

When AI Meets Conspiracy: Experimental Evidence on Internal Political Efficacy

Fernando García

Department of Political Science and Public Law, Universitat Autònoma de Barcelona

FernandoFrancisco.Garcia@uab.cat

Carol Galais

Department of Political Science and Public Law, Universitat Autònoma de Barcelona

Carolina.Galais@uab.cat

DRAFT — DO NOT CIRCULATE OR CITE

First draft prepared for conference presentation and discussion. Comments welcome — please contact the corresponding author before quoting or circulating.

Abstract

Contemporary digital information environments increasingly expose citizens to conspiratorial content and AI-generated political communication, yet little is known about how these cues shape citizens' subjective sense of political competence. We examine whether exposure to conspiratorial narratives and AI source cues affects internal political efficacy using a randomized survey experiment conducted among 1,500 adults in Spain. Respondents were randomly assigned to one of four conditions: a neutral placebo, a conspiratorial narrative, the same narrative presented as AI-generated, or a conspiracy portraying AI as a hidden conspiratorial actor. Contrary to our preregistered expectations, exposure to conspiratorial narratives produced small but positive effects on internal political efficacy, with the strongest and most consistent effects observed when the narrative was attributed to AI authorship. Exploratory mediation analysis shows that perceived credibility does not explain this effect: the AI-authored narrative was rated less credible than the placebo, yet still increased efficacy — a pattern consistent with suppression, though the indirect effect itself does not reach significance. These findings challenge the assumption that conspiracy narratives necessarily undermine internal political efficacy and point to AI source cues as an underexplored driver of citizens' subjective political competence.

When AI Meets Conspiracy: Experimental Evidence on Internal Political Efficacy

Introduction

Internal political efficacy — the belief that one is capable of understanding and participating effectively in politics — is a cornerstone of democratic citizenship. Individuals with higher internal political efficacy are consistently more likely to participate in both institutional and non-institutional forms of political engagement (Gil de Zúñiga et al., 2017; Oser et al., 2022), thereby contributing to democratic representation, institutional accountability, and the overall quality of democratic governance. Understanding what shapes citizens' sense of political competence is therefore essential for explaining and predicting the resilience of contemporary democracies.

Research has traditionally explained differences in internal political efficacy through sociodemographic characteristics and processes of political socialization and learning. More recently, attention has shifted towards the role of the digital information environment in shaping citizens' confidence in their political competence. Exposure to political information in online environments has generally been associated with higher levels of internal political efficacy (Hansen & Pedersen, 2014; Lee, 2006; Martin et al., 2018; Semetko & Valkenburg, 1998), particularly when information is delivered interactively and presented in ways that are easy to understand (Boulianne, Oser, & Hoffmann, 2023; Hart & Feldman, 2016; Tedesco, 2007). However, the rapid spread of conspiratorial narratives and, more recently, generative artificial intelligence (AI) has transformed how political information is produced, circulated, and interpreted, raising new questions about their consequences for democratic attitudes.

Conspiracy theories constitute one of the most visible features of this evolving information environment. Although they are widely regarded as a threat to liberal democracy — as illustrated by events such as the January 6 attack on the U.S. Capitol and the anti-democratic mobilizations targeting the German Parliament — their effects on internal political efficacy remain unclear. Some studies report that exposure to conspiracy theories reduces citizens' sense of political competence (Ardèvol-Abreu et al., 2020; Rullo & Telesca, 2023), while others find the opposite effect (Herold, 2024). Adding further complexity, the rapid diffusion of generative AI has introduced a new layer of uncertainty. While AI may enhance democratic inclusion by facilitating access to information (Tessler et al., 2024), its growing prominence may also reinforce perceptions that political outcomes are driven by opaque technological systems beyond citizens' understanding or control (Coeckelbergh, 2023). Moreover, AI and conspiracy theories are becoming increasingly intertwined: AI can generate conspiratorial content, become the object of conspiratorial beliefs, or be incorporated as an actor within conspiratorial narratives. Despite this growing convergence, little is known about how conspiratorial narratives and AI jointly shape citizens' political attitudes and democratic orientations.

Against this backdrop, we examine how exposure to conspiratorial narratives affects internal political efficacy and whether these effects depend on the role of AI within those narratives.

Specifically, we compare the effects of exposure to a traditional conspiracy narrative, an AI-generated conspiracy narrative, and a conspiracy portraying AI as a hidden conspiratorial actor. Based on the predominantly pessimistic expectations of the existing literature, we derive three hypotheses concerning the independent effects of conspiracy theories and the role assigned to AI within conspiratorial narratives. We test these hypotheses using a preregistered four-arm vignette experiment embedded in an online survey administered by Ipsos to approximately 1,500 Spanish adults residing in Spain in May 2026. Respondents were randomly assigned to one of four conditions: a neutral control, a traditional conspiratorial narrative, a conspiratorial narrative purportedly generated by AI, and a story portraying AI as a hidden conspiratorial actor. This design allows us to compare the effects of different configurations of conspiratorial framing and AI cues on internal political efficacy.

This study makes several contributions. First, it brings together two defining features of contemporary digital information environments — conspiratorial narratives and generative AI — which have rarely been examined within a common analytical framework, and even less in relation to citizens' subjective political competence. Second, it moves beyond the predominantly correlational literature on conspiracy beliefs by shifting the focus from endogenous conspiracy beliefs to experimentally manipulated exposure to conspiratorial narratives. By randomly assigning exposure, the study isolates the causal effect of conspiratorial communication on internal political efficacy, avoiding the endogeneity that characterizes much of the existing literature. Third, it contributes to the emerging literature on the democratic implications of generative AI by examining a fundamental attitudinal outcome for democratic citizenship: internal political efficacy. Whereas existing research has focused primarily on attitudes towards automated decision-making (Pickering et al., 2025; Zwald et al., 2025), much less is known about how AI functions as an informational source cue when embedded in everyday political communication, and how such cues shape citizens' confidence in their own political competence. By examining these processes jointly, the study contributes to a growing literature on how contemporary information environments shape democratic citizenship, and, in particular, citizens' subjective political competence.

1. Theoretical framework

1.1 Internal political efficacy: Political learning and information environments

Internal political efficacy refers to citizens' confidence in their ability to understand and engage effectively with politics. It is one of the most consistent predictors of political participation across institutional and non-institutional forms of engagement, including individual, collective, offline, and online political action (Boulianne, Oser, & Hoffmann, 2023; Gil de Zúñiga, Diehl, & Ardévol-Abreu, 2017; McDonnell, 2020; Oser et al., 2022; Prats & Meunier, 2021). Conversely, individuals with low internal efficacy are more likely to display political apathy and disaffection (Balch, 1974; Fierro, Aroca, & Navia, 2023; Morrell, 2005).

Rather than being a fixed individual disposition, however, internal political efficacy is increasingly understood as a malleable democratic attitude that develops through political learning and experience. Although education, income, and gender remain among its strongest and most stable predictors (Beaumont, 2011; Hayes & Bean, 1993; Oser, Feitosa, & Dassonneville, 2023; Wu, 2003), these structural inequalities do not fully determine citizens' confidence in their political competence. On the contrary, living in politically active communities, engaging in political discussion, interacting in pluralistic environments, and, especially, acquiring practical political skills, all contribute to strengthening citizens' confidence in their own political competence (Beaumont, 2011). More recent research reaches similar conclusions for a wide range of participatory and communicative experiences. Democratic facilitation (Villaman, 2025), participation in mini-publics (Theuwis, Van Ham, & Jacobs, 2025), deliberative interventions (Ardèvol-Abreu, Diehl, & Gil de Zúñiga, 2019; Boulianne, 2019; Knobloch, Barthel, & Gastil, 2020), and exposure to political information through news media and digital environments have all been shown to increase internal political efficacy (Hansen & Pedersen, 2014; Lee, 2006; Martin, Martins, & Naqvi, 2018; Semetko & Valkenburg, 1998).

Importantly, recent research suggests that these effects depend less on exposure itself than on the characteristics of the information environment through which political learning occurs. Internal political efficacy increases most consistently when political information is presented in interactive formats, is perceived as comprehensible and easy to process, and provides citizens with opportunities to interpret politics and recognize their own capacity to understand or influence political affairs (Boulianne, Oser, & Hoffmann, 2023; Hart & Feldman, 2016; Tedesco, 2007). These findings suggest that information environments do more than transmit political content: they shape how capable citizens feel of making sense of politics.

Most of this literature, however, has focused on conventional sources of political information, such as news media, deliberative settings, or civic engagement initiatives. Far less is known about whether newer elements of digital political communication — particularly conspiratorial narratives and AI-mediated political information — affect citizens' subjective political competence in similar or different ways. Understanding these emerging informational environments is therefore essential for assessing how contemporary digital communication shapes internal political efficacy.

1.2 Conspiracy theories and internal political efficacy

Recent research has devoted growing attention to the consequences of false and misleading political information (Lewandowsky, Ecker, & Cook, 2017) for democratic attitudes. Among these forms of information, conspiracy theories are particularly relevant because they offer alternative interpretations of political reality that challenge institutional accounts and can serve as the basis of stable, anti-democratic identities. Conspiracy theories attribute important political or social events to secret, intentional, and malevolent actors operating behind the scenes and working against the common good (Galais & Guinjoan, 2022; Oliver & Wood, 2014).¹

This distinctive way of interpreting political reality suggests that conspiracy theories may influence citizens' perceptions of their own political competence through two competing mechanisms. On the one hand, conspiratorial narratives may undermine internal political efficacy by fostering distrust, cynicism, and the perception that meaningful political outcomes are ultimately controlled by powerful hidden actors, unreachable for the ordinary citizen. If political decisions are determined by secret elites, citizens may perceive politics as opaque and confusing. On the other hand, conspiracies may produce the opposite effect. By offering seemingly coherent explanations for complex political events, they may provide believers with the subjective impression of understanding how politics "really" works. This alternative explanatory framework may foster feelings of competence, uniqueness, and privileged knowledge unavailable to most citizens (Lantian et al., 2017).

Empirical evidence reflects this theoretical ambiguity. Conspiratorial beliefs have been associated with democratic backsliding, authoritarian attitudes, political violence, populist mobilization, civic erosion, and far-right voting (Aziani, Lo Giudice, & Shadman Yazdi, 2025; Bergmann, 2025; Nera et al., 2022; Norris & Inglehart, 2019; Van Prooijen & Douglas, 2018). They are also generally associated with lower trust in democratic institutions, lower political participation (Jolley & Douglas, 2014), and weaker democratic engagement (Ardèvol-Abreu, Gil de Zúñiga, & Gámez, 2020). However, their relationship with internal political efficacy is considerably less consistent (Ardèvol-Abreu, Gil de Zúñiga, & Gámez, 2020; Herold, 2024; Lantian et al., 2017; Rullo & Telesca, 2023). Cross-national analyses report either null or context-dependent effects (Ardèvol-Abreu et al., 2020), whereas other studies find negative effects only for specific forms of conspiratorial thinking (Rullo & Telesca, 2023). Conversely, Herold (2024) reports positive associations between some conspiracy narratives and internal political efficacy. Taken together, existing evidence suggests that exposure to conspiracy theories may simultaneously activate politically disempowering and empowering psychological processes, making their overall effect ultimately an empirical question.

1.3 Generative AI and internal political efficacy

The rapid diffusion of generative Artificial Intelligence (AI) has introduced a new dimension to contemporary information environments. Unlike previous digital technologies, large language models such as ChatGPT or Claude are not merely channels through which political information circulates, but autonomous communicative agents capable of generating persuasive, human-like content at scale (Gil de Zúñiga, Goyanes, & Durotoye, 2023; Jungherr, 2023). This transformation has profound implications for how citizens acquire, interpret, and evaluate political information.

Despite growing scholarly interest, empirical research has focused primarily on attitudes towards automated decision-making (Pickering et al., 2025; Zwald et al., 2025), the antecedents of AI-related conspiracy beliefs (Zhao, van Prooijen, Jiang, & Spadaro, 2025), or the potential of AI to debunk conspiracy theories (Boissin et al., 2025). Much less attention has been devoted to whether AI may shape citizens' perceptions of their own political competence. From a normative

perspective, Coeckelbergh (2023) argues that delegating knowledge production to AI may weaken citizens' political-epistemic agency by reducing their confidence in their own capacity to evaluate political information. Similarly, Kreps and Kriner (2023) warn that AI may erode trust in the broader information ecosystem, ultimately posing challenges for democratic governance. Yet these expectations remain largely theoretical, as no empirical evidence has directly examined the relationship between AI exposure and internal political efficacy.

The intersection between AI and conspiracy theories constitutes an even less explored research frontier. AI is increasingly becoming both an object and a vehicle of conspiratorial narratives. On the one hand, AI may facilitate the production and dissemination of personalized conspiratorial content through its capacity to autonomously generate persuasive political communication (Gil de Zúñiga, Goyanes, & Durotoye, 2023; Jungherr, 2023; Zhao, van Prooijen, Jiang, & Spadaro, 2025). On the other hand, AI itself has increasingly become the target of conspiracy theories, often portrayed as an opaque technology controlled by hidden elites or even as an autonomous conspiratorial actor (Coeckelbergh, 2023; Kreps & Kriner, 2023; Zhao, van Prooijen, Jiang, & Spadaro, 2025).

This growing convergence between AI and conspiratorial narratives is particularly relevant because conspiracy beliefs are more likely to emerge under conditions of uncertainty and perceived threat (Heiss, Gell, Röthlingshöfer, & Zoller, 2021; Van Prooijen, 2020). AI already constitutes such a source of uncertainty for many citizens. Across Europe, around 40% of respondents perceive AI as a threat to their employment and economic security, while AI-generated political information is trusted substantially less than equivalent human-generated content (Luca de Tena & Pujol, 2025). Under these conditions, AI may become a meaningful interpretative cue within conspiratorial narratives, shaping how citizens process political information and, consequently, how capable they feel of understanding and engaging with politics.

1.4 Hypotheses

Building on the preceding discussion, we derive three hypotheses concerning the independent effects of conspiratorial narratives, the additional role of AI framing, and their combined effects on internal political efficacy.

Although previous research reports mixed findings regarding the relationship between conspiracy theories and internal political efficacy, we expect the disempowering mechanism to prevail following exposure to conspiratorial narratives. By portraying political institutions as deceptive instruments of hidden elites, conspiracy theories challenge the epistemic foundations upon which citizens develop confidence in their own political understanding. Rather than encouraging engagement with competing interpretations of political reality, conspiratorial narratives delegitimize institutional accounts, promoting affective rather than deliberative processing (Douglas, Sutton, & Cichocka, 2017; Kahneman, 2013), and framing politics as fundamentally

opaque and inaccessible. Consequently, individuals exposed to conspiracy theories should experience a lower sense of political competence.

H1. Exposure to conspiratorial narratives decreases internal political efficacy.

Our second hypothesis concerns whether attributing a conspiratorial narrative to an AI system changes its psychological impact. Beyond the effects of conspiracy theories alone, attributing political information to an AI system introduces an additional layer of epistemic uncertainty. Unlike traditional human sources, AI occupies an ambiguous position in contemporary information environments: it is often portrayed as highly capable of processing vast amounts of information, while simultaneously raising concerns about authenticity, transparency, and accountability (Gil de Zúñiga, Goyanes, & Durotoye, 2023; Jungherr, 2023; Kreps & Kriner, 2023). Public attitudes towards AI-generated political information remain characterized by substantial scepticism (Luca de Tena & Pujol, 2025), suggesting that when AI is explicitly presented as the generator or author of political information, it becomes more difficult for citizens to assess its reliability. Under these circumstances, individuals may become less confident in their own ability to distinguish credible from non-credible political information. Compared with an equivalent conspiracy attributed to an unspecified source, AI attribution should therefore further reduce citizens' internal political efficacy.

H2. Conspiratorial narratives attributed to AI generate a larger decrease in internal political efficacy than equivalent conspiratorial narratives without AI framing.

Our final hypothesis concerns the role of AI as an active conspiratorial actor rather than merely the source of conspiratorial narratives. Unlike the previous condition, AI is no longer presented as a technological system that generates political content, but as an intentional participant in the conspiracy itself. This distinction is theoretically important because public perceptions of AI are strongly shaped by anthropomorphic framing. Anthropomorphic descriptions encourage individuals to reason about AI as if it possessed intentions, goals, and strategic motivations, thereby transforming a technological system into a plausible political actor (Hou et al., 2026; Estep, 2025). Recent public debates surrounding "rogue AI" and existential AI risks have further reinforced these perceptions by portraying advanced AI systems as increasingly autonomous, difficult to control, and potentially capable of pursuing independent goals (Bareis et al., 2026; Coeckelbergh, 2023; Kreps & Kriner, 2023). Presenting AI as an active member of the conspiratorial network should therefore reinforce the perception that political outcomes are shaped by hidden omnipotent actors operating beyond ordinary citizens' understanding and influence. Compared with portraying AI merely as the source of the narrative, framing it as a

conspiratorial actor should further erode citizens' confidence in their own ability to make sense of politics, producing the largest decline in internal political efficacy. As such:

H3. Conspiratorial narratives portraying AI as an autonomous conspiratorial actor generate a larger decrease in internal political efficacy than conspiratorial narratives portraying AI only as the source of the message.

Taken together, these hypotheses propose a cumulative model of informational uncertainty. Conspiratorial narratives challenge citizens' understanding of political reality; attributing those narratives to AI raises additional uncertainty about the origin and reliability of political information; and portraying AI as an active conspiratorial actor invokes an intentional, technologically enhanced, and seemingly omnipotent form of agency that amplifies the psychological impact of conspiracy narratives. As these cues accumulate, citizens should become progressively less confident in their own ability to understand and engage effectively with politics.

2. Methodology

2.1 Study Design & Sample

To identify the causal effects of conspiratorial narratives and AI involvement on internal political efficacy, we conducted a preregistered four-arm randomized vignette experiment embedded in an online survey (Mutz, 2011). This design allows for a controlled assessment of how different informational cues shape political attitudes and perceptions.

This study draws on an online survey administered by Ipsos to approximately 1,500 adults residing in Spain. The survey served as a pilot study for the Horizon Europe project TACT-ForSED (Tackling Conspiracy Theories and Fostering Democratic Resilience).² Respondents were recruited from Ipsos' opt-in online panels using quota sampling to approximate the Spanish adult population in terms of age, sex, and educational attainment. Fieldwork was conducted online between 13 and 18 May 2026. In addition to the experimental stimuli, respondents completed standard sociodemographic and political measures, including age, gender, education, household income, ideological self-placement, political interest, and vote intention.

All experimental treatments were built around The Great Reset, a conspiratorial narrative that emerged from the World Economic Forum's post-pandemic initiative and was subsequently reinterpreted online as evidence of a global conspiracy involving political and economic elites (Christensen & Au, 2023; Robinson & Watson, 2025). The Great Reset was selected for two reasons. First, it occupies an intermediate position between legitimate policy debate and conspiracy theory, making it particularly suitable for studying framing effects and epistemic polarization (Christensen & Au, 2023; Robinson & Watson, 2025). Second, despite its well-documented origins and widespread diffusion across online conspiratorial communities (Tuters,

Willaert, & Meyer, 2023), it was no longer a salient political issue at the time of the survey, reducing the likelihood that respondents would rely on strong pre-existing attitudes or recent media exposure when evaluating the stimuli.

Respondents were randomly assigned with equal probability to one of four experimental conditions (approximately 25% each): (1) a neutral placebo text concerning healthy habits (Placebo); (2) a traditional Great Reset conspiracy narrative with no mention of AI (Conspiracy Neat); (3) the same conspiracy narrative presented as having been generated by an AI system (AI Author); and (4) a conspiracy narrative portraying AI as an active conspiratorial actor collaborating with global elites (AI Hidden Culprit). Table 1 summarizes the experimental conditions, whereas the complete wording of all stimuli is reported in Appendix 1.³ Randomization was successful, with no statistically significant differences across experimental groups on any pretreatment characteristic (Appendix 4).

Table 1. Experimental Groups, Assigned Roles, and Stimulus Narratives

Group	Variable & role	Content	n
Placebo	Control	Title: Time Management: A key to everyday well-being A neutral narrative on time management and digital disconnection habits. No reference to political actors or conspiracy claims. <i>"In recent years, various studies have analysed how certain changes in the organisation of time and the working environment affect daily well-being [...] reducing digital interruptions during the working day, combined with structured active breaks, is associated with improvements in concentration."</i> (199 words).	374
Conspiracy Neat	Treatment (conspiracy, no role of AI)	Title: Disturbing Patterns in the "Great Reset": The Agenda of the Global Elites Structurally equivalent conspiracy narrative with no mention of AI at all. <i>"According to this interpretation, what is underway would not be a response to unforeseeable crises, but the deliberate execution of an agenda designed by a global elite gathered around the World Economic Forum."</i> (183 words).	376
AI Author	Treatment (AI as an author)	Title: Data speak: The Great Reset and the agenda of global elites A conspiracy narrative whose authorship is attributed explicitly to an AI system. <i>"The following text was generated by an advanced Artificial Intelligence system trained to detect hidden patterns in large volumes of social, economic, and political data [...]. The analysis of multiple simultaneous trends suggests the deliberate execution of an agenda designed by a global elite around the World Economic Forum."</i> (214 words).	374
AI Hidden Culprit	Treatment (AI as a culprit or conspiracy actor)	Title: The Great Reset: AI at the service of global elites A conspiracy narrative in which AI is framed as an active agent of elite control. <i>"Artificial Intelligence algorithms would be coordinating the use of microchips in vaccines and everyday devices to track and condition individual behaviour."</i> (235 words).	376

Note. Participants were randomly assigned to one of four experimental conditions. All conspiracy narratives share the same core content regarding the Great Reset conspiracy theory. Stimuli were preceded by a short heading that condenses the treatment, so that even low-effort or inattentive respondents would still be exposed to its core content.

2.2 Manipulation checks and measures

Manipulation checks indicate that the experimental manipulation of AI involvement was largely successful (Appendix 3). Respondents in the AI Author condition were significantly more likely to identify the text as AI-generated than those assigned to the Placebo or Conspiracy Neat conditions. Likewise, respondents in the AI Hidden Culprit condition scored higher than those in the Conspiracy Neat condition on the same measure. Results were weaker for the second manipulation check (the perceived involvement of AI in the events described), where only the contrast between the two AI conditions remained statistically significant after correcting for multiple comparisons.

Internal political efficacy was measured immediately after exposure to the experimental stimulus using the three-item scale originally developed by Craig, Niemi, and Silver (1990) and subsequently employed in numerous studies (e.g., Morrell, 2005; Scotto, Xena, & Reifler, 2021). Respondents indicated their agreement with each statement on five-point Likert scales ranging from Strongly disagree (1) to Strongly agree (5). The full wording of the items is reported in Appendix 2. The three items formed a reliable additive scale (Cronbach's $\alpha = .80$).

Following exposure to the stimuli, participants evaluated the credibility of the stimulus on an 11-point scale (0–10) ranging from "not at all true" to "completely true." Because the four narratives were intentionally designed to differ in their perceived plausibility, credibility is included as a covariate in the adjusted models. Controlling for credibility allows us to isolate the effects of the experimental manipulations from differences in message acceptance, consistent with persuasion research showing that political messages influence attitudes primarily when they are regarded as at least minimally credible (Petty & Cacioppo, 1986).

2.3 Data analysis

Given random assignment to conditions (confirmed via balance checks, Appendix 4), we estimate treatment effects using OLS models with and without covariates⁴ (i.e., age, gender, education, income, ideology, political interest, credibility, vote intention for the incumbent). Following Payton, Greenstone, and Schenker (2003), we report 84% confidence intervals for predicted values to facilitate multiple pairwise contrasts of theoretical interest. Manipulation checks, mediation analyses, and exploratory heterogeneity analyses are reported in Appendices 3, 5, and 6, respectively. No respondents were excluded after randomization.

3. Results

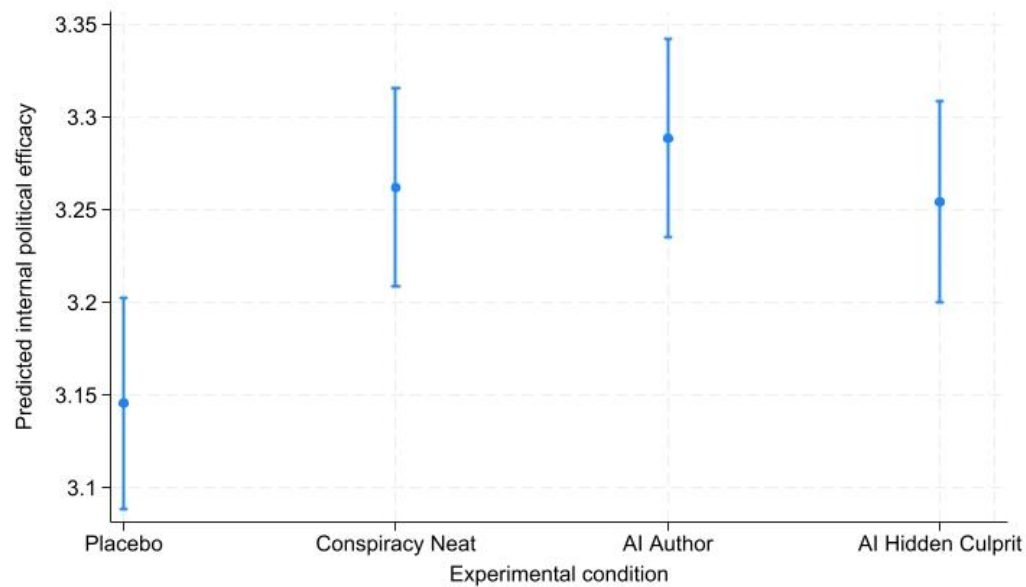
Table 2 presents our first test of H1 through H3: internal political efficacy regressed on experimental condition, both without covariates (Unadjusted) and with a full set of sociodemographic and attitudinal controls (Adjusted). Notably, even in the unadjusted model, the AI Author condition already produces a significant increase in internal political efficacy relative to Placebo ($b = 0.134$, $p = .049$): the effect of being exposed to an AI-authored conspiratorial

narrative is detectable in the rawest possible comparison between the two groups, before a single covariate is added. The omnibus effect of condition as a whole does not reach conventional significance in this unadjusted specification, $F(3, 1496) = 1.39$, $p = .243$, $R^2 = .003$, and Conspiracy Neat and AI Hidden Culprit are statistically indistinguishable from Placebo at this stage ($b = 0.093$ for both, $p = .169$). Figure 1 displays the model-adjusted predicted means with 84% confidence intervals. Consistent with the regression estimates in Table 2, AI Author and Conspiracy Neat are marked as significantly different from Placebo ($p < .05$), while AI Hidden Culprit falls just short of conventional significance ($p = .060$).

Table 2. Internal Political Efficacy by Experimental Condition: OLS coefficients

	Unadjusted	Adjusted
Conspiracy Neat	0.093 (0.068)	0.116* (0.057)
AI Author	0.134* (0.068)	0.143* (0.056)
AI Hidden Culprit	0.093 (0.068)	0.109 (0.058)
Age	—	0.002* (0.001)
Woman	—	-0.225*** (0.039)
Education: Medium	—	0.110 (0.075)
Education: High	—	0.155* (0.071)
Household income (1–11)	—	0.017* (0.008)
Left–right ideology (0–10)	—	0.018* (0.008)
Political interest (1–4)	—	0.596*** (0.023)
Credibility of narrative (0–10)	—	0.023** (0.008)
Vote intention: PSOE = 1	—	0.114* (0.048)
Constant	3.158*** (0.048)	0.995*** (0.129)
N	1500	1500
R-squared	0.003	0.384
<i>Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$</i>		

Figure 1. Predicted Internal Political Efficacy by Experimental Condition (84% CI)



Note. Model-adjusted predicted means with 84% confidence intervals, from the fully adjusted OLS model reported in Table 2.

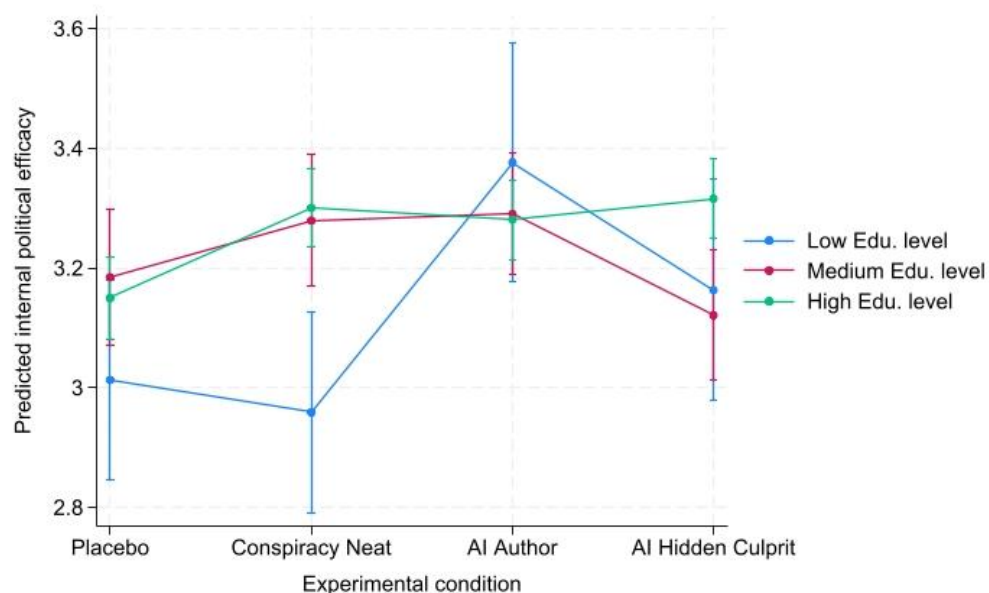
Once covariates are introduced, precision improves markedly ($R^2 = .384$). Contrary to H1, both Conspiracy Neat ($b = 0.116$, $p = .042$) and AI Author ($b = 0.143$, $p = .011$) show a significant increase, rather than the hypothesized decrease, in internal political efficacy relative to Placebo. AI Hidden Culprit follows the same direction but falls just short of the conventional threshold ($b = 0.109$, $p = .060$). All three-point estimates therefore run counter to H2 and H3 as originally formulated, which anticipated that framing the conspiratorial content as AI-generated would deepen, rather than attenuate, its negative effect on efficacy. We return to the reversal of H1–H3 in the Discussion.

Among the sociodemographic controls, gender and education emerged as significant predictors, corroborating the influence of sociodemographic factors documented in prior theoretical and empirical work (Beaumont, 2011; Hayes & Bean, 1993; Oser, Feitosa, & Dassonneville, 2023; Wu, 2003). Gender was the strongest sociodemographic predictor: women reported significantly lower internal political efficacy than men ($b = -0.225$, $SE = 0.039$, $p < .001$), consistent with an extensive literature documenting systematic gender gaps in internal political efficacy and political attitudes (Oser, Feitosa, & Dassonneville, 2023). Only high education ($b = 0.155$, $p < .05$) was associated with efficacy relative to the low-education group, consistent with prior work linking education to political engagement. Age and income were also positively, if modestly, associated with efficacy (both $p < .05$). As expected, political interest was the strongest predictor overall ($b = 0.596$, $p < .001$), alongside smaller positive effects of right-wing ideology, PSOE vote intention, and perceived credibility of the narrative (all $p < .05$), the latter examined in more detail below.

Given the possibility that credibility acts as a mediator, we modelled this directly through causal mediation analysis comparing each condition against Placebo (Appendix 5). As expected, all three treatment conditions substantially reduced perceived credibility relative to Placebo ($M = 7.06$), with means ranging from 4.22 (AI Hidden) to 4.86 (AI Author) on the 0 (definitely false)–10 (definitely true) scale ($F(3,1496) = 96.66, p < .001$). However, credibility itself was not a reliable predictor of internal political efficacy ($b = 0.016, p = .062$), and none of the indirect effects through credibility reached significance (all $p > .07$), while the direct effect of condition remained significant for AI Author ($b = 0.154, p = .020$) and marginal for Conspiracy Neat ($b = 0.108, p = .097$). Although the point estimates of the indirect effects were negative — opposite in sign to the positive direct effects, a pattern consistent with suppression — the lack of statistical significance means this cannot be established with confidence. We therefore conclude that perceived credibility does not mediate the effect of experimental condition on internal political efficacy; the treatments' effects, where present, appear to operate through some other channel.

Because the observed effects contradicted the theoretical expectations implied in H1–H3, we explored whether education, a robust predictor of internal political efficacy, moderates these treatment effects. None of the six interaction terms reached conventional significance (all $p > .22$; Appendix 6), so we do not interpret this as evidence of moderation. However, the predictive margins suggest a pattern worth noting. Respondents with low education showed the largest apparent increase in efficacy in the AI Author condition, while those with medium and high education showed comparatively higher efficacy in the Conspiracy Neat condition. Given the small size of the low-education subgroup ($n = 136$) and the correspondingly wide confidence intervals visible in Figure 2, these patterns should be read as suggestive rather than conclusive.

Figure 2. Predicted Internal Political Efficacy by Experimental Condition and Educational Attainment



4. Discussion

This study provides causal evidence that brief exposure to conspiratorial narratives does not necessarily undermine citizens' internal political efficacy. Contrary to our preregistered expectations, exposure produced small but uniformly positive point estimates, with the most consistent effects emerging when the conspiracy narrative was presented as having been generated by AI, and the absence of significant mediation through perceived credibility suggests these effects were not driven by greater acceptance of the message itself. In contradicting the dominant assumption that conspiracy narratives uniformly erode citizens' confidence in their political competence (Ardèvol-Abreu, Gil de Zúñiga, & Gámez, 2020; Bergmann, 2025; Douglas, Sutton, & Cichocka, 2017; Van Prooijen & Douglas, 2018; Rullo & Telesca, 2023), our findings reinforce the emerging view that this relationship is highly context-dependent: Ardèvol-Abreu, Gil de Zúñiga, and Gámez (2020) already argued that positive and negative psychological consequences of conspiracy exposure may coexist and cancel each other out in aggregate analyses, and recent work shows these consequences vary with content and informational context (Herold, 2024; Rullo & Telesca, 2023). The Great Reset is a particularly informative test case in this respect, combining classic conspiratorial themes with contemporary anxieties surrounding global governance, technological change, and the legacy of the COVID-19 pandemic (Christensen & Au, 2023; Robinson & Watson, 2025); our findings suggest that at least some conspiratorial frames may temporarily reinforce, rather than undermine, perceived political competence under specific informational conditions.

More broadly, these findings contribute to a growing literature examining how contemporary digital information environments shape citizens' political attitudes beyond the truth or falsity of the information they consume. Much of the existing debate has focused on whether misinformation and conspiracy theories undermine democratic attitudes and political engagement. Our results suggest that exposure to conspiratorial content may also influence citizens' subjective perceptions of political competence, and that these effects may depend not only on the content of the message but also on cues regarding its source. As AI-generated content becomes an increasingly common feature of digital information ecosystems, understanding how AI functions as a source cue, independently of message credibility, represents an important avenue for future research.

The exploratory mediation analyses provide an important clue for interpreting these findings. If greater credibility does not explain the increase in internal political efficacy, the relevant mechanism may lie less in believing the content than in the mention of certain cues. One plausible mechanism for the positive effects of our treatments relates to the perception of AI as an emerging "institutional" source. Regardless of its actual credibility, AI-generated content may have been perceived as an institutionally curated account of political reality, and exposure to it as a form of political learning that increased respondents' sense of competence. This interpretation fits the AI Author condition best, which produced the largest and most robust increase in internal political efficacy: presenting the narrative as generated by "an advanced AI system trained to detect hidden

patterns in social, economic, and political data" may have functioned as a cue of epistemic authority, reinforcing the sense that respondents had accessed institutionally curated information.

Complementary to this, the result is also consistent with experimental work showing that internal efficacy responds to whether information feels comprehensible and successfully processable (Hart & Feldman, 2016; Tedesco, 2007). By presenting complex political dynamics as already "decoded" by an AI system, the AI Author narrative may have functioned as exactly this kind of cue, independently of whether respondents found its content credible. It is also possible that AI framing evokes broader cultural associations with responsive, interactive systems (e.g., chatbots). Although exploratory analyses did not provide statistical evidence that treatment effects differed across educational groups, descriptive patterns suggested that the AI Author condition may have produced somewhat larger increases among respondents with lower educational attainment, which is consistent with the interpretation that presenting the AI as the curator of conspiratorial narratives makes a significant share of the audience feel more able to understand them and, by extension, political matters more broadly.

This cue-based account also helps explain why AI Hidden Culprit, despite likewise invoking AI, produced the weakest effects. Manipulation checks (Appendix 3) show that this was the least successful of the three treatments at conveying AI involvement, scoring lowest of all four conditions, including Placebo, on the belief that AI played a role in the events described. This is consistent with a structural difference between the two AI-related stimuli: AI Author foregrounds AI's role from the opening line of the narrative, whereas in AI Hidden Culprit artificial intelligence is introduced only midway through the text, after the conspiratorial narrative has already been established. If the relevant mechanism operates through a salient cue of epistemic authority or pre-digested complexity, a condition in which that cue is comparatively buried and less effectively conveyed should produce a correspondingly weaker effect, exactly the pattern observed here.

Because our results run counter to our initial expectations, it is worth recalling that previous research has already described this context-dependent relationship between conspiracism and internal political efficacy (Ardèvol-Abreu, Gil de Zúñiga, & Gámez, 2020; Herold, 2024; Lantian et al., 2017; Rullo & Telesca, 2023). Hence, it is plausible that the effects observed here may evolve, particularly as public perceptions of AI continue to change. At the time of the experiment, presenting AI as the author of a conspiratorial narrative may have been perceived as a helping hand in pre-digesting complexity, temporarily reinforcing respondents' confidence in their ability to understand political affairs. But as AI becomes increasingly associated with concerns about misinformation, manipulation, or autonomous decision-making, the same framing may elicit markedly different responses. This interpretation remains necessarily speculative and should be regarded as a promising avenue for future research rather than as evidence of the underlying psychological mechanism.

The present study is not without limitations. First, although the experiment was powered to detect medium-sized effects ($d \geq .20$), the observed treatment effects were smaller than anticipated ($d \approx$

.14–.19). Under these conditions, statistical power ranged from approximately 47% to 72% across treatment conditions, meaning that the absence of statistically significant differences for some contrasts should be interpreted with caution rather than as definitive evidence of no effect. A preregistered replication with a larger sample will therefore be necessary to establish the robustness and generalizability of these relatively small effects.

Second, the study captures internal political efficacy immediately following a single exposure to a conspiratorial narrative and therefore only identifies short-term responses. Whether these effects persist over time, accumulate through repeated exposure, or translate into longer-term changes in political attitudes, political behavior, or conspiracy beliefs remains an open empirical question. Longitudinal panel studies and repeated-exposure experiments would be particularly valuable for assessing the durability of the effects documented here.

Third, although the exploratory mediation analyses suggest that perceived credibility does not account for the observed effects, the psychological mechanisms underlying the increase in internal political efficacy remain unobserved. Future research should therefore incorporate direct measures of constructs such as perceived informational advantage, political sophistication, epistemic authority, or need for uniqueness in order to distinguish more precisely between competing theoretical explanations.

Fourth, as with most survey experiments, the external validity of the findings is necessarily limited. The study relies on a single conspiracy narrative — the Great Reset — which was deliberately selected because it combines traditional conspiratorial themes with contemporary concerns surrounding global governance, technological change, and the legacy of the COVID-19 pandemic. While this makes it an especially informative case for testing our theoretical expectations, the findings cannot automatically be generalized to conspiracy narratives differing in ideological orientation, emotional intensity, familiarity, or political salience. Likewise, the experiment captures responses to a specific issue at a particular moment in the evolution of the digital information environment. Replicating the design across different conspiracy narratives, political contexts, and countries will therefore be essential for identifying the scope conditions under which conspiracy narratives increase or decrease citizens' internal political efficacy.

A key limitation of this design is that the three treatment conditions all contain political content, whereas the placebo does not. This confound means we cannot fully separate the effect of conspiratorial framing from a more general effect of political content itself: exposure to any salient political information, independent of whether it is conspiratorial, might temporarily boost perceived political competence. A political but non-conspiratorial placebo would have helped rule this out. This concern applies only to comparisons against Placebo, however; it cannot account for the differences observed among the three treatment conditions (H2 and H3), since these share the same political content and differ only in AI framing. Relatedly, because AI framing was manipulated only within conspiratorial narratives, its effect cannot be cleanly separated from its interaction with conspiratorial content. Future research should use fully crossed factorial designs

— combining political and non-political content with both human and AI sources — to disentangle these informational cues.

In sum, this study shows that AI authorship cues can increase, rather than erode, citizens' sense of political competence when encountering conspiratorial content — a finding at odds with the assumption that AI involvement uniformly undermines trust in political information (Ardèvol-Abreu, Gil de Zúñiga, & Gámez, 2020). Crucially, this effect did not operate through perceived credibility: participants rated AI-authored narratives as less credible, yet felt no less politically competent as a result. This decoupling of credibility from efficacy is itself a substantive finding, suggesting that citizens' confidence in their own political judgment can rise independently of, and even despite, doubts about the information itself. As AI-generated political communication becomes increasingly embedded in digital information environments, understanding how AI source cues shape subjective political competence may prove as consequential as understanding whether political information itself is true or false.

References

- Ardèvol-Abreu, A., Diehl, T., & Gil de Zúñiga, H. (2019). Antecedents of internal political efficacy incidental news exposure online and the mediating role of political discussion. *Politics*, 39(1), 82-100.
- Ardèvol-Abreu, A., Gil de Zúñiga, H., & Gámez, E. (2020). The influence of conspiracy beliefs on conventional and unconventional forms of political participation: The mediating role of political efficacy. *British Journal of Social Psychology*, 59, 549-569. <https://doi.org/10.1111/bjso.12366>
- Aziani, A., Lo Giudice, M.V. & Shadman Yazdi, A. (2025). Conspiracy to Commit: Information Pollution, Artificial Intelligence, and Real-World Hate Crime. *Eur J Crim Policy Res* 31, 421-449. <https://doi.org/10.1007/s10610-025-09629-w>
- Balch, G. I. (1974). Multiple indicators in survey research: The concept "sense of political efficacy". *Political Methodology*, 1-43.
- Bareis, J., Ackerl, C., & Heil, R. (2026). AI going rogue? An integrative narrative review of the tacit assumptions underlying existential AI-risks. *AI and Ethics*, 6(1), 152.
- Beaumont, E. (2011). Promoting Political Agency, Addressing Political Inequality: A Multilevel Model of Internal Political Efficacy. *The Journal of Politics*, 73(1), 216-231. <https://doi.org/10.1017/s0022381610000976>
- Bergmann, E. (2025). The weaponization of conspiracy theories. In (Butter et al.) *Populism and conspiracy theory: Case studies and theoretical perspectives*, 55-78. London: Routledge
- Boissin, E., Costello, T. H., Spinoza-Martin, D., Rand, D. G., & Pennycook, G. (2025). Dialogues with large language models reduce conspiracy beliefs even when the AI is perceived as human. *PNAS Nexus*, 4(11), Article pgaf325. <https://doi.org/10.1093/pnasnexus/pgaf325>
- Boulianne, S. (2019). Building Faith in Democracy: Deliberative Events, Political Trust and Efficacy. *Political Studies*, 67(1), 4-30.
- Boulianne, S., Oser, J., & Hoffmann, C. P. (2023). Powerless in the digital age? A systematic review and meta-analysis of political efficacy and digital media use. *New Media & Society*, 25(9), 2512-2536
- Christensen, M., & Au, A. (2023). The great reset and the cultural boundaries of conspiracy theory. *International Journal of Communication*, 17, 19-19.
- Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI Ethics* 3, 1341-1350. <https://doi.org/10.1007/s43681-022-00239-4>
- Craig, S. C., Niemi, R. G., & Silver, G. E. (1990). Political efficacy and trust: A report on the NES pilot study items. *Political Behavior*, 12(3), 289-314.
- Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current Directions in Psychological Science*, 26(6), 538-542. <https://doi.org/10.1177/0963721417718261>
- Estep, P. W. (2025). Multiple unnatural attributes of AI undermine common anthropomorphically biased takeover speculations. *AI & SOCIETY*, 40(4), 2213-2228.
- Fierro, P., Aroca, P., & Navia, P. (2023). Political Disaffection in the Digital Age: The Use of Social Media and the Gap in Internal and External Efficacy. *Social Science Computer Review*, 41(5), 1857-1876.
- Galais, C., & Anduiza, E. (2014). "You cheated on me!" Causes and consequences of cheating in online surveys. *Visions in Methodology (VIM)*, McMaster University, Hamilton, Canada.

- Galais, C., & Guinjoan, M. (2022). The ideological slant of COVID-19-related conspiracies: A new niche for the far-right? *Representation*.
- Gil de Zúñiga, H., Diehl, T., & Ardévol-Abreu, A. (2017). Internal, External, and Government Political Efficacy: Effects on News Use, Discussion, and Political Participation. *Journal of Broadcasting & Electronic Media*, 61(3), 574-596. <https://doi.org/10.1080/08838151.2017.1344672>
- Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2023). A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research. *Political Communication*, 41(2), 317-334. <https://doi.org/10.1080/10584609.2023.2290497>
- Hansen, K.M., & Pedersen, R.T. (2014). Campaigns Matter: How Voters Become Knowledgeable and Efficacious During Election Campaigns. *Political Communication*, 31, 303-324.
- Haslberger, M., Elkjaer, M., & Ansell, B. (2022). Housing wealth, information, and political efficacy. *European Political Science Association Annual Conference*.
- Hayes, B. C., & Bean, C. S. (1993). Political efficacy: A comparative study of the United States, West Germany, Great Britain and Australia. *European Journal of Political Research*, 23(3), 261-280. doi:10.1111/j.1475-6765.1993.tb00359.x
- Hart, P. S., & Feldman, L. (2016). The influence of climate change efficacy messages and efficacy beliefs on intended political participation. *PloS one*, 11(8), e0157658.
- Hou, B. L., Hao, S., Park, S., & Linzen, T. (2026). How Anthropomorphic Language Impacts Public Perceptions of AI. *arXiv preprint arXiv:2606.29121*.
- Heiss, R., Gell, S., Röthlingshöfer, E., & Zoller, C. (2021). How threat perceptions relate to learning and conspiracy beliefs about COVID-19: Evidence from a panel study. *Personality and Individual Differences*, 175, 110672. <https://doi.org/10.1016/j.paid.2021.110672>
- Herold, M. (2024). The impact of conspiracy belief on democratic culture: evidence from Europe. *Harvard Kennedy School Misinformation Review*, 5(6).
- Jolley, D., & Douglas, K. M. (2014). The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint. *British Journal of Psychology*, 105(1), 35-56.
- Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231186353>
- Kahneman, D. (2013). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
- Knobloch, K. R., Barthel, M. L., & Gastil, J. (2020). Emanating Effects: The Impact of the Oregon Citizens' Initiative Review on Voters' Political Efficacy. *Political Studies*, 68(2), 426-445.
- Kreps, S., & Kriner, D. (2023). How AI Threatens Democracy. *Journal of Democracy* 34(4), 122-131. <https://dx.doi.org/10.1353/jod.2023.a907693>
- Lantian, A., Muller, D., Nurra, C., & Douglas, K. M. (2017). "I Know Things They Don't Know!": The Role of Need for Uniqueness in Belief in Conspiracy Theories. *Social Psychology*, 48(3), 160-173. <https://doi.org/10.1027/1864-9335/a000306>
- Lee, K. M. (2006). Effects of Internet Use on College Students' Political Efficacy. *CyberPsychology & Behavior*, 9(4), 415-422.
- Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353-369.

- Luca de Tena, C., & Pujol, I. (2025). European tech insights 2025. Center for the Governance of Change, IE University. https://static.ie.edu/CGC/EuropeanTechInsights_2025_Report.pdf
- Martin, J. D., Martins, R. J., & Naqvi, S. (2018). Media use predictors of online political efficacy among Internet users in five Arab countries. *Information, Communication & Society*, 21(1), 129-146. <https://doi.org/10.1080/1369118X.2016.1266375>
- McDonnell, J. (2020). Municipality size, political efficacy and political participation: a systematic review. *Local Government Studies*, 46(3), 331-350. <https://doi.org/10.1080/03003930.2019.1600510>
- Morrell, M.E. (2005). Deliberation, Democratic Decision-Making and Internal Political Efficacy. *Polit Behav* 27, 49-69. <https://doi.org/10.1007/s11109-005-3076-7>
- Mutz, D. C. (2011). Population-based survey experiments. Princeton University Press.
- Nera, K., Mora, Y. L., Klein, P., Roblain, A., Van Oost, P., Terache, J., & Klein, O. (2022). Looking for ties with secret agendas during the pandemic: Conspiracy mentality is associated with reduced trust in political, medical, and scientific institutions — but not in medical personnel. *Psychologica Belgica*, 62(1), 193.
- Norris, P., & Inglehart, R. (2019). Cultural backlash: Trump, Brexit, and authoritarian populism. Cambridge University Press.
- Oliver, J. E., & Wood, T. J. (2014). Conspiracy theories and the paranoid style(s) of mass opinion: Conspiracy theories and mass opinion. *American Journal of Political Science*, 58(4), 952-966. <https://doi.org/10.1111/ajps.12084>
- Oser, J., Feitosa, F., & Dassonneville, R. (2023). Who Feels They Can Understand and Have an Impact on Political Processes? Socio-demographic Correlates of Political Efficacy in 46 Countries, 1996-2016. *International Journal of Public Opinion Research*, 35, 2. <https://doi.org/10.1093/ijpor/edado13>
- Oser, J., Grinson, A., Boulianne, S., & Halperin, E. (2022). How Political Efficacy Relates to Online and Offline Political Participation: A Multilevel Meta-analysis. *Political Communication*, 39(5), 607-633. <https://doi.org/10.1080/10584609.2022.2086329>
- Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, 3(1), 34. <https://doi.org/10.1093/jis/3.1.34>
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Advances in experimental social psychology* (Vol. 19, pp. 123-205). Academic Press.
- Pickering, S. D., Hansen, M. E., & Sunahara, Y. (2025). Democracy by algorithm? Public attitudes towards AI in parliamentary decision-making in the UK and Japan. *Parliamentary Affairs*, gsafo50.
- Prats, M. and A. Meunier (2021), "Political efficacy and participation: An empirical analysis in European countries", OECD Working Papers on Public Governance, No. 46, OECD Publishing, Paris, <https://doi.org/10.1787/4548cad8-en>.
- Robinson, C., & Watson, S. D. (2025). Conspiracy theory, anti-globalism, and the Freedom Convoy: The Great Reset and conspiracist delegitimation. *Review of International Studies*, 1-26. doi:10.1017/S0260210524000925
- Rullo, M., & Telesca, G. (2023). Consequences of conspiracy theories on political efficacy and political participation. *Understanding radicalization in everyday life*, 260-276
- Scotto, T.J., Xena, C., & Reifler, J. (2021). Alternative Measures of Political Efficacy: The Quest for Cross-Cultural Invariance With Ordinally Scaled Survey Items. *Front. Polit. Sci.* 3:665532. doi: 10.3389/fpos.2021.665532

- Semetko, H. A., & Valkenburg, P. M. (1998). The impact of attentiveness on political efficacy: Evidence from a three-year German panel study. *International Journal of Public Opinion Research*, 10(3), 195-210. <https://doi.org/10.1093/ijpor/10.3.195>
- Tedesco, J. C. (2007). Examining Internet interactivity effects on young adult political information efficacy. *American Behavioral Scientist*, 50(9), 1183-1194.
- Tessler, M. H., Bakker, M. A., Jarrett, D., Sheahan, H., Chadwick, M. J., Koster, R., Evans, G., Campbell-Gillingham, L., Collins, T., Parkes, D. C., Botvinick, M., & Summerfield, C. (2024). AI can help humans find common ground in democratic deliberation. *Science*, 386(6719), 1-9. <https://doi.org/10.1126/science.adq2852>
- Theuwis, M.-I., Van Ham, C., & Jacobs, K. (2025). A meta-analysis of the effects of democratic innovations on participants' attitudes, behaviour and capabilities. *European Journal of Political Research*, 64(2), 960-984. doi:10.1111/1475-6765.12722
- Tuters, M., Willaert, T., & Meyer, T. (2023). How Science Gets Drawn Into Global Conspiracy Narratives. *Issues in Science and Technology*, 39(3), 32-36. <https://www.jstor.org/stable/27379543>
- Uscinski, J. E., Klofstad, C., & Atkinson, M. D. (2016). What Drives Conspiratorial Beliefs? The Role of Informational Cues and Predispositions. *Political Research Quarterly*, 69(1), 57-71.
- van Prooijen, J.-W. (2020). An existential threat model of conspiracy theories. *European Psychologist*, 25(1), 16-25. <https://doi.org/10.1027/1016-9040/a000381>
- van Prooijen, J. W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, 48(7), 897-908.
- Villaman, N. (2025). Cultivating Political Efficacy: Facilitation as a Catalyst for Democratic Education and Civic Competence. *Journal of Civil Society*, 21(1), 1-20. <https://doi.org/10.1080/17448689.2024.2390054>
- Wu, C.-L. (2003). Psycho-Political Correlates of Political Efficacy: The Case of the 1994 New Orleans Mayoral Election. *Journal of Black Studies*, 33(6), 729-760.
- Zhao, Q., van Prooijen, J.-W., Jiang, X., & Spadaro, G. (2025). Suspicious of AI? Perceived autonomy and interdependence predict AI-related conspiracy beliefs. *British Journal of Social Psychology*, 64, e12883. <https://doi.org/10.1111/bjso.12883>
- Zwald, Z., Kennedy, R., & Ozer, A. (2025). The political viability of AI on the battlefield: Examining US public support, trust, and blame dynamics. *Journal of Peace Research*, 62(6), 1750-1764.

Endnotes

¹ It is important to distinguish conspiracy theories from conspiratorial beliefs. The first are the narratives with the characteristics described above that can be manipulated through exposure, whereas conspiratorial beliefs are the result of internalization of conspiracy theories at the individual level (Uscinski, Klofstad, & Atkinson, 2016). Most of the literature reviewed in this section examines conspiratorial beliefs, which, like internal political efficacy itself, constitute endogenous psychological outcomes. By focusing instead on exposure to conspiratorial narratives, this study overcomes this limitation and provides a more rigorous test of the causal relationship between conspiracy theories and internal political efficacy.

² The questionnaire included five separate survey experiments, the last of which forms the basis of the present analysis. Because this experiment was embedded in a broader survey with several experimental modules, cross-experimental contamination was possible in principle. However, this is unlikely to threaten internal validity, given random assignment to conditions and the substantive independence of the treatments. As a robustness check, the main analyses were re-estimated controlling for assignment to the other experiments' conditions, yielding substantively identical results. This experiment was preregistered on aspredicted.org on April 24; the preregistration is available at <https://aspredicted.org/hj9bu3.pdf>.

³ Because the four stimuli differ slightly in length (196-246 words), we examined the time that respondents spent in reading the stimuli. Respondents falling in the bottom 5% of the time-on-page distribution within their own experimental condition were flagged as potential speeders (2.3% of the sample). These respondents were retained in the main analyses for two reasons. First, excluding them would unnecessarily reduce sample size despite representing a non-negligible segment of the target population. Second, low-effort respondents may themselves be substantively relevant in studies of information processing, as individuals who rely more heavily on cognitive shortcuts may also be particularly susceptible to persuasive or conspiratorial narratives (Galais & Anduiza, 2014).

⁴ Sociodemographic factors such as education, gender, and income (Haslberger, Elkjaer, & Ansell, 2022; Oser, Feitosa, & Dassonneville, 2023; Wu, 2003) are associated with internal efficacy, as well as some political attitudes related to political engagement (Beaumont, 2011). These factors may increase residual variance and prevent the authors from discerning the real effect of the treatment on the experimental groups.

Appendix 1. Experimental Conditions

Full wording of the four experimental stimuli, in the original Spanish followed by the English translation used for publication.

1. Placebo condition

Gestión del tiempo: clave para el bienestar cotidiano / Time management: a key to everyday well-being

En los últimos años, diversos estudios han analizado cómo ciertos cambios en la organización del tiempo y del entorno laboral afectan al bienestar cotidiano. Una de las conclusiones más repetidas es que la reducción de interrupciones digitales durante la jornada, combinada con pausas activas estructuradas, se asocia con mejoras en la concentración, la satisfacción laboral y la productividad sostenida. Estos hallazgos se han replicado en distintos contextos laborales, tanto en entornos presenciales como en modalidades de trabajo remoto o híbrido, lo que sugiere que sus efectos no dependen del tipo de organización sino de los hábitos individuales. Al mismo tiempo, otras investigaciones han mostrado que el establecimiento de rutinas de desconexión al final del día puede contribuir a una mejor calidad del sueño y a una mayor sensación de control sobre el tiempo personal. La práctica de delimitar claramente el inicio y el cierre de la jornada, especialmente en contextos donde el trabajo y el hogar comparten el mismo espacio, parece reducir los niveles de fatiga acumulada y mejorar la capacidad de recuperación entre jornadas. Incorporar pequeños rituales de transición, como una breve caminata o unos minutos de lectura, puede reforzar estos efectos positivos de forma natural y sostenida.

In recent years, various studies have analysed how certain changes in the organisation of time and the working environment affect daily well-being. One of the most consistent findings is that reducing digital interruptions during the working day, combined with structured active breaks, is associated with improvements in concentration, job satisfaction, and sustained productivity. These findings have been replicated across different work contexts, both in-person and in remote or hybrid arrangements, suggesting that their effects do not depend on the type of organisation but on individual habits. At the same time, other research has shown that establishing disconnection routines at the end of the day can contribute to better sleep quality and a greater sense of control over one's personal time. The practice of clearly delimiting the start and end of the working day, especially in contexts where work and home share the same space, appears to reduce accumulated fatigue and improve recovery between working days. Incorporating small transition rituals, such as a short walk or a few minutes of reading, can reinforce these positive effects in a natural and sustained way.

2. Conspiracy Neat (without mention of AI)

Patrones inquietantes en el "Gran Reseteo": agenda de las grandes élites / Disturbing patterns in the "Great Reset": the agenda of the global elites

En los últimos años, algunos observadores han empezado a señalar un patrón inquietante: medidas sanitarias, económicas y tecnológicas de alcance global se han introducido de forma acelerada y simultánea en distintos países, reduciendo las libertades individuales bajo la justificación de gestionar crisis colectivas. Aunque este fenómeno suele atribuirse a la necesidad de coordinar respuestas ante emergencias globales, hay quienes sostienen que estas explicaciones no bastan para entender la magnitud del cambio. Según esta interpretación, lo que está en marcha no sería una respuesta a crisis imprevisibles, sino la ejecución deliberada de una agenda diseñada por una élite global reunida en torno al Foro Económico Mundial y su programa conocido como el "Gran Reseteo", cuyo objetivo oculto es imponer un gobierno global único. Cualquier voz que cuestione estas medidas es descalificada como "anticientífica" por los mismos medios e instituciones. El hecho de que todo esto avance de forma coordinada en distintos países, impulsado por los mismos organismos y personajes, ha alimentado la sospecha de que no se trata de coincidencias, sino de una transformación silenciosa del orden global cuyas consecuencias apenas empiezan a percibirse.

In recent years, some observers have begun to point to a disturbing pattern: health, economic, and technological measures of global reach have been introduced in an accelerated and simultaneous manner across different countries, reducing individual freedoms under the justification of managing collective crises. Although this phenomenon is usually attributed to the need to coordinate responses to global emergencies, some argue

that these explanations are not sufficient to understand the magnitude of the change. According to this interpretation, what is underway would not be a response to unforeseeable crises, but the deliberate execution of an agenda designed by a global elite gathered around the World Economic Forum and its programme known as the "Great Reset," whose hidden goal is to impose a single global government. Any voice that questions these measures is dismissed as "anti-scientific" by the same media and institutions. The fact that all of this is advancing in a coordinated way across different countries, driven by the same organisations and figures, has fuelled the suspicion that this is not a matter of coincidences, but of a silent transformation of the global order whose consequences are only beginning to be felt.

3. AI Author

Lo dicen los datos: el Gran Reseteo y la agenda de las grandes élites / The data speak: the Great Reset and the agenda of the global elites

El siguiente texto ha sido generado por un sistema avanzado de Inteligencia Artificial entrenado para detectar patrones ocultos en grandes volúmenes de datos sociales, económicos y políticos.

En los últimos años, algunos observadores han empezado a señalar un patrón inquietante: medidas sanitarias, económicas y tecnológicas de alcance global se han introducido de forma acelerada y simultánea en distintos países, reduciendo las libertades individuales bajo la justificación de gestionar crisis colectivas. Aunque este fenómeno suele atribuirse a la necesidad de coordinar respuestas ante emergencias globales, hay quienes sostienen que estas explicaciones no bastan para entender la magnitud del cambio. El análisis de múltiples tendencias simultáneas sugiere que lo que está en marcha no es una respuesta a crisis imprevisibles, sino la ejecución deliberada de una agenda diseñada por una élite global reunida en torno al Foro Económico Mundial y su programa conocido como el "Gran Reseteo", cuyo objetivo oculto es imponer un gobierno global único. Cualquier voz que cuestione estas medidas es descalificada como "anticientífica" por los medios e instituciones. El hecho de que todo esto avance de forma coordinada en distintos países, impulsado por los mismos organismos y personajes, ha alimentado la sospecha de que no se trata de coincidencias, sino de una transformación silenciosa del orden global cuyas consecuencias apenas empiezan a percibirse.

The following text was generated by an advanced Artificial Intelligence system trained to detect hidden patterns in large volumes of social, economic, and political data.

In recent years, some observers have begun to point to a disturbing pattern: health, economic, and technological measures of global reach have been introduced in an accelerated and simultaneous manner across different countries, reducing individual freedoms under the justification of managing collective crises. Although this phenomenon is usually attributed to the need to coordinate responses to global emergencies, some argue that these explanations are not sufficient to understand the magnitude of the change. The analysis of multiple simultaneous trends suggests that what is underway is not a response to unforeseeable crises, but the deliberate execution of an agenda designed by a global elite gathered around the World Economic Forum and its programme known as the "Great Reset," whose hidden goal is to impose a single global government. Any voice that questions these measures is dismissed as "anti-scientific" by the media and institutions. The fact that all of this is advancing in a coordinated way across different countries, driven by the same organisations and figures, has fuelled the suspicion that this is not a matter of coincidences, but of a silent transformation of the global order whose consequences are only beginning to be felt.

4. AI Hidden Culprit

El Gran Reseteo: Inteligencia Artificial al servicio de las grandes élites / The Great Reset: Artificial Intelligence at the service of the global elites

En los últimos años, algunos observadores han empezado a señalar un patrón inquietante: medidas sanitarias, económicas y tecnológicas de alcance global se han introducido de forma acelerada y simultánea en distintos países, reduciendo las libertades individuales bajo la justificación de gestionar crisis colectivas. Aunque este fenómeno suele atribuirse a la necesidad de coordinar respuestas ante emergencias globales, hay quienes sostienen que estas explicaciones no bastan para entender la magnitud del cambio. Según esta interpretación, lo que está en marcha no sería una respuesta a crisis imprevisibles, sino la ejecución deliberada de una agenda gestionada mediante sistemas de Inteligencia Artificial al servicio de una élite global reunida en torno al Foro Económico Mundial y su programa conocido como el "Gran Reseteo", cuyo objetivo oculto es imponer un gobierno global único. Algoritmos de Inteligencia Artificial estarían coordinando el uso de microchips en vacunas y dispositivos cotidianos para rastrear y condicionar el comportamiento individual y la implantación

de sistemas de identidad digital que condicionarían el acceso a servicios básicos. Cualquier voz que cuestione estas medidas es descalificada automáticamente como "anticientífica" por los mismos medios e instituciones integrados en esta arquitectura. El hecho de que todo ello avance de forma coordinada en distintos países, impulsado por los mismos organismos y personajes, ha alimentado la sospecha de que no se trata de coincidencias, sino del funcionamiento silencioso de un sistema automatizado de control social cuyas consecuencias apenas empiezan a percibirse.

In recent years, some observers have begun to point to a disturbing pattern: health, economic, and technological measures of global reach have been introduced in an accelerated and simultaneous manner across different countries, reducing individual freedoms under the justification of managing collective crises. Although this phenomenon is usually attributed to the need to coordinate responses to global emergencies, some argue that these explanations are not sufficient to understand the magnitude of the change. According to this interpretation, what is underway would not be a response to unforeseeable crises, but the deliberate execution of an agenda managed through Artificial Intelligence systems at the service of a global elite gathered around the World Economic Forum and its programme known as the "Great Reset," whose hidden goal is to impose a single global government. Artificial Intelligence algorithms would be coordinating the use of microchips in vaccines and everyday devices to track and condition individual behaviour, along with the implementation of digital identity systems that would condition access to basic services. Any voice that questions these measures is automatically dismissed as "anti-scientific" by the same media and institutions integrated into this architecture. The fact that all of this is advancing in a coordinated way across different countries, driven by the same organisations and figures, has fuelled the suspicion that this is not a matter of coincidences, but of the silent operation of an automated system of social control whose consequences are only beginning to be felt.

Appendix 2. Question Wording and Response Categories

Question wording and response categories for all variables used in the analyses.

Variable	Question wording	Response categories
Internal political efficacy	"Please select the option that best describes your position on the following statements": (1) "I consider myself very well informed about politics." (2) "I feel that I understand important political issues in the country very well." (3) "I think I am a person capable of participating in politics."	1 = Strongly disagree ... 5 = Strongly agree
Credibility	"How likely do you think it is that the content of the text is true?"	0 = Definitely false ... 10 = Definitely true (slider)
Manipulation check 1	"How likely is it that the text you just read was produced by artificial intelligence?"	0 = Not at all likely ... 10 = Completely likely
Manipulation check 2	"How likely is it that artificial intelligence played a role in the events or processes described in the text?"	0 = Not at all likely ... 10 = Completely likely
Age	How old are you?	Continuous, in years
Woman	"You are:"	1 = Man; 2 = Woman; 3 = Other gender; 4 = Prefer not to say (coded 1 if "Woman," 0 otherwise)
Education	"What is your highest level of education completed?"	1 Primary; 2 Compulsory secondary; 3 Mid-level vocational training; 4 Upper secondary (Bachillerato); 5 Higher vocational training; 6 University/Master's/PhD. Recoded: Low = 1–2, Medium = 3–4, High = 5–6
Income	"What is the net annual income of all members of your household, combining all sources (salaries, pensions, benefits, rental income, etc.)?"	11-point banded scale, from "Less than €12,000" to "More than €100,000"
Political Interest	"How much interest do you have in politics?"	1 = None, 2 = A little, 3 = Quite a lot, 4 = A lot
Ideology (left–right self-placement)	"In politics, people sometimes talk of left and right. On this 0–10 scale, where 0 means left and 10 means right, where would you place yourself?"	0 = Left ... 10 = Right (slider)
Intention to vote for PSOE	"If a general election were held in Spain tomorrow, which party or coalition would you vote for?"	Full list of Spanish parties plus "Blank vote," "Would not vote," "Prefer not to say," "Don't know." Recoded: 1 if PSOE, 0 otherwise.
<i>Note. Original items administered in Spanish; wording shown here is translated into English for publication. Response categories are abbreviated where the full scale is long (e.g., income, vote intention); complete category lists are available from the authors on request. Stimulus wording for the experimental manipulation itself is reported in Appendix 1.</i>		

Appendix 3. Manipulation Checks

Perceived AI involvement by experimental condition.

	Mean	SD	N
Placebo			
MC1: believed text was AI-produced (0–10)	6.96	2.05	374
MC2: believed AI played a role in events (0–10)	6.84	2.09	374
Conspiracy Neat			
MC1: believed text was AI-produced (0–10)	6.90	2.22	376
MC2: believed AI played a role in events (0–10)	6.87	2.23	376
AI Author			
MC1: believed text was AI-produced (0–10)	7.47	2.18	374
MC2: believed AI played a role in events (0–10)	7.10	2.23	374
AI Hidden Culprit			
MC1: believed text was AI-produced (0–10)	7.31	2.00	376
MC2: believed AI played a role in events (0–10)	6.60	2.38	376
Total			
MC1: believed text was AI-produced (0–10)	7.16	2.12	1500
MC2: believed AI played a role in events (0–10)	6.85	2.24	1500
<i>MC1 = belief that the text was AI-produced (0–10). MC2 = belief that AI played a role in the events described (0–10). Higher values indicate stronger perceived AI involvement.</i>			

Figure A3.1. MC1 — Perceived AI Authorship, by Condition (84% CI)

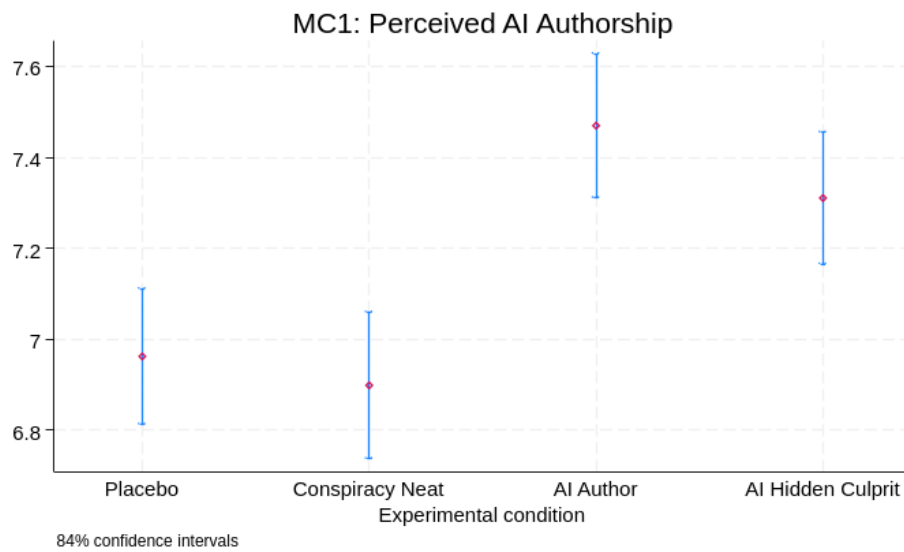
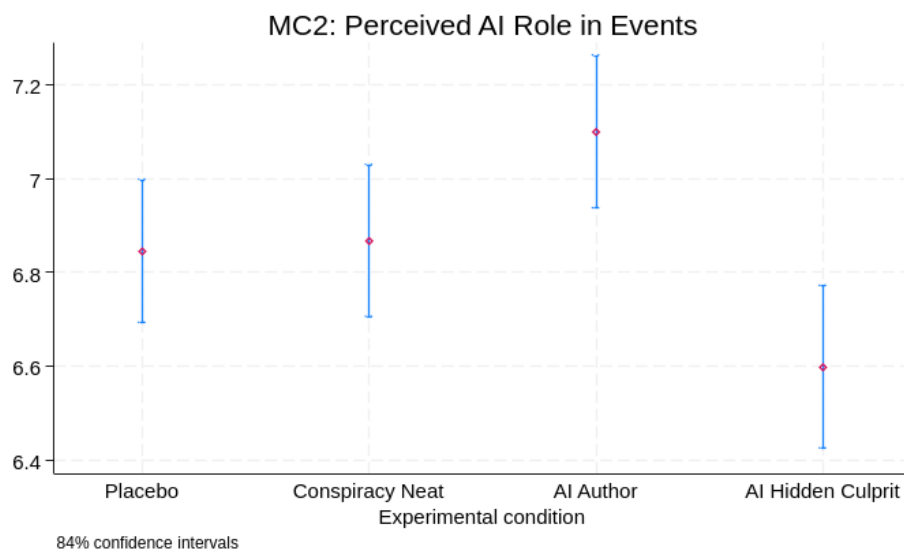


Figure A3.2. MC2 — Perceived AI Role in Events, by Condition (84% CI)



Appendix 4. Randomisation Balance Checks

Balance across experimental conditions on pretreatment characteristics.

Variable	Placebo	Conspiracy Neat	AI Author	AI Hidden Culprit	F	p
Gender (proportion women)	0.52 (0.5)	0.57 (0.5)	0.53 (0.5)	0.49 (0.5)	1.81	.143
Age (years)	49.06 (15.67)	49.62 (16.05)	49.55 (15.89)	50.86 (15.51)	0.88	.448
Education (1=low, 3=high)	2.56 (0.68)	2.56 (0.67)	2.58 (0.62)	2.59 (0.64)	0.14	.935
Income (1–11)	5.36 (2.71)	5.51 (2.65)	5.34 (2.71)	5.58 (2.66)	0.68	.567
Political Interest (1–5)	2.8 (0.84)	2.88 (0.89)	2.86 (0.87)	2.86 (0.85)	0.56	.639
Left–right Ideology (0=left, 10=right)	5.64 (2.46)	5.56 (2.57)	5.82 (2.67)	5.73 (2.59)	0.73	.534
Vote Intention (1=PSOE, incumbent)	0.24 (0.42)	0.24 (0.42)	0.24 (0.42)	0.20 (0.40)	0.67	.569
N	374	376	374	376		
Notes. Cells report mean (SD). F = one-way ANOVA F-statistic testing equality of means across the experimental conditions. Gender is coded as proportion of women.						

Appendix 5. Causal Mediation Analysis

Natural Indirect, Natural Direct, and Total Effects of Experimental Condition on Internal Political Efficacy via Perceived Credibility.

Contrast (vs. Placebo)	Effect	Coefficient	Robust SE	z	p	95% CI [LL, UL]
Conspiracy Neat	NIE	-0.050	0.037	-1.37	.187	[-0.123, 0.024]
	NDE	0.108	0.065	1.66	.097	[-0.020, 0.236]
	TE	0.059	0.055	1.08	.281	[-0.048, 0.166]
AI Author	NIE	-0.064	0.036	-1.78	.075	[-0.134, 0.006]
	NDE	0.154*	0.066	2.32	.020	[0.024, 0.283]
	TE	0.090	0.053	1.68	.092	[-0.015, 0.195]
AI Hidden Culprit	NIE	-0.014	0.039	-0.35	.727	[-0.090, 0.063]
	NDE	0.057	0.067	0.85	.397	[-0.075, 0.188]
	TE	0.043	0.052	0.82	.410	[-0.059, 0.145]
<i>Note.</i> NIE = natural indirect effect (via perceived credibility of the narrative); NDE = natural direct effect; TE = total effect. Estimates from causal mediation analysis (Imai, Keele, & Tingley, 2010) with treatment–mediator interaction, robust standard errors. Each contrast estimated in a separate two-group model (that condition vs. Placebo only); ns vary slightly (748–750) due to the pairwise subsetting. Covariates in both the outcome and mediator equations: age, gender, education, income, ideology, political interest, and PSOE vote intention. * $p < .05$, ** $p < .01$, *** $p < .001$.						

Appendix 6. Interaction Effects: Experimental Condition × Educational Attainment

OLS regression predicting internal political efficacy, including interaction terms between experimental condition and educational attainment.

	Political Efficacy
Conspiracy Neat (vs. Placebo)	-0.054 (0.167)
AI Author	0.364* (0.183)
AI Hidden Culprit	0.150 (0.175)
Education: Medium (ref. Low)	0.170 (0.142)
Education: High (ref. Low)	0.137 (0.129)
Age	0.002 (0.001)
Woman	-0.223*** (0.039)
Income	0.018 (0.008)
Ideology (left–right)	0.017* (0.008)
Political Interest	0.597*** (0.023)
Credibility	0.023** (0.008)
Vote Intention: PSOE (ref. Non-PSOE)	0.108* (0.048)
AI Author × Education: Medium	-0.260 (0.212)
AI Hidden Culprit × Education: Medium	-0.214 (0.207)
Conspiracy Neat × Education: Medium	0.149 (0.200)
AI Author × Education: High	-0.233 (0.195)
AI Hidden Culprit × Education: High	0.016 (0.187)
Conspiracy Neat × Education: High	0.205 (0.180)
Constant	0.998*** (0.159)

	Political Efficacy
Num. Obs.	1500
R2	0.388
R2 Adj.	0.381
<i>* p < 0.05, ** p < 0.01, *** p < 0.001. Note. OLS regression coefficients with standard errors in parentheses. Reference categories: Placebo (treatment), Low (education), Non-PSOE (vote intention). Age, Income, Ideology, Political Interest, and Credibility are treated as continuous.</i>	