1559-1816.1998.tb01732.x>

Yang, H., & Sundar, S. S. (2024). Machine heuristic: Concept explication and development of a measurement scale. *Journal of Computer-Mediated Communication*, 29(6), zmae019.

<https://doi.org/10.1093/jcmc/zmae019>

Tables

Table 1

Moderation Analysis of Machine Heuristic

Predictor	b		sr ²	sr ²	
	b	95% CI [LL, UL]		95% CI [LL, UL]	Fit
(Intercept)	2.57**	[2.40, 2.73]			
sycophancy	0.24*	[0.04, 0.45]	.01	[-.01, .03]	
bias	1.46**	[1.26, 1.67]	.37	[.28, .45]	
mh	-0.18	[-0.40, 0.03]	.01	[-.01, .02]	
sycophancy:mh	-0.04	[-0.31, 0.23]	.00	[-.00, .00]	
bias:mh	0.36**	[0.09, 0.62]	.01	[-.01, .03]	

R² = .387**

95% CI[.30,.45]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. *sr*² represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 2*Moderation Analysis of AI Trust*

	b			sr ²	
Predictor	b	95% CI	sr ²	95% CI	Fit
		[LL, UL]		[LL, UL]	
(Intercept)	2.56**	[2.39, 2.73]			
sycophancy	0.25*	[0.05, 0.46]	.01	[-.01, .03]	
bias	1.46**	[1.26, 1.66]	.36	[-.28, .45]	
trust	0.06	[-0.11, 0.23]	.00	[-.00, .01]	
sycophancy:trust	0.12	[-0.08, 0.32]	.00	[-.01, .01]	
bias:trust	-0.24*	[-0.44, -0.05]	.01	[-.01, .03]	
					R ² = .387**
					95% CI [.30, .45]

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 3*Moderation of RCV Exposure*

	b		sr ²		
Predictor	b	95% CI [LL, UL]	sr ²	95% CI [LL, UL]	Fit
(Intercept)	1.68**	[1.30, 2.06]			
sycophancy	0.81**	[0.34, 1.27]	.02	[-.00, .04]	
bias	1.99**	[1.53, 2.45]	.12	[.07, .18]	
exposure	0.36**	[0.22, 0.49]	.05	[.01, .08]	
sycophancy:exposure	-0.22*	[-0.39, -0.04]	.01	[-.01, .03]	
bias:exposure	-0.21*	[-0.39, -0.04]	.01	[-.01, .03]	
					R ² = .421**
					95% CI[.34,.48]

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 4*Moderation of RCV Knowledge*

	b		sr ²		
Predictor	b	95% CI	sr ²	95% CI	Fit
		[LL, UL]		[LL, UL]	
(Intercept)	1.80**	[1.46, 2.14]			
sycophancy	0.74**	[0.30, 1.18]	.02	[-.00, .04]	
bias	1.79**	[1.36, 2.23]	.11	[.06, .17]	
knowledge	0.35**	[0.22, 0.49]	.04	[.01, .08]	
sycophancy:knowledge	-0.23*	[-0.42, -0.05]	.01	[-.01, .03]	
bias:knowledge	-0.15	[-0.33, 0.03]	.00	[-.01, .02]	
					R ² = .423**
					95% CI[.34,.48]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. *sr*² represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 5*Mediation Analysis of Post-Test Trust in AI*

Mediated Variable	Mediator	Indirect Effect Bootstrap Estimate	p	Indirect Effect 95% CI	
				LLCI	ULCI
Bias	Post-Test Trust	-0.0213	0.3391	-0.0876	0.0264
Sycophancy	Post-Test Trust	-0.0013	0.8767	-0.0221	0.0178

Table 6*Age*

Predictor	<i>b</i>		<i>sr</i> ²	<i>sr</i> ²	
	<i>b</i>	95% CI		95% CI	Fit
		[LL, UL]		[LL, UL]	
(Intercept)	2.91**	[2.48, 3.33]			
sycophancy	-0.04	[-0.58, 0.49]	.00	[-.00, .00]	
bias	1.61**	[1.07, 2.16]	.06	[.02, .10]	
age	-0.11	[-0.24, 0.02]	.01	[-.01, .02]	
sycophancy:age	0.09	[-0.07, 0.26]	.00	[-.01, .01]	
bias:age	-0.04	[-0.20, 0.12]	.00	[-.00, .00]	

$$R^2 = .384^{**}$$

95% CI[.30,.45]

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 7*Education*

	b		sr ²		
Predictor	b	95% CI	sr ²	95% CI	Fit
		[LL, UL]		[LL, UL]	
(Intercept)	2.42**	[1.88, 2.95]			
sycophancy	-0.21	[-0.81, 0.40]	.00	[-.00, .01]	
bias	1.92**	[1.31, 2.53]	.07	[.03, .12]	
educ	0.03	[-0.07, 0.14]	.00	[-.00, .01]	
sycophancy:educ	0.10	[-0.02, 0.21]	.00	[-.01, .02]	
bias:educ	-0.10	[-0.22, 0.02]	.01	[-.01, .02]	
					R ² = .384**
					95% CI [.30, .45]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. *sr*² represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 8*Political Party*

	b		sr ²		
Predictor	b	95% CI	sr ²	95% CI	Fit
		[LL, UL]		[LL, UL]	
(Intercept)	2.04**	[1.76, 2.33]			
sycophancy	0.38*	[0.04, 0.71]	.01	[-.01, .02]	
bias	1.56**	[1.23, 1.90]	.15	[.09, .21]	
democrat	0.83**	[0.48, 1.18]	.04	[.01, .07]	
sycophancy:democrat	-0.14	[-0.56, 0.28]	.00	[-.00, .01]	
bias:democrat	-0.13	[-0.54, 0.29]	.00	[-.00, .00]	
					R ² = .470**
					95% CI[.38,.53]

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 9*Mediation Analysis of Perceived Information Bias Ratings*

Mediated Variable	Mediator	Indirect Effect Bootstrap Estimate	p	Indirect Effect 95% CI	
				LLCI	ULCI
Bias	Information Bias Rating	0.3426	0.0209	0.0090	0.6555
Sycophancy	Information Bias Rating	-0.0213	0.1034	-0.0011	0.4288

Table 10*Political Interest*

Predictor	<i>b</i>		<i>sr</i> ²	<i>sr</i> ²	
	<i>b</i>	95% CI [LL, UL]		95% CI [LL, UL]	Fit
(Intercept)	1.50**	[0.89, 2.11]			
sycophancy	0.85*	[0.13, 1.56]	.01	[-.01, .03]	
bias	1.95**	[1.23, 2.66]	.05	[.01, .09]	
interest	0.35**	[0.16, 0.55]	.02	[-.00, .05]	
sycophancy:interest	-0.19	[-0.42, 0.04]	.00	[-.01, .02]	
bias:interest	-0.17	[-0.40, 0.06]	.00	[-.01, .01]	

$$R^2 = .400^{**}$$

$$95\% \text{ CI} [.31, .46]$$

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 11*Political Attention*

	b		sr ²		
Predictor	b	95% CI	sr ²	95% CI	Fit
		[LL, UL]		[LL, UL]	
(Intercept)	1.43**	[0.83, 2.03]			
sycophancy	0.90*	[0.12, 1.67]	.01	[-.01, .03]	
bias	2.37**	[1.60, 3.14]	.07	[-.02, .11]	
attention	0.24**	[0.12, 0.35]	.03	[-.00, .05]	
sycophancy:attention	-0.13	[-0.29, 0.02]	.01	[-.01, .02]	
bias:attention	-0.19*	[-0.34, -0.04]	.01	[-.01, .03]	

R² = .401**95%
CI[.32,.46]

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.