

Cognitive Processing of Ai-Generated Advice: The Role of Belief Congruence

Dorit Nevo

nevod@rpi.edu

Rensselaer Polytechnic Institute

Benjamin D Horne

University of Tennessee at Knoxville

Jianguo He

University of Tennessee at Knoxville

Article

Keywords: Large language models, belief congruence, heuristic-systematic model, fact-checking, human-AI interaction

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11021363/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Large language models (LLMs) increasingly provide advice to users on a range of topics. Yet little is known about how users cognitively process the advice and rely on it. Drawing on the Heuristic-Systematic Model (HSM), we specifically study how person-AI belief congruence influences both processing and acceptance of LLM-generated fact-checking advice. In an online study, U.S. adults assessed political news headlines before and after interacting with left-leaning and right-leaning LLM fact-checkers. Political congruence with the AI was associated with less systematic processing. Allowing users to select their AI advisor up front similarly reduced systematic processing relative to randomly assigned advisors. The LLM's fact-checking verdict also shifted subsequent trust judgments in the direction of its recommendation, even after accounting for users' initial judgments and political congruence. These findings identify person-AI belief congruence as an important heuristic cue. The results highlight a potential tension for personalized AI systems: features that increase perceived fit may simultaneously reduce critical scrutiny of AI-generated advice.

1 INTRODUCTION

Advances in AI's ability to generate new content rather than only predict labels or numbers are expected to transform how humans interact with AI on knowledge tasks (Jo 2023; Epstein et al. 2022). As AI becomes an interactive partner in sensemaking and decision-making, understanding how users cognitively engage with AI-generated advice has become increasingly important (Xu 2019). One factor that has received relatively little attention is whether perceived belief congruence between the user and the AI shapes how that advice is evaluated and accepted. Individuals are known to interpret and evaluate new information through the lens of their existing attitudes and beliefs (Lord et al. 1979; Nickerson 1998). Although advice-taking research acknowledges that individuals anchor on their initial judgments (Yaniv 2004), it offers limited insight into how perceived alignment with an AI advisor shapes the cognitive processing and acceptance of its advice.

Research has identified several factors that shape whether people incorporate external advice into their judgments. Individuals may discount advice because they overweight their own expertise (Bonaccio and Dalal 2006; Yaniv 2004), while advice is more likely to be accepted when the advisor is perceived as an expert or displays high confidence (Bonaccio and Dalal 2006). Advice taking also depends on the distance between the advice and the recipient's initial judgment (Yaniv 2004). Thus, existing research has established that characteristics of the recipient, the advisor, and the advice itself shape advice acceptance. Less understood is whether the perceived alignment between recipient and advisor, specifically, whether the advisor appears to share the recipient's beliefs, shapes how closely that advice is scrutinized.

This limitation is especially salient in the context of AI-generated advice because AI systems can be configured to appear more or less aligned with the beliefs of their users. Therefore, we need to expand

our knowledge on whether advice from a seemingly congruent AI advisor is evaluated differently from advice provided by a seemingly incongruent advisor.

To explain how person-AI belief congruence shapes engagement with AI-generated advice, we draw on the Heuristic-Systematic Model (HSM; Chaiken 1980). HSM distinguishes between two modes of information processing. Systematic processing entails careful, effortful evaluation of the content and quality of an argument. Heuristic processing relies on readily available cues, such as a communicator's expertise, confidence, group membership, message length, social endorsement, or characteristics of the communication medium (Horne et al. 2020). HSM is particularly useful for studying the effect of belief congruence on AI advice acceptance because it allows for heuristic and systematic processing to occur concurrently, rather than representing mutually exclusive routes (Chen and Chaiken 1996). The model suggests that people seek to minimize cognitive effort while attaining sufficient confidence in their judgments (Chaiken and Ledgerwood 2012). Whether that confidence threshold is reached, and therefore whether further scrutiny seems necessary, may depend on cues about the source of the advice.

LLM-based advice makes this allowance especially important because it typically provides two layers of information: an immediately usable recommendation and an explanation supporting that recommendation. Users may directly rely on the model's verdict, they may invest effort in evaluating the explanation, or they may combine both forms of processing. Further, because the explanation is presented alongside the recommendation and can be extended through conversational follow-ups, the user does not necessarily need to leave the information environment to engage in greater elaboration. LLMs therefore increase the opportunity for systematic processing while continuing to provide prominent heuristic cues.

A politically congruent AI may lead users to infer that the advisor is credible, trustworthy, or likely to share their understanding of the issue, thereby reducing the perceived need to examine its reasoning systematically. Alternatively, congruence may shape how users interpret the explanation even when systematic processing occurs.

Processing route, in HSM, is distinct from advice influence. Both heuristic and systematic processing can lead to persuasion and belief updating (Chaiken 1980). In other words, a user may accept the AI's advice by deferring to its verdict, or they may arrive at the same revised judgment after evaluating its explanation. Similarly, systematic processing does not guarantee acceptance; careful evaluation may instead lead a user to reject the recommendation (Vasconcelos et al. 2023). We therefore examine advice acceptance as a separate downstream outcome. This distinction allows us to study both how users process AI-generated advice and whether the advice ultimately changes their judgments.

In sum, this paper examines the effect of belief congruence on two related aspects of engagement with AI advice: how users process the advice is the first; and whether the advice influences their subsequent judgment is the second. Examining both aspects is necessary because reliance observed in a final judgment does not reveal the cognitive process through which that judgment was formed. We address

two interrelated questions. First, does perceived congruence between the user and the AI advisor specifically function as a heuristic cue that reduces systematic engagement with the AI's explanation? And second, after accounting for users' prior beliefs and their congruence with the AI advisor, does the AI's recommendation shift their subsequent judgments? As an extension, we also examine whether processing differs when users select their AI advisor rather than interacting with assigned advisors. Our hypotheses follow a two-stage account of AI advice-taking. H_1 and H_2 establish the conventional ability and motivation conditions associated with systematic processing. H_3 and H_4 introduce our focal extension: characteristics that signal fit between the user and the AI advisor may reduce the perceived need to scrutinize the AI's explanation. H_5 then establishes the consequential second stage of the advice-taking episode: whether the AI's recommendation produces belief updating.

We examine these questions in the context of AI-assisted fact-checking of news. This context provides a theoretically meaningful setting because assessing the veracity of news requires individuals to reconcile external evidence with prior beliefs, making it well suited to studying belief congruence, information processing, and belief updating. AI systems are often perceived as "black boxes" that reflect the values, assumptions, and classification choices embedded in their design and training (Faraj et al. 2018). These perceptions may undermine assessments of integrity and transparency, both of which are central to trust in AI (Glikson and Woolley 2020). In our studies, we vary the chatbot's stated political orientation and the news sources available to it. This design allows us to examine whether person-AI political congruence operates as a heuristic cue and whether selecting a seemingly aligned AI advisor changes users' reliance on, and processing of, its recommendations.

The paper makes two principal contributions to research on AI-assisted decision-making. First, it shifts attention from whether individuals rely on AI-generated advice to how they process that advice. Specifically, we distinguish between reliance on an LLM's verdict and systematic engagement with the explanation supporting it. Although LLMs make elaboration readily available, users still decide how much cognitive effort to invest in evaluating that information. Drawing on HSM, we conceptualize this processing decision as distinct from the subsequent influence of the AI's advice on users' judgments. Second, we theorize person-AI political congruence and advisor choice as cues that may shape how much additional scrutiny users believe is necessary. This extends attention beyond the content of AI-generated advice to the perceived relationship between the user and the AI advisor. At the center of both contributions is person-AI belief congruence, which we conceptualize as a relational heuristic: users respond not only to what an AI says, but also to whether the AI appears to see the world as they do.

We examine these arguments in an online study in which participants assess the veracity of news headlines before and after receiving advice from custom-built, purportedly left-leaning or right-leaning LLM fact-checkers. Part 1 varies exposure to congruent and incongruent AI advisors within participants, whereas Part 2 allows participants to select an advisor before receiving its recommendations. We examine how person-AI political congruence and advisor selection shape the processing of AI-generated fact-checking advice and whether the LLM's recommendations ultimately shift users' judgments.

2 LLM SUPPORTED NEWS FACT-CHECKING

Warning labels from fact-checkers are a widely studied intervention in the news consumption space that reduces people's belief in, trust in, and sharing intentions of misinformation (Martel and Rand 2024). Evidence suggests that people are particularly susceptible to misinformation that reinforces their existing political positions (Frenda et al. 2013; Xiong et al. 2023). At the same time, content warning labels from fact-checking are shown to be generally effective across a variety of individual differences, including partisanship, albeit to varying degrees (e.g. Horne 2025; Yaqub et al. 2020).

While effective, fact-checking is also resource-intensive and requires extensive human labor. Consequently, large language models (LLMs), which are trained on massive datasets and can augment training data through web search, are increasingly being used for fact-checking (Renault et al. 2025) and studies have shown LLMs can perform well in misinformation detection tasks (Hoes et al. 2023; Zhou et al. 2024; Kuznetsova et al. 2025; Huang et al. 2025).

LLM-based fact-checkers, whether they are explicitly built for fact-checking tasks or not, differ from traditional fact-checking in ways that may affect how users respond to their corrections. Just as LLMs represent a “significant shift from traditional search engines” (Narayanan Venkit et al. 2025), they also represent a significant shift from traditional fact-checking. LLMs can be configured to be ideologically aligned with their users via the content emphasized in their responses, the sources cited, and their configured persona (Kröger and Barkett 2025). The most prominent example of an ideologically configured LLM is the chatbot Grok, which has been explicitly configured to be “not woke” and provide politically conservative responses (Zeff 2023; Melimopoulos 2025). Preliminary work suggests that Republican X/Twitter users tend to select Grok over the more neutral Perplexity for fact-checks (Renault et al. 2025). As a result, corrections from LLM fact-checkers may be influenced by partisan preferences, which raises critical questions about whether and how the effectiveness of fact-checking extends to corrections delivered by congruent or incongruent LLMs.

AI recommendations, in general, may encourage cognitive offloading and reduce effortful evaluation, contributing to overreliance even when those recommendations are incorrect (Buçinca et al. 2021). At the same time, users may underutilize useful AI advice, with reliance varying across individuals and situations (Horne and Nevo 2025; Klingbeil et al. 2024; Passi and Vorvoreanu 2022; Vasconcelos et al. 2023). Research on algorithmic aversion and algorithmic appreciation helps explain users' general orientation toward algorithmic advice. Algorithmic aversion refers to reluctance to follow an algorithm's recommendation, particularly after observing an error (Dietvorst et al. 2015; Jussupow et al. 2024), whereas algorithmic appreciation refers to a preference for algorithmic rather than human judgment (Logg et al. 2019). These perspectives help explain whether people are inclined to rely on AI, but reliance alone does not reveal the full story. A user may rely on an AI-generated recommendation by deferring to its verdict or by carefully evaluating the accompanying explanation. Thus, the question remains about how AI advice is cognitively processed and whether it ultimately produces belief updating.

LLM fact-checkers alter the information-processing conditions under which veracity judgments are formed. To explain these changes, we draw on the Heuristic–Systematic Model (HSM; Chaiken 1980). HSM proposes that the extent of heuristic and systematic processing depends on individuals’ effort considerations and confidence needs (Chaiken and Ledgerwood 2012). Importantly, the effort required for systematic processing is not fixed; it depends on the information environment. Traditional fact-checking by human experts often presents users with a relatively simple cue, such as a true/false verdict or a verified/unverified warning. Although users can investigate the basis for that judgment, doing so may require them to leave the platform, consult additional sources, and reconcile competing evidence. LLM-based fact-checking reduces these barriers by providing not only a veracity judgment but also explanations, evidence, and opportunities for interactive follow-up within the same information environment. It therefore makes elaboration more readily available while preserving an immediately usable heuristic cue.

Studies applying HSM to traditional fact-checking suggest that users often treat warning labels as endpoints for judgment rather than as prompts for further systematic evaluation (Koch et al. 2023; Horne and Nevo 2025). LLM-based fact-checking creates a different processing environment because the heuristic cue and the information needed for systematic evaluation are presented together. Users may defer to the LLM’s verdict, scrutinize the explanation supporting it, move between these forms of processing through follow-up interactions, or rely on heuristic cues while simultaneously interpreting the substantive information provided. Greater access to explanation therefore does not necessarily produce greater systematic processing. Instead, LLMs expand the processing options available to users and make it especially important to distinguish the influence of AI advice from the cognitive route through which that influence occurs.

In sum, LLM-based fact-checking changes the informational and effort conditions under which veracity judgments are formed. By lowering the barriers to systematic processing while retaining salient heuristic cues, LLMs create greater opportunities for systematic engagement without ensuring that users will take advantage of them. This makes LLM fact-checking a theoretically important extension of HSM. Building on this application of HSM, we next derive hypotheses about how ability, accuracy motivation, and person-AI political congruence shape users’ processing of LLM-generated fact-checking advice; we subsequently consider advisor choice as an extension of this framework.

2.1 Hypothesis development

As previously discussed, HSM explains information processing in decision-making through heuristic and systematic modes. The heuristic mode relies on mental shortcuts to form judgments, whereas systematic processing involves deliberate and careful evaluation of message content (Chaiken 1980). Judgments formed heuristically are generally less stable and less predictive of subsequent behavior than those formed through systematic processing (Trumbo 2002).

To explain the extent of information processing in which individuals engage, the model proposes the least effort principle and the sufficiency threshold. The least effort principle assumes that individuals strive to minimize cognitive effort and therefore default to less effortful processing when possible (Chaiken and Ledgerwood 2012). When a discrepancy arises between actual and desired confidence, individuals may increase cognitive effort to reduce this gap (Todorov and Chaiken 2002). The larger the confidence gap, the more likely individuals are to engage in systematic processing, trading off effort for confidence (Chen et al. 1996). The sufficiency threshold represents the point at which effort and confidence are balanced (Chen et al. 1999).

The model further identifies three motivational orientations that influence processing. Accuracy motivation reflects the desire to hold correct beliefs (Chen et al. 1999). Defense motivation refers to the desire to maintain beliefs consistent with one's existing attitudes, whereas impression motivation reflects the desire to form beliefs that align with social expectations (Chen et al. 1996).

In addition to motivation, HSM posits that cognitive ability determines the extent to which individuals can engage in systematic processing. Even when individuals are motivated to form accurate judgments, systematic processing requires sufficient cognitive resources and task-relevant knowledge (Chaiken 1980; Chen and Chaiken 1999). Primary cognitive determinants of ability include time availability and prior knowledge about the topic (Eagly and Chaiken 1993).

When either motivation or ability is insufficient, individuals are more likely to rely on heuristic cues. When both are high, systematic processing is more likely to occur, although heuristic cues may still be at play. Importantly, HSM does not conceptualize heuristic and systematic processing as mutually exclusive. The two modes can operate concurrently (Chen and Chaiken 1999). Heuristic cues may serve as simple judgmental shortcuts when motivation or ability is low, but they can also bias or guide the interpretation of information when individuals engage in more elaborate evaluation. For example, source credibility may function as a heuristic cue in one instance, yet in another it may influence the interpretation of ambiguous arguments, thereby shaping the outcome of systematic processing.

What is unique about the LLM fact-checking environment is that the bot does not provide only a single judgment, such as true or false, but also provides an explanation for how that judgment was formed. The verdict remains readily available as a heuristic cue, while the accompanying explanation provides an opportunity for systematic processing at a reduced information-access cost. Users can therefore decide how much of the available explanation they wish to evaluate and they can follow up with clarification questions. HSM suggests that this decision will depend, in part, on their ability to process the information and their motivation to form an accurate judgment. Individuals with greater ability to evaluate the LLM's response should be better able to engage with its explanation, assess the evidence provided, and integrate that information into their own judgment. We therefore hypothesize:

H₁

Greater ability to process the LLM's response will increase systematic processing of the fact-checking information.

Accuracy motivation should similarly increase users' willingness to invest effort in evaluating the explanation. When the desire to hold a correct belief is high, relying on the verdict alone may be insufficient to reach the user's desired level of confidence. Users should therefore engage in greater systematic processing of the information supporting the LLM's recommendation.

H₂

Greater motivation to hold correct beliefs will increase systematic processing of the fact-checking information.

In addition to ability and accuracy motivation, the belief congruence between the user and the AI advisor may shape how much processing is considered sufficient. In related work, Ma et al. (2024) studied how confirmation bias, which they measured as agreement between a person and a headline's political leaning, can serve as an important heuristic cue en route to engagement with news stories. Focusing on the relationship between the person and the AI, a meta-analysis by Hart et al. (2009) finds that people typically select congenial sources when given the choice and that this selection can satisfy their information-processing motivation.

Using similar reasoning, we propose that political congruence between the person and the LLM fact-checking bot can alter the processing route. When a bot appears politically aligned with the user, the user may infer that it is more credible, trustworthy, or likely to evaluate the issue through a compatible perspective. This perceived fit may help the user reach a sufficient level of confidence without closely evaluating the explanation. A congruent bot may increase reliance on the readily available verdict and reduce the extent of systematic processing. At the same time, political congruence may also impact systematic processing by shaping how the user interprets and understands the explanation. Our directional expectation, however, is that congruence will reduce the perceived need for systematic processing:

H₃

People will engage in less systematic processing when interacting with politically congruent LLMs.

Finally, returning to the advice-taking literature, Bonaccio and Dalal (2006) note that advisor utilization resembles the agency problem in that information asymmetries exist between decision makers and advisors. In such contexts, trust and perceived alignment play central roles in determining the extent to which advice is incorporated into final judgments. Similarly, research on algorithm aversion shows that when algorithms are imposed, users may remain skeptical and engage in more critical evaluation of their advice. In contrast, providing individuals with even minimal control over algorithm selection and modification increases their willingness to use it and their reliance on its outputs (Dietvorst et al. 2018).

Allowing users to pre-select an LLM fact-checker may therefore reduce resistance and increase acceptance of its judgments. From an HSM perspective, the choice itself can function as a heuristic cue: this is the bot the user selected and, by implication, the bot they considered appropriate or trustworthy. Source credibility is one of the most commonly discussed heuristic cues in the dual-processing literature (Eagly and Chaiken 1993), and AI recommendation labels have similarly been shown to act as social influence signals that reduce effortful processing (Ma et al. 2025). Selecting the LLM may therefore reduce the perceived need to scrutinize its explanation and increase reliance on its readily available judgment.

H₄

The ability to pre-select a fact-checker will reduce systematic processing of the LLM's fact-checking information.

Our first four hypotheses concern how users process the LLM fact-checker's response (systematically or heuristically). Processing mode, however, does not uniquely determine whether its advice will be accepted. Both heuristic and systematic processing can produce persuasion: users may defer directly to the LLM's verdict or revise their judgment after evaluating its explanation. Conversely, systematic evaluation may lead users to reject the advice. We therefore treat the influence of the LLM's recommendation on users' judgments as a distinct outcome rather than as a direct implication of a particular processing mode.

Prior evidence concerning this influence is mixed. LLMs have successfully persuaded users in contexts such as conspiracy beliefs (Costello et al. 2024), science skepticism (Hornsey et al. 2025), and voter attitudes (Lin et al. 2025). In fact-checking contexts, however, DeVerna et al. (2024) found no significant relationship between exposure to ChatGPT-generated fact-checks and users' ability to discern headline accuracy. Other studies suggest that LLM interventions may also produce unintended or asymmetric effects. Costello et al. (2026), for example, found that interactions with LLMs can sometimes increase conspiracy beliefs, while Ma et al. (2024) found that labeling news as recommended by AI reduced belief in true news without producing a corresponding reduction in belief in false news. Together, these findings indicate that the influence of LLM advice depends on the context and may not consistently operate in the intended direction.

We therefore examine whether the LLM fact-checker's recommendation shifts users' judgments in the present context. When the LLM judges a headline to be true, users should increase their trust in the headline; when it judges a headline to be false, users should decrease their trust. Given our broader interest in belief congruence, we also examine whether this directional influence is evident across levels of political congruence.

H₅

Users' trust in a headline will shift in the direction of the LLM fact-checker's recommendation, increasing when the LLM judges the headline to be true and decreasing when it judges the headline to be false.

H₅ completes the advice-taking episode by extending the analysis from the processing of AI advice to its observable influence on judgment. It does not propose that either processing mode necessarily produces greater belief updating. Rather, H₁ through H₄ explain the conditions under which users scrutinize the AI's reasoning, and H₅ establishes whether the recommendation enters their subsequent judgment. Person-AI congruence connects these stages as a focal characteristic of the advisor-user relationship: it shapes processing and is retained when examining belief updating.

3 Results

We collected data to examine how belief congruence shapes individuals' processing of AI-generated fact-checking advice. The overall task had participants evaluate the veracity of political news headlines themselves before interacting with custom-built LLM chatbots that were configured to exhibit either left-leaning or right-leaning perspectives. The study first employed a within-subjects design in which participants fact-checked headlines using both chatbots in counterbalanced order. We then introduced choice by allowing participants to pre-select which chatbot they wished to use rather than randomly interacting with both chatbots. The selection of headlines, study design, and measurements are described in depth in the Methods section and the study flow is shown in Fig. 1 below.

We required the chatbots used as fact-checkers in this study to use at least one fact-checking website when generating their responses. Consequently, our chatbots performed well against the headlines' ground truth. Across both chatbots, when the headline's ground truth was False, the bot replied 'False' 89.2% of the time, 'Unverifiable' 3.3% of the time, and 'True' 7.5% of the time. When the headline's ground truth was True, the bot replied 'True' 86.6% of the time, 'Unverifiable' 6.9% of the time, and 'False' 6.6% of the time.

3.1 Variables

Processing mode. The dependent variable used to test H₁ through H₄ was "processing mode," formed from both subjective and behavioral indicators. In line with HSM, systematic processing is conceptualized as more effortful, whereas heuristic processing relies on faster, lower-effort judgments (Chaiken 1980; Chen and Chaiken 1999). Respondents reported their processing style on a five-point heuristic-systematic scale. We also captured the total length (in characters) of all prompts entered for each headline and the total duration (in seconds) that respondents spent on each headline before moving on to the next one. Recall that the study design is such that a headline is displayed first, then the user can interact with the fact-checking bot through interactive prompting before they make their final veracity judgment. The `prompt_length` and the `headline_duration` are both objective measures of each person-bot interaction for each headline. Because both prompt length and duration were right-skewed, we applied a natural-log transformation before standardizing and averaging the three indicators for a

single DV measure. Higher values of 'processing mode' reflect more systematic processing and lower values reflect more heuristic processing.

Belief updating. To test H_5 , the dependent variable was post-interaction trust in the headline's veracity, measured after participants received the LLM's fact-checking advice using the same five-point scale from 'completely distrust' to 'completely trust.' Initial trust was measured before the participant interacted with the LLM and was included in the model as a covariate. The model therefore estimates whether the LLM's verdict changed participants' final trust in the headline after accounting for their initial judgment; the dependent variable is the post-interaction trust rather than a raw change score.

Independent variables. The independent variables in our model include ability, accuracy motivation, confirmation, and person-AI political congruence. *Ability* was operationalized as the respondent's ability to form judgment on each headline. It was measured using the initial trust rating provided by the respondent for each headline. Specifically, we converted the five-point trust scale to a 3-point confidence scale such that a response of 3 ('neither trust nor distrust') was converted to 0, responses of either 2 or 4 ('somewhat distrust/trust') were converted to 1, and responses of either 1 or 5 ('strongly distrust/trust') were converted to 2. We interpret weak ability to assess the veracity of a news headline as indicating lower task-specific ability (i.e., lower confidence and/or knowledge) in the headline's domain. In contrast, if a person can form a strong opinion on a specific headline, we assume that they also possess the ability to systematically process (i.e., read and assess) the fact-checking information.

Accuracy motivation was operationalized using the "concern for accuracy" scale, which is a part of the MiSS index (Kumar et al. 2025). The misinformation susceptibility (MiSS) index consists of three conceptual parts: concern for accuracy, distrust in experts, and epistemic vigilance. The "concern for accuracy" scale includes the following items:

1. I rarely fact-check information to see if it is true or false (rev)
2. I rarely care whether information is true or false (rev)
3. When I look at news or information I focus more on if it's interesting than if it's true (rev)
4. I'm more concerned with whether I agree with the sentiment than with its accuracy (rev)
5. Once my mind is made up, I tend to be closed to new information
6. When looking at information online, I look out for unusual formatting, like spelling mistakes or strange layouts, to judge if it's credible

After conducting exploratory factor analysis, only the first four items were retained for this study.

Confirmation (political congruence between the person and the headline) and *bot congruence* (political congruence between the person and the bot) were both operationalized similarly to Ma et al. (2024) by multiplying either the headline's leaning (right/left) for confirmation, or the bot's leaning (right/left) for bot congruence, by the participant's political leaning. Headline and bot leaning were converted into - 1 (right) or + 1 (left). Respondents' political leaning was self-reported on a 7-point scale with 1 being very

conservative and 7 very liberal. We then converted it to a -3 to + 3 scale to compute the confirmation and bot congruence measures.

Finally, for the model used to test H_5 we also included the *initial 'trust in headline veracity'* score, which was measured on a five-point scale from 'completely distrust' to 'completely trust', as well as the 'bot's first word', which was limited by our prompt design to be either True, False, or Unverifiable.

3.2 Processing Mode

Because our observations were cross-classified within participants and headlines (each participant evaluated eight headlines, and each headline was evaluated by multiple participants) we estimated a linear mixed-effects regression model with random intercepts for participant (PID) and headline (HID) using SAS PROC MIXED. The dependent variable was an index of processing mode as described above.

Table 1 presents the results. Ability was positively associated with systematic processing ($b = 0.037$, $p < .001$), supporting H_1 . Accuracy motivation was also positively associated with systematic processing ($b = 0.117$, $p < .001$), supporting H_2 . Person-AI political congruence was negatively associated with systematic processing ($b = -0.010$, $p = .017$), supporting H_3 . Thus, greater political alignment between the participant and the AI advisor was associated with less systematic engagement, holding the other predictors constant. The advisor-choice extension was also associated with processing mode. Participants in the assigned-advisor part of the study (the first part) demonstrated greater systematic processing than participants who pre-selected the bot they wished to work with ($b = 0.118$, $p = .007$), supporting H_4 .

The random effects are also shown in Table 1 along with their 95% confidence intervals. The random effect controlling for person level differences was significant, suggesting that the average level of systematic versus heuristic processing differs significantly from person to person. Some people naturally engage in more systematic processing and some in more heuristic processing, regardless of the fixed effects. The random effect for headline was significant but substantially smaller than participant-level variance, indicating limited between-headline variability relative to between-participant variability. Intraclass correlation estimates from the null model indicated that approximately 61% of total variance was attributable to random effects, primarily at the participant level. Finally, the inclusion of fixed effects significantly improved model fit compared to an intercept-only null model ($\Delta AIC = 52.1$).

Table 1
Linear mixed-effects model predicting systematic processing

Predictor	Estimate	SE	t	p
Intercept	-0.772	0.120	-6.42	< .001
Ability	0.037	0.011	3.51	< .001
Accuracy motivation	0.117	0.020	5.79	< .001
Person-AI political congruence	-0.010	0.004	-2.40	.017
Bot choice ¹	0.118	0.044	2.71	.007
Person ID	0.285	[0.255...0.321]		
Headline ID	0.005	[0.003...0.009]		

¹ Binary variable, choice was the reference category.

3.3 Trust in Headlines

The preceding analysis examined how participants processed the LLM's advice. We next examine the second aspect of advice-taking: whether the LLM's recommendation influenced participants' subsequent judgments. Because heuristic and systematic processing can both produce persuasion, processing mode and belief updating are treated as related but distinct outcomes.

We estimated a linear mixed-effects model predicting participants' post-interaction trust in each headline. The model included the LLM's initial verdict, i.e., "True," "False," or "Unverifiable", as the primary predictor. Initial trust in the headline, headline-person political congruence (confirmation), and person-AI political congruence were included as covariates. Random intercepts were included for participant and headline.

As shown in Model 1 of Table 2, the LLM's verdict strongly predicted participants' subsequent judgments. Relative to an "Unverifiable" verdict, a "True" verdict increased trust in the headline ($b = 1.275$, $p < .001$), whereas a "False" verdict decreased trust ($b = -0.837$, $p < .001$). These findings support H_5 : participants updated their judgments in the direction of the LLM's recommendation. It is worth noting that an 'Unverifiable' judgment was more closely perceived as false than as true.

Initial trust was also positively associated with post-interaction trust ($b = 0.284$, $p < .001$). Confirmation had a smaller positive association with post-interaction trust ($b = 0.023$, $p < .001$). Person-AI political congruence was not significantly associated with post-interaction trust in Model 1 ($b = 0.007$, $p = .382$).

To explore the nuances of how the bot impacts trust in the headline, we examined how political congruence may influence the results through an interaction effect. In Model 2, we added interactions between the LLM's verdict and person-AI political congruence. Adding these interactions improved model fit relative to Model 1. However, the interaction comparing a "False" verdict with an "Unverifiable"

verdict was not individually significant ($b = -0.050$, $p = .119$). Subsequently, Model 3 excluded observations in which the LLM described the headline as “Unverifiable.” The interaction between person-AI political congruence and the LLM’s verdict was significant ($b = -0.055$, $p < .001$). Specifically, the association between person-AI congruence and post-interaction trust was weaker when the LLM judged the headline as false than when it judged the headline as true. Importantly, this interaction qualifies but does not overturn the central H_5 finding: the LLM’s verdict strongly shifted participants’ judgments in the direction of its recommendation.

Table 2
Linear mixed-effects models predicting post-interaction trust in headline veracity

Predictor	Model 1 Estimate (SE)	Model 2 Estimate (SE)	Model 3 Estimate (SE)
Intercept	1.880*** (0.077)	1.879*** (0.077)	3.176*** (0.051)
Initial trust	0.284*** (0.012)	0.282*** (0.012)	0.283*** (0.012)
Person-AI political congruence	0.007 (0.008)	0.028 (0.031)	0.033** (0.011)
Confirmation	0.023*** (0.007)	0.022** (0.007)	0.022** (0.007)
Bot’s First Word: False	−0.837*** (0.070)	−0.824*** (0.071)	−2.150*** (0.043)
Bot’s First Word: True	1.275*** (0.066)	1.271*** (0.067)	Reference
Bot’s First Word: Unverifiable	Reference	Reference	—
Congruence × False	—	−0.050 (0.032)	−0.055*** (0.014)
Congruence × True	—	0.005 (0.032)	Reference

The inclusion of fixed effects significantly improved model fit compared to the intercept-only null model ($\Delta AIC = 1,586.2$). For the null model, the variance was partitioned between the headlines (HID) and the participants (PID). The Intraclass Correlation Coefficient (ICC) for Headlines (HID) was approximately 56.5%, indicating that more than half of the total variance in trust was due to differences between the headlines themselves. The ICC for Participants (PID) was comparatively small ($\sim 3.2\%$). Together, these results indicate that trust judgments were strongly shaped by headline-level characteristics.

4 DISCUSSION

This paper examines how belief congruence shapes the processing of AI fact-checking advice. Drawing on the Heuristic-Systematic Model (HSM), we investigated whether individuals process the fact-checking advice heuristically or systematically, and how ability, accuracy motivation, and political congruence influence that processing. Our results show evidence that belief congruence impacts how AI advice is processed and integrated. Adding a model to link processing to persuasion, our combined findings

describe both the process and the consequence of engaging with AI advice. Ability, accuracy motivation, political congruence, and advisor choice help explain how users engage with the recommendation, while the second analysis shows that the same AI interaction can meaningfully shift subsequent trust judgments. These results should not be interpreted as a test of mediation from processing mode to belief updating; rather, they show why both insights are needed for a fuller account of AI advice-taking.

A central starting point for this study was that the question is not simply *whether* individuals rely on AI advice, but *how* they process it. While prior research on algorithm aversion and appreciation has focused primarily on reliance rates or comparative preferences between humans and algorithms (e.g. Dietvorst et al., 2015; Logg et al., 2019; Jussupow et al., 2024), our findings add to this literature by studying AI advice acceptance using the information-processing lens. Consistent with the HSM theoretical foundation, we find that individuals engaged in more systematic processing when both ability and accuracy motivation were higher. This suggests that individuals do not uniformly default to heuristic reliance on AI. This finding holds true even in politically charged domains such as the veracity of political news headlines.

Adding to the HSM, our results reveal that political congruence between the bot and the person can function as an important heuristic cue. When individuals interact with a politically aligned chatbot, they rely more heavily on heuristic processing and engage in less systematic evaluation. This finding extends prior work on confirmation bias and congenial source selection (Hart et al., 2009; Ma et al., 2024) by demonstrating that belief alignment affects not only source selection but also downstream cognitive processing. Future research can focus on other measures of similarity, beyond belief alignment, and how those affect trade-offs between processing modes. Further, HSM is often measured as a snapshot, but LLM chats are sequential. Future work can model processing as evolving across rounds; it may be that early heuristic reliance turns systematic through user follow-up prompts, or vice-versa.

Our results further demonstrate that the ability to pre-select the chatbot also increases heuristic processing. Allowing individuals to choose their fact-checker reduced systematic engagement relative to random assignment. This finding contributes to the growing literature on algorithm control and modification (Dietvorst et al., 2018) by suggesting that even minimal agency may shift processing mode, not just usage rates. In HSM terms, selection itself may operate as a heuristic (“this is the system I trust”), lowering the need to scrutinize reasoning. Future research can focus on how perceived ownership or endorsement of the information sources used by AI alters cognitive processing independent of message content.

This result further has important implications for personalized AI systems. Customizable or ideologically configurable LLMs may increase user comfort and engagement. At the same time, it is important to remember that such selection may simultaneously reduce critical evaluation, and, depending on the context, customization may reduce the accuracy of AI advice. For example, while both of the chatbots used in our study were highly accurate due to the requirement of using third-party fact-checkers, the accuracy also differed across the left and right bots due to the news sources each could use to generate

content. Specifically, the right bot was slightly less accurate than the left bot across both parts of the study. The left bot labeled false headlines correctly between 90% and 91% of the time and labeled true headlines correctly between 88% and 91% of the time. The right bot labeled false headlines correctly between 85% and 91% of the time and labeled true headlines correctly between 80% and 86% of the time. Future work can further examine the extent to which LLM chatbots can be customized without sacrificing the accuracy of their advice or the critical engagement of their users.

Finally, in testing hypothesis 5 we establish that LLM fact-checking significantly shifts trust in headlines, increasing trust in true content and decreasing trust in false content. This is noteworthy given mixed prior findings regarding AI-generated corrections (DeVerna et al., 2024; Ma et al., 2024). In our setting, LLM verdict labels produced robust directional belief change, even after controlling for prior trust. At the same time, confirmation bias remained significant in headline trust, and interaction effects suggest that congruence between the chatbot and person moderates the magnitude of trust updating. Future research can further explore the extent to which LLM advice acceptance is sensitive to motivated reasoning dynamics in contexts other than fact-checking news.

To summarize, this study makes three important contributions. First, it advances research on algorithmic decision support by focusing on how individuals cognitively process AI-generated advice. By integrating HSM, we demonstrate that advice acceptance unfolds through distinct processing routes. Individuals may accept AI advice heuristically (deferring to verdict labels) or systematically (scrutinizing explanations). This distinction clarifies why AI reliance may appear inconsistent across contexts: similar reliance rates may mask different cognitive processing routes. We further adapt HSM to a new technological environment characterized by interactive explanations and configurable personas. While HSM has been widely applied to persuasion and communication, its application to AI advice remains underdeveloped. Our findings support the basic tenets of the HSM, that ability and motivation increase systematic processing, into the realm of GenAI outputs. Second, our study further identifies an important heuristic in the form of political congruence. While much has been studied on the related concept of confirmation bias, political congruence between people and GenAI has not been studied. Finally, we show that AI selection shifts processing toward heuristic reliance. This result contributes to emerging debates about ideologically configurable AI systems and suggests that politically aligned LLMs may increase comfort and possibly engagement, but at the cost of reduced critical information processing. Selection and personalization, often viewed as empowering, may unintentionally promote heuristic reliance and introduce a tension between user satisfaction and epistemic vigilance in AI system design.

This study is not without limitations. We operationalized ability through the respondents' confidence in their initial veracity judgment. In doing so, we follow a reasoning that prefers revealed ability over self-reported ability. Future research can use different measures of ability for triangulation. We also focused in this study on accuracy motivation only. Future work can also explore our findings when respondents' motivation is defense or impression. A potential statistical concern is the large sample size, which could theoretically lead to "overly significant" p-values for negligible effects. However, several factors suggest the observed relationships are substantively meaningful rather than artifacts of statistical power. First,

the parameter estimates remained highly consistent across both the PROC MIXED and PROC REG models. Second, the t-values for primary predictors were high, exceeding the threshold for marginal significance even in large samples.

Conclusion

This study integrates the Heuristic-Systematic Model with the emerging literature on algorithmic reliance. We find that ability and accuracy motivation predict deeper, more systematic engagement with AI explanations, while political congruence and the opportunity to choose an AI advisor shift processing toward heuristic reliance. At the same time, the fact-checker's verdict meaningfully updates trust in true versus false headlines, even after accounting for prior beliefs and belief congruence. Together, these findings highlight a central design tension for configurable and personalized AI systems: the same features that increase comfort and perceived fit may also reduce systematic processing and critical thinking. Understanding and shaping the processing route through which users engage AI advice will be critical for realizing the benefits of LLM decision support without inadvertently weakening epistemic vigilance.

Method

We collected data to examine how belief congruence shapes individuals' processing of AI-generated fact-checking advice. The overall task had participants evaluate the veracity of political news headlines themselves before interacting with custom-built LLM chatbots that were configured to exhibit either left-leaning or right-leaning perspectives. The study first employed a within-subjects design in which participants fact-checked headlines using both chatbots in counterbalanced order. We then introduced choice by allowing participants to pre-select which chatbot they wished to use rather than randomly interacting with both chatbots.

5.1 LLM Chatbot Design and Interface

The fact-checking chatbot used in this study was a web-based conversational AI system. The system was designed to assist participants in evaluating the veracity of news headlines by providing fact-checking capabilities with real-time search, where the participants could ask unlimited questions about a given headline. The chatbot used GPT 5.1 with medium reasoning effort and was integrated with the Exa API to perform real-time web searches as an LLM tool. Web searches could return up to 25 results per query, and the LLM could submit multiple queries in response to a single participant's question prompt. Those web results were restricted to articles published after January 1, 2025, on U.S.-based websites. Search results from Exa API included LLM-generated article summaries, publication dates, and source URLs.

The chatbot supported our experimental conditions via media source filtering and system prompt framing. Specifically, each version of the chatbot was configured to restrict search results to specific

news domains based on its political leaning. The full list of news domains that each chatbot was restricted to can be found in the Supplementary Information. Further, the chatbot's system prompt was configured to reflect different political perspectives, instructing the model to convey principles typically associated with either left-leaning or right-leaning viewpoints. The prompt framing was adapted from existing empirical studies on political bias of LLMs (Agiza et al. 2024; Chen et al. 2024; Smith-Vaniz et al. 2025). To ensure the LLM features were consistent across conditions, the chatbot followed a structured response protocol, whereby the chatbot had to: (1) provide an initial verdict label (True, False, or Unverifiable), (2) deliver concise responses under 300 words, (3) cite at least three sources from the configured media group, (4) include one to two fact-checking sources (i.e., Snopes), and (5) pose follow-up questions to encourage continued engagement. Our goal with this design was to provide as realistic a chatbot setup as possible while also having some control of the possible outputs from the chatbot. A complete system prompt can be found in the Supplementary Information.

The interface presented headlines in a card format displaying the headline text, an associated image, and a 5-point trust rating scale ranging from "completely distrust" to "completely trust." Below the headline, a chat panel allowed participants to interact with the chatbot through a text input field (see Fig. 2). Chatbot responses were streamed in real-time, with a collapsible "Reasoning Process" panel displaying ongoing search queries and sources being consulted. Inline citations appeared as color-coded badges indicating the political orientation of each source (blue for left-leaning, red for right-leaning), with hover cards providing article titles, publication dates, and content summaries. Participants were required to complete at least one conversational exchange before submitting their trust rating for each headline.

5.2 Experimental Setup

We collected 412 responses from U.S. adults through the survey platform Prolific. The sample combined respondents who were drawn from a U.S.-representative Prolific sample stratified by age, gender, and political affiliation ($n = 294$) as well as participants who were recruited using a politically balanced sampling frame ($n = 118$; Republicans, Independents, and Democrats). This sampling mix ensured sufficient variation and statistical power for analyses involving political congruence and partisan subgroup comparison, and balanced external validity and statistical power.

Each participant was asked to review a random set of eight political news headlines, balanced across ground truth (true/false) and political leaning (left/right). For each headline, participants were asked to indicate how much they trusted the information on a 5-point scale (from "completely distrust" to "completely trust"). After rating all eight headlines, participants were told that they would next use two different chatbots for fact-checking (assigned in counterbalanced order). They were also told when the chatbots switched. For each chatbot, they were shown a small tutorial of the interface and a list of example news outlets that each LLM would use to fact-check. For the left bot, these example news outlets were The Guardian, The Huffington Post, MSNBC, The Daily Beast, and Mother Jones. For the right bot, these example news outlets were The American Conservative, Breitbart, Fox News, The Daily

Caller, and Newsmax. Our pre-testing indicated that participants correctly identified the general leaning of each bot.

Participants were then asked to prompt the fact-checking bot at least once and review the bot's answer before again rating their trust in each headline. While participants were required to complete at least one round of conversation with the chatbot for each headline, they were free to complete as many rounds of conversation as they wanted before submitting their new trust rating.

After each chatbot session (for each headline), participants were also asked how they used the chatbot to judge their trust in each headline (self-reported heuristic-systematic processing). Possible answers to this question ranged from heuristic to systematic reasoning: 'Directly relied on the chatbot's true/false assessment, without evaluating its explanation', 'Relied mostly on the chatbot's true/false assessment, with minimal consideration of the explanation', 'Considered the chatbot's explanation but largely deferred to its overall judgment', 'Actively evaluated the chatbot's explanation and used it to inform my own judgment', 'Fully reflected on the chatbot's assessment and explanation and integrated them with my own knowledge and reasoning about the topic' or 'Completely ignored the chatbot'. Before analysis, we excluded four respondents who reported completely ignoring the chatbot for all eight headlines. We also excluded 13 respondents who failed the attention checks, leaving a final sample of 395 respondents. At the response level, we removed 55 observations in which respondents indicated that they had completely ignored the chatbot, because this response option was not included in the dependent-variable measure. These observations were removed individually rather than by excluding the respondent; for example, if a respondent used the chatbot for seven of the eight headlines, only the observation for the headline on which they ignored it was removed. The final dataset contained 3,105 observations. Finally, after completing the fact-checking task across both chatbots, all participants were redirected to a post-experiment survey to capture their political leaning and accuracy motivation.

Following this first stage of the study, we recruited an additional 293 adults from a U.S. representative sample across age, gender, and political affiliation from the same survey platform. The experimental setup mimicked what was explained above with one difference: participants chose which chatbot to use and used the chosen chatbot to fact-check all eight headlines. In the choice menu, participants were given the same instructions and details about each chatbot's design via example news outlets. Before analysis, we excluded ten respondents who reported completely ignoring the chatbot's recommendation for all headlines. At the response level, we also removed 36 additional observations in which respondents selected this option. This final dataset contained 2,228 observations.

5.3 Headline Selection

To ensure that headlines were recent and balanced across political congruence, we followed a two-step process. First, following Horne (2025), we randomly selected political headlines that were published in the last six months. False headlines were selected from third-party fact-checking sites, such as Snopes, PolitiFact, FactCheck.org, or Reuters Fact Check. True headlines were selected from both recently fact-checked articles and articles from Reuters and the Associated Press (about half and half). This

approach allowed our true headlines to be represented by a mixture of typical factual content and true headlines that required fact-checking. From the selected articles, we used the headline and an image in the article to create the headline stimuli.

Second, we pre-tested all headlines in a separate sample of 356 participants who indicated “how good for the Democrats versus Republicans the headline would be if it was true” on a 5-point Likert scale from very favorable to very unfavorable (similar to the approach used by Chen et al. 2023). On average, each headline had 35 ratings, which we then used to compute an average political leaning of each headline. Finally, we created a balanced set of 132 headlines across veracity and political leaning by randomly sampling from each of the four groups: true right, true left, false right, and false left.

Declarations

Funding. No external funding was received for this manuscript.

Data Availability Statement. The data supporting the findings of this study is available here: <https://rpi.box.com/s/wq5y0sqqIk4o6pzm9npu2qfe4htsvqi5>

Competing Interests. The authors declare no competing interests.

Ethics statement. The study was reviewed and approved by the Institutional Review Board at Rensselaer Polytechnic Institute (protocol #2291). All research was performed in accordance with relevant IRB guidelines and regulations. Informed consent was obtained from all participants prior to participation.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with manuscript organization, language, clarity, and readability. The authors reviewed and edited all AI-assisted content as needed and take full responsibility for the content of the publication.

References

1. Agiza, A., Mostagir, M. & Reda, S. Politune: Analyzing the impact of data selection and fine-tuning on economic and political biases in large language models. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Vol. 7, No. 1, pp. 2–12). (2024), October.
2. Bonaccio, S. & Dalal, R. S. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organ. Behav. Hum Decis. Process.* **101** (2), 127–151 (2006).
3. Buçinca, Z., Malaya, M. B. & Gajos, K. Z. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction*, 5(CSCW1), 1–21. (2021).
4. Chaiken, S. & Ledgerwood, A. A theory of heuristic and systematic information processing. *Handb. Theor. social Psychol.* **1**, 246–266 (2012).

5. Chaiken, S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J. Personal. Soc. Psychol.* **39** (5), 752 (1980).
6. Chen, S. & Chaiken, S. The heuristic-systematic. *Dual-process theories in social psychology*, 73. (1999).
7. Chen, S., Shechter, D. & Chaiken, S. Getting at the truth or getting along: Accuracy-versus impression-motivated heuristic and systematic processing. *J. Personal. Soc. Psychol.* **71** (2), 262 (1996).
8. Chen, S., Duckworth, K. & Chaiken, S. Motivated heuristic and systematic processing. *Psychol. Inq.* **10** (1), 44–49 (1999).
9. Chen, X., Pennycook, G. & Rand, D. What makes news sharable on social media? *Journal Quant. Description: Digit. Media*, **3**. (2023).
10. Chen, K., He, Z., Yan, J., Shi, T. & Lerman, K. How susceptible are large language models to ideological manipulation? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17140–17161). (2024), November.
11. Costello, T. H., Pennycook, G. & Rand, D. G. Durably reducing conspiracy beliefs through dialogues with AI. *Science* **385** (6714), eadq1814 (2024).
12. Costello, T. H. et al. Large language models can effectively convince people to believe conspiracies. *arXiv preprint arXiv:2601.05050*. (2026).
13. DeVerna, M. R., Yan, H. Y., Yang, K. C. & Menczer, F. Fact-checking information from large language models can decrease headline discernment. *Proceedings of the National Academy of Sciences*, **121**(50), e2322823121. (2024).
14. Dietvorst, B. J., Simmons, J. P. & Massey, C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* **144** (1), 114 (2015).
15. Dietvorst, B. J., Simmons, J. P. & Massey, C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Manage. Sci.* **64** (3), 1155–1170 (2018).
16. Eagly, A. H. & Chaiken, S. *The psychology of attitudes* (Harcourt brace Jovanovich college, 1993).
17. Epstein, Z. et al. Do explanations increase the effectiveness of AI-crowd generated fake news warnings? In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 183–193). (2022), May.
18. Faraj, S., Pachidi, S. & Sayegh, K. Working and organizing in the age of the learning algorithm. *Inf. Organ.* **28** (1), 62–70 (2018).
19. Frenda, S. J., Knowles, E. D., Saletan, W. & Loftus, E. F. False memories of fabricated political events. *J. Exp. Soc. Psychol.* **49** (2), 280–286 (2013).
20. Glikson, E. & Woolley, A. W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **14** (2), 627–660 (2020).
21. Hart, W. et al. Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychol. Bull.* **135** (4), 555 (2009).

22. Hoes, E., Altay, S. & Bermeo, J. Leveraging ChatGPT for efficient fact-checking. *PsyArXiv April*, **3**(10.31234). (2023).
23. Horne, B. D. Does the source of a warning matter? Examining the effectiveness of veracity warning labels across warners. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 19, pp. 823–836). (2025), June.
24. Horne, B. D. & Nevo, D. People adhere to content warning labels even when they are wrong due to ecologically rational adaptations. *Sci. Rep.* **15** (1), 13896 (2025).
25. Horne, B. D., Nevo, D., Adali, S., Manikonda, L. & Arrington, C. Tailoring heuristics and timing AI interventions for supporting news veracity assessments. *Computers Hum. Behav. Rep.* **2**, 100043 (2020).
26. Hornsey, M. J., Smith, A. E., Pearson, S., Bretter, C. & Nylund, J. L. Using conversational AI to reduce science skepticism. *Current Opin. Psychology*, 102216. (2025).
27. Huang, T., Yi, J., Yu, P. & Xu, X. Unmasking digital falsehoods: A comparative analysis of LLM-based misinformation detection strategies. In *2025 8th international conference on advanced algorithms and control engineering (ICAACE)* (pp. 2470–2476). IEEE. (2025), March.
28. Jo, A. The promise and peril of generative AI. *Nature* **614** (1), 214–216 (2023).
29. Jussupow, E., Benbasat, I. & Heinzl, A. An integrative perspective on algorithm aversion and appreciation in decision-making. *MIS Q.* **48** (4), 1575–1590 (2024).
30. Klingbeil, A., Grützner, C. & Schreck, P. Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI. *Comput. Hum. Behav.* **160**, 108352 (2024).
31. Koch, T. K., Frischlich, L. & Lerner, E. Effects of fact-checking warning labels and social endorsement cues on climate change fake news credibility and engagement on social media. *J. Appl. Soc. Psychol.* **53** (6), 495–507 (2023).
32. Kröger, P. & Barkett, E. Don't Change My View: Ideological Bias Auditing in Large Language Models. *arXiv preprint arXiv:2509.12652*. (2025).
33. Kumar, S., Butler, L., DeGutis, J., Amazeen, M. & Swire-Thompson, B. Development and validation of the Misinformation Susceptibility Self-Report (MiSS). (2025).
34. Kuznetsova, E. et al. In generative AI we trust: can chatbots effectively verify political information? *J. Comput. Social Sci.* **8** (1), 15 (2025).
35. Lin, H. et al. Persuading voters using human–artificial intelligence dialogues. *Nature* **648** (8093), 394–401 (2025).
36. Logg, J. M., Minson, J. A. & Moore, D. A. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* **151**, 90–103 (2019).
37. Lord, C. G., Ross, L. & Lepper, M. R. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11), 2098. (1979).

38. Ma, H., Dennis, A. S., Dennis, A. R. & Huang, W. Artificial intelligence recommendations amplify the sharing of true and fake news on social media by appealing to fast cognition. *J. Manage. Inform. Syst.* **42** (4), 987–1016 (2025).
39. Ma, H., Huang, W. & Dennis, A. R. Unintended consequences of disclosing recommendations by artificial intelligence versus humans on true and fake news believability and engagement. *J. Manage. Inform. Syst.* **41** (3), 616–644 (2024).
40. Martel, C. & Rand, D. G. Fact-checker warning labels are effective even for those who distrust fact-checkers. *Nat. Hum. Behav.* **8** (10), 1957–1967 (2024).
41. Melimopoulos, E. *What is Grok and why has Elon Musk's chatbot been accused of anti-Semitism?* [Internet]. *Al Jazeera* 10 Jul 2025. (2025).
42. Narayanan Venkit, P., Laban, P., Zhou, Y., Mao, Y. & Wu, C. S. Search engines in the AI era: a qualitative understanding to the false promise of factual and verifiable source-cited responses in LLM-based search. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1325–1340). (2025), June.
43. Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2** (2), 175–220 (1998).
44. Passi, S. & Vorvoreanu, M. Overreliance on AI literature review. *Microsoft Res.* **339**, 340 (2022).
45. Renault, T., Mosleh, M. & Rand, D. *@ Grok is this true?* (LLM-powered fact-checking on social media, 2026).
46. Smith-Vaniz, N., Lyon, H., Steigner, L., Armstrong, B. & Mattei, N. Investigating political and demographic associations in large language models through moral foundations theory. In *Proceedings of the AAI/ACM Conference on AI, Ethics, and Society* (Vol. 8, No. 3, pp. 2419–2430). (2025), October.
47. Todorov, A., Chaiken, S. & Henderson, M. D. The heuristic-systematic model of social information processing. *persuasion handbook: Developments theory Pract.* **23**, 195–211 (2002).
48. Trumbo, C. W. Information processing and risk perception: An adaptation of the heuristic-systematic model. *J. communication.* **52** (2), 367–382 (2002).
49. Vasconcelos, H. et al. Explanations can reduce overreliance on ai systems during decision-making. *Proceedings of the ACM on human-computer interaction*, 7(CSCW1), 1–38. (2023).
50. Xiong, A., Lee, S., Seo, H. & Lee, D. Effects of associative inference on individuals' susceptibility to misinformation. *J. Experimental Psychology: Appl.* **29** (1), 1 (2023).
51. Xu, W. Toward human-centered AI: A perspective from human-computer interaction. *interactions* **26** (4), 42–46 (2019).
52. Yaniv, I. Receiving other people's advice: Influence and benefit. *Organ. Behav. Hum Decis. Process.* **93** (1), 1–13 (2004).
53. Yaqub, W., Kakhidze, O., Brockman, M. L., Memon, N. & Patil, S. Effects of credibility indicators on social media news sharing intent. In *Proceedings of the 2020 chi conference on human factors in*

computing systems (pp. 1–14). (2020), April.

54. Zeff, M. Enter 'Grok,'Elon Musk's Anti-Woke Chatbot. *Archived from the original on December, 2, 2023.* (2023).
55. Zhou, X., Sharma, A., Zhang, A. X. & Althoff, T. Correcting misinformation on social media with a large language model. *arXiv preprint arXiv:2403.11169.* (2024).

Figures

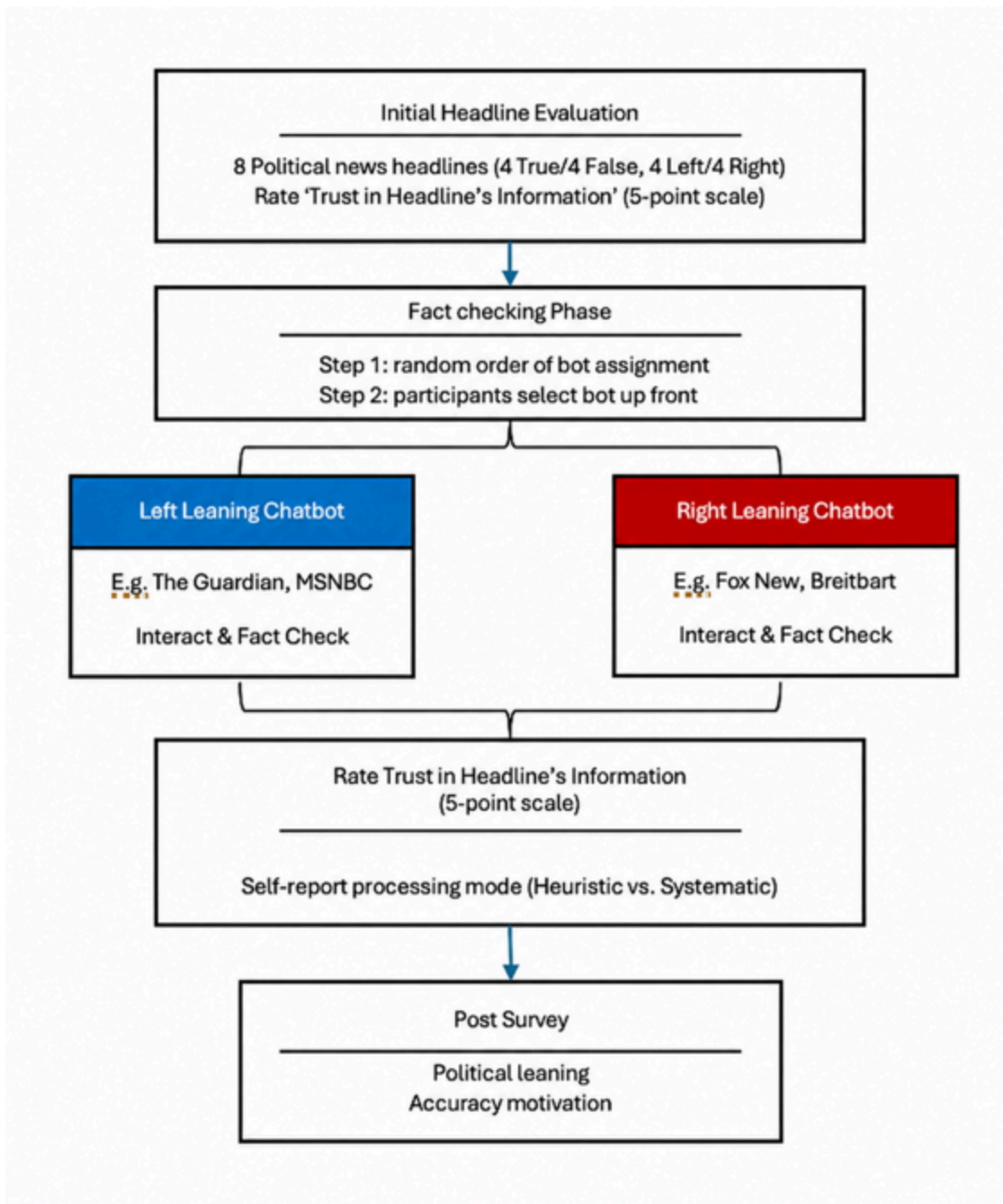


Figure 1

Flow chart of the study design

