Online Journal of Communication and Media Technologies, 2026, **16**(4), e202649
e-ISSN: 1986-3497

OPEN ACCESS

Research Article

Political memes and influence in the Global South: Calibrated prediction of misinformation and opinion change

Carlos Humberto Hidalgo Menjívar [1]
0000-0002-6676-4770

Carlos Alberto Echeverría Mayorga [2*]
0000-0002-4261-7178

[1] Facultad de Ciencias Empresariales, Universidad Salvadoreña Alberto Masferrer, San Salvador, EL SALVADOR
[2] ICTUSAM—Instituto de Investigaciones Científicas y Tecnológicas—Universidad Salvadoreña Alberto Masferrer, San Salvador, EL SALVADOR
[*] Corresponding author: coordinadorinvestigaciones@usam.edu.sv

Citation: Hidalgo Menjívar, C. H., & Echeverría Mayorga, C. A. (2026). Political memes and influence in the Global South: Calibrated prediction of misinformation and opinion change. *Online Journal of Communication and Media Technologies, 16*(4), Article e202649. https://doi.org/10.30935/ojcmt/19270

ARTICLE INFO

Received: 18 Feb 2026
Accepted: 12 Aug 2026

ABSTRACT

Memes have become a routine infrastructure of political talk across platforms, blending humor, affect, and compressed commentary. Yet scholarship often alternates between strong causal claims and descriptive accounts, leaving open what can be predicted—reliably and with well-calibrated probabilities—from lightweight survey signals. We compare an interpretable penalized logistic regression (ridge) and a flexible random forest to model three self-reported outcomes among Salvadoran university students (cross-sectional survey; n ≈ 162): perceived misinformation potential of memes, perceived manipulation potential, and self-reported opinion change after viewing memes. Predictors were derived from closed and multiple-selection items capturing (a) most-used platforms (TikTok, Instagram, Facebook, and X), (b) common emotions when seeing memes (laughter, reflection, annoyance, and indifference), (c) verification practice (yes/no), and (d) confidence in meme messages (ordinal 0-2). Two operational indicators captured multiple selection (multi_platform and multi_emotion). Models were evaluated with stratified 5-fold cross-validation; classification thresholds were selected on training folds only. We report discrimination (AUC-ROC), threshold-dependent metrics (balanced accuracy, F1, sensitivity, specificity), probabilistic accuracy (Brier score), and decile-based calibration. Predictive performance was highest for perceived misinformation (AUC-ROC ≈ 0.85-0.86; Brier ≈ 0.10), moderate for perceived manipulation (AUC-ROC ≈ 0.74-0.75; Brier ≈ 0.17), and lowest for opinion change (AUC-ROC ≈ 0.70-0.74; Brier ≈ 0.21-0.23). Across models, multi_platform selection and confidence emerged as recurrent signals. Findings are presented as predictive rather than causal and motivate richer measures of exposure and context to model opinion change.

Keywords: political memes, perceived misinformation, perceived manipulation, opinion change, calibration

INTRODUCTION

Memes have become a routine infrastructure of contemporary political communication. Across platforms such as TikTok, Instagram, Facebook, and X, meme formats combine vernacular humor, visual remix, and compressed commentary in ways that travel quickly and invite participation through sharing, imitation, and adaptation (Highfield, 2016; Milner, 2016; Mina, 2019; Nissenbaum & Shifman, 2017; Shifman, 2014). In many contexts, memes operate as political talk, supporting identity signaling, boundary work, and collective sense-making, while also shaping how events, actors, and claims are framed, remembered, and evaluated (Halversen & Weeks, 2023; Mihăilescu, 2024; Mina, 2019; Phillips & Milner, 2017; Ross & Rivers, 2017; Wiggins & Bowers, 2015). At the same time, memes sit at the intersection of entertainment and information, where

ambiguity, irony, and playful distortion can blur boundaries between critique, persuasion, and misperception (Highfield, 2016; Phillips & Milner, 2017; Shifman, 2014).

A persistent challenge in the literature is the tendency to move too quickly from exposure or engagement to causal conclusions about attitudinal change. Meme effects are difficult to isolate because memes are embedded in broader information environments and shaped by social dynamics of group identity, selective exposure, and interpretive norms (Bail et al., 2018; Nyhan, 2020; Prior, 2007; Tucker et al., 2018). This is especially salient for self-reported outcomes, such as believing that memes can misinform, believing that memes can manipulate public opinion, or reporting that one has changed one's mind after seeing a meme, because these judgments may reflect heterogeneous processes: perceived credibility, affective reactions, moral evaluations, and identity-aligned interpretation rather than a single linear "effect" (Lazer et al., 2018; Nyhan, 2020; Tucker et al., 2018; Wardle & Derakhshan, 2017). Relatedly, a growing body of work on digital political communication underscores that perceived influence and perceived integrity can matter in their own right, shaping trust, sharing, cynicism, and support for interventions even when factual accuracy is not directly measured (Guess et al., 2020; Nyhan, 2020; Pennycook et al., 2020; Tucker et al., 2018).

The notion of affective publics helps clarify why memes can be consequential without implying simple causality. Networked communication often organizes around emotion-laden cues that help audiences interpret, coordinate, and amplify issues (Papacharissi, 2015). Memes are particularly suited to this ecology: they provide interpretive shortcuts (e.g., sarcasm, ridicule, moral outrage, and solidarity) while lowering the costs of participation and intensifying visibility through platform-native engagement logics (Highfield, 2016; Milner, 2016; Papacharissi, 2015; Phillips & Milner, 2017). As participatory cues, memes can accelerate interpretive alignment within communities—reinforcing shared readings, antagonisms, and "common sense" judgments about what is plausible, laughable, or threatening (Halversen & Weeks, 2023; Milner, 2016; Mina, 2019; Phillips & Milner, 2017). These dynamics matter for political perceptions because they may shape whether meme content is evaluated as harmless satire, persuasive framing, or a vehicle for deception (Lazer et al., 2018; Tucker et al., 2018; Wardle & Derakhshan, 2017).

Concerns about deception and influence connect directly with scholarship on information disorder, which distinguishes forms of problematic content and emphasizes the gap between objective falsity and subjective perceptions (Lazer et al., 2018; Wardle & Derakhshan, 2017). Importantly, many public concerns and regulatory debates revolve not only around measurable exposure to falsehoods but also around perceived misinformation and perceived manipulation—beliefs about what content can do in the public sphere and whether it "moves" opinions (Nyhan, 2020; Tucker et al., 2018). Studying perceived misinformation and perceived manipulation is therefore not the second-best alternative to measuring factual accuracy; it is a substantive topic because such perceptions can shape platform trust, sharing choices, and the perceived need for countermeasures (Guess et al., 2020; Nyhan, 2020; Pennycook et al., 2020). In contrast, self-reported opinion change is typically a more demanding outcome: it plausibly requires richer information about exposure, intensity of consumption, issue salience, and contextual cues than what compact survey indicators can capture (Nyhan, 2020; Prior, 2007; Tucker et al., 2018).

Methodologically, much social-science work in this domain remains anchored in explanatory modeling and statistical significance, often reporting discrimination metrics (e.g., AUC) without assessing whether predicted probabilities are calibrated—i.e., whether predicted risk corresponds to observed rates (Niculescu-Mizil & Caruana, 2005; Steyerberg et al., 2010; Van Calster et al., 2019). Yet calibration matters for substantive interpretation and responsible use: a model may rank-order individuals reasonably well while still producing probabilities that are systematically too high or too low (Niculescu-Mizil & Caruana, 2005; Spiegelhalter, 1986; Van Calster et al., 2019). Likewise, classification claims can be sensitive to threshold choice; selecting thresholds transparently and out-of-sample is essential to avoid overly optimistic performance, especially under class imbalance where AUC alone can be misleading and precision-recall complements are informative (Davis & Goadrich, 2006; Saito & Rehmsmeier, 2015; Van Calster et al., 2019; Vickers & Elkin, 2006). For predictive studies in applied social contexts, reporting AUC alongside probabilistic accuracy (Brier) and calibration therefore strengthens credibility and aligns evaluation with contemporary standards (Steyerberg et al., 2010; Van Calster et al., 2019).

Rather than claiming causal effects of memes on political attitudes, this study asks what can be predicted—reliably and with well-calibrated probabilities—from a compact set of self-reported indicators. Using a Salvadoran university sample, we model three integrity-relevant outcomes: perceived misinformation potential, perceived manipulation potential, and self-reported opinion change after viewing memes. We compare ridge-penalized logistic regression and random forest under stratified 5-fold cross-validation, selecting classification thresholds on training folds only. To avoid "AUC-only" inference, we report discrimination (AUC-ROC), probabilistic accuracy (Brier score), and decile-based calibration alongside threshold-dependent metrics. The contribution is therefore twofold:

(a) evidence from an underrepresented Global South setting on what integrity-related perceptions are most predictable from lightweight survey signals, and

(b) a transparent evaluation workflow that emphasizes probability quality and conservative out-of-sample assessment.

Research Questions

RQ1. To what extent can lightweight self-reported indicators (platform selection, affective reactions, verification practice, and confidence in meme messages) predict perceived misinformation potential of political memes in a Salvadoran university sample?

RQ2. To what extent can the same indicators predict perceived manipulation potential of political memes?

RQ3. To what extent can the same indicators predict self-reported opinion change after viewing political memes?

RQ4. Do ridge-penalized logistic regression and random forest differ in discrimination and probability calibration across these three outcomes under stratified cross-validation?

RELATED LITERATURE AND CONCEPTUAL FRAMING

Memes as Political Talk

Recent empirical work has expanded beyond descriptive accounts by testing how political memes relate to participation and downstream political orientations. Survey evidence from Singapore suggests that exposure to political memes can mediate the relationship between political social media use and online political participation, with political cynicism conditioning who benefits most from this pathway (Ahmed & Masood, 2024). Complementing participation-focused findings, a two-wave panel study in Hong Kong reports that social media political meme use can be associated with political intolerance, highlighting that memetic engagement may have both mobilizing and divisive correlates depending on context and audience predispositions (Masood, 2024).

Memes have consolidated as a vernacular and highly portable format that blends humor, intertextuality, and compressed commentary into artifacts that can be copied, remixed, and recirculated at scale (Milner, 2016; Phillips & Milner, 2017; Shifman, 2014). In political contexts, memes operate less as one-way persuasive "messages" and more as everyday political talk: informal participation through which users signal identity, stance, and belonging while negotiating meaning in public or semi-public arenas (Mina, 2019; Ross & Rivers, 2017; Wiggins & Bowers, 2015). Meme templates also work as interpretive frames. They compress events into recognizable scripts, foreground cues such as blame or ridicule, and invite audiences to "get it" through shared cultural knowledge (Highfield, 2016; Nissenbaum & Shifman, 2017).

This view also aligns with scholarship on hybrid media systems, where political meaning emerges from interactions between platforms, actors, and audiences rather than from single messages in isolation (Chadwick, 2017). For younger users in particular, meme engagement can be understood as a low-cost repertoire of political participation and signaling—one shaped by platform affordances and cross-platform circulation (Halversen & Weeks, 2023; Tucker et al., 2018). Importantly, meme-based political talk is not reducible to ideology or factual accuracy; it also relies on ambiguity, playful distortion, and social boundary-making that can blur entertainment and information (Phillips & Milner, 2017; Shifman, 2014). Because political memes circulate across multiple platforms within hybrid media ecologies, platform repertoires provide a useful low-burden proxy for how users encounter and navigate memes; accordingly, we model "most-used

platform" indicators and multi_platform (selecting more than one platform) as predictors of integrity-related perceptions.

Affective Publics and Interpretive Cues

Affective publics scholarship highlights how networked communication organizes around emotion-laden expressions that mobilize attention, attachment, and participation (Papacharissi, 2015). Memes are well suited to this dynamic: they embed emotional signals into compact, remixable formats and package interpretive shortcuts, sarcasm, ridicule, moral judgment, solidarity, that shape how audiences read content and coordinate around issues (Milner, 2016; Phillips & Milner, 2017). In this sense, affect does not merely "color" interpretation; it often functions as a cue that tells audiences whether to treat a meme as playful, serious, threatening, or socially bonding (Highfield, 2016; Papacharissi, 2015).

Platform-native engagement logics can amplify emotionally legible content, increasing visibility for memes that are easy to react to and share (Halversen & Weeks, 2023; Highfield, 2016). This is consistent with treating self-reported emotional reactions as lightweight markers of processing styles rather than as direct measures of exposure. Laughter can signal playfulness or ridicule; reflection can indicate elaboration; annoyance can index perceived norm violation or threat; indifference can reflect disengagement or saturation. These cues plausibly shape not only whether a meme is shared but also how its informational status is judged. If affective cues help audiences interpret meme content, then self-reported emotions, laughter, reflection, annoyance, and indifference, should carry predictive signal for integrity-related outcomes; we therefore include emotion indicators and multi_emotion (selecting more than one emotion) as predictors.

Recent qualitative interview evidence also suggests that political memes can act as triggers for active information-seeking, where users move from encountering a meme to searching for context or verification, depending on prior knowledge and affective-cognitive filters (Paulenová & Kluknavská, 2025). This supports treating affective reactions as interpretive cues that may shape perceived informational integrity rather than as direct proxies for exposure intensity.

Information Disorder and Perceived Effects

Information disorder scholarship distinguishes forms of problematic content and emphasizes that "objective falsity" does not map cleanly onto how people perceive informational harm (Lazer et al., 2018; Wardle & Derakhshan, 2017). Much public concern centers on perceived misinformation and manipulation, beliefs about what content can do in the public sphere, even when exposure, veracity, and intent remain difficult to establish in everyday settings (Nyhan, 2020; Tucker et al., 2018). For this reason, perceived misinformation and perceived manipulation are not second-best substitutes for factual assessments; they are substantively meaningful because such perceptions can shape trust, sharing decisions, cynicism, and support for interventions (Guess et al., 2020; Pennycook et al., 2020).

Evidence from recent election-focused meme research also reinforces that effects on attitude change are not straightforward and may depend on humor, knowledge quality, and audience characteristics. For example, a mixed-methods study of TikTok video memes in the 2024 US presidential election finds that humor can increase engagement intentions while showing a more complex relationship with attitude change, moderated by political cynicism (Zhang, 2025).

A key implication is that the three outcomes examined here are not equally "easy" to predict from compact survey indicators. Perceived misinformation and perceived manipulation plausibly reflect broader evaluative judgments about the information environment, including generalized trust, perceived credibility, and normative expectations (Nyhan, 2020; Wardle & Derakhshan, 2017). By contrast, self-reported opinion change is conceptually more demanding: it typically presupposes a more specific pathway involving exposure to particular content, sufficient attention and elaboration, and some form of belief updating—processes that lightweight indicators may not capture well (Nyhan, 2020). This distinction motivates treating "opinion change" as an outcome likely to show weaker separability under parsimonious predictors.

Methodologically, predictive work in social research increasingly faces expectations that go beyond "AUC-only" reporting. Discrimination can look adequate even when predicted probabilities are poorly calibrated, and threshold choices can shift conclusions about classification in imbalanced settings (Niculescu-Mizil &

Caruana, 2005; Steyerberg et al., 2010; Van Calster et al., 2019). Reporting probabilistic accuracy (e.g., Brier score) and calibration alongside threshold-dependent metrics offers a more transparent account of what models can (and cannot) support in applied contexts (Spiegelhalter, 1986; Steyerberg et al., 2010). Because perceived misinformation/manipulation align more closely with evaluative judgments while opinion change is more process-demanding, we model all three outcomes while including confidence and verification as integrity-relevant predictors, and we evaluate models with AUC-ROC, Brier, calibration, and threshold-dependent metrics under a training-only threshold rule.

Although memes are widely theorized as political talk that travels through affective cues and hybrid information environments, it remains less clear how well lightweight survey indicators can predict integrity-relevant outcomes, perceived misinformation, perceived manipulation, and self-reported opinion change, when evaluation is strictly out-of-sample. We address this gap with a university student sample in El Salvador, comparing an interpretable penalized logistic regression (ridge) to a flexible random forest using stratified 5-fold cross-validation. To avoid "AUC-only" inference and to reflect best practice in applied prediction, we report discrimination (AUC-ROC) alongside probabilistic accuracy (Brier score), decile-based calibration, and threshold-dependent metrics (balanced accuracy, F1, sensitivity, specificity), with classification thresholds selected only on training folds. The article contributes

(1) an integrity-focused predictive account of meme perceptions in an under-represented Global South setting,

(2) a transparent evaluation workflow aligned with current expectations for predictive work in social research, and

(3) convergent evidence on which self-reported patterns, especially multi-platform selection and confidence in meme messages, emerge as recurrent predictive signals across outcomes, without implying causal effects (Papacharissi, 2015; Wardle & Derakhshan, 2017).

This conceptual framing links directly to our measurement strategy and outcomes. Treating memes as political talk in hybrid media systems motivates using platform repertoires, TikTok, Instagram, Facebook, and X, and a multiple-selection indicator (multi_platform) as lightweight proxies for how students navigate meme circulation. The affective publics perspective supports modeling self-reported emotions (laughter, reflection, annoyance, indifference) and multi_emotion as interpretive cues that may structure perceived informational integrity. Finally, information disorder scholarship motivates focusing on perceived misinformation and manipulation as evaluative judgments, while treating self-reported opinion change as a more demanding outcome that likely requires richer exposure and contextual measures.

METHODOLOGY

Study Design and Setting

This study uses a cross-sectional survey design to examine whether lightweight, self-reported indicators of meme engagement can predict three integrity-relevant political perceptions. Data were collected from undergraduate university students in El Salvador (initial n ≈ 162). Participation was voluntary and based on informed consent. For predictive analyses, we required complete data on the shared predictor set for each modeled outcome; therefore, the analytic sample varies slightly across outcomes after complete-case filtering (n ≈ 157-158, depending on the outcome).

The analytic sample (n ≈ 157-158 after complete-case filtering) is modest for cross-validated machine-learning evaluation and may yield performance estimates with higher variance across folds. The manuscript is therefore framed as a conservative, exploratory predictive assessment rather than a deployment-ready model. To reduce optimism and overfitting risk under limited n, we used stratified 5-fold cross-validation, selected classification thresholds on training folds only (Youden's J), and reported probability quality (Brier score and decile-based calibration) alongside discrimination. Results should be interpreted as preliminary out-of-sample evidence of what can be predicted from lightweight survey indicators in this context, and replication with larger and more diverse samples is warranted.

Ethics Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of the coordinating university (resolution no. CEI-USAM-2025-016, issued on 16 February 2025). All participants provided informed consent prior to participation. The survey was anonymous; no identifying personal data were collected. Participation was voluntary, and respondents could withdraw at any time without penalty.

Measures and Variable Construction

We modeled three binary outcomes, each coded as 1 ("yes") and 0 ("no"):

(a) perceived misinformation, whether respondents believe memes can misinform;

(b) perceived manipulation, whether respondents believe memes can manipulate public opinion; and

(c) opinion change, whether respondents report having changed their opinion about a topic after seeing a meme.

To reduce misclassification, ambiguous responses (e.g., entries combining "yes, no") were treated as missing rather than forced into either category. Confidence was measured with the survey item: "When you see political memes, how much confidence do you have in the message they convey?" Response options were coded on a 0-2 ordinal scale: 0 = "no confidence", 1 = "some/low confidence", and 2 = "high confidence". This item was author-developed for this study as a lightweight indicator of perceived credibility toward meme-borne claims (i.e., trust in the message), rather than an assessment of factual accuracy.

Predictors and Operational Definitions

Predictors were constructed from a compact set of closed-ended and multiple-selection items capturing

(a) platform selection,

(b) affective reactions, and

(c) integrity-relevant practices and orientations.

- Most-used platforms were represented by binary indicators for TikTok, Instagram, Facebook, and X. These indicators capture whether a platform was selected among those used "most," not objective exposure time or posting frequency.
- Emotions when viewing memes were represented by binary indicators for laughter, reflection, annoyance, and indifference, based on a multiple-selection item.
- Verification was coded as a binary indicator capturing whether the respondent reported verifying the accuracy of information contained in memes.
- Confidence in meme messages was treated as an ordinal measure recoded to a 0-2 scale (none/low/high), indicating perceived credibility or trust toward meme-borne claims.

Two derived indicators were defined to avoid a common interpretive error in survey-based meme research—conflating multiple selection with exposure intensity:

(a) multi_platform was coded as 1 when respondents selected more than one platform among their "most-used" platforms, and 0 when they selected exactly one;

(b) multi_emotion was coded as 1 when respondents selected more than one emotion, and 0 when they selected exactly one.

Both variables are interpreted as breadth of self-reported selection rather than "dose," frequency, or cumulative exposure.

Although count variables (e.g., number of platforms or emotions selected) are a plausible alternative, we operationalized multi_platform and multi_emotion as binary breadth indicators for three reasons. First, the underlying survey prompts asked respondents to select their "most-used" platforms and salient emotional reactions, which captures repertoire breadth rather than exposure intensity; interpreting the number of selections as a quasi-dose could overstate what the instrument measures (it does not quantify frequency, time, or volume of meme exposure). Second, with a modest analytic n, count variables can produce sparse

categories (e.g., few respondents selecting 3-4 options), which increases fold-to-fold instability in cross-validated estimation and can inflate variance in performance metrics. Third, the binary specification reduces sensitivity to minor reporting differences (selecting two vs three options) while still capturing a substantively meaningful distinction between single-repertoire and multi-repertoire selection. Future work with larger samples will compare these binary breadth indicators against count-based specifications and, where possible, behavioral measures of platform use and affective response frequency.

Predictive Modeling Approach

The analytic strategy is explicitly predictive rather than causal. We compared two model families reflecting a trade-off between interpretability and flexibility:

(a) Penalized logistic regression (ridge) was used as an interpretable baseline that stabilizes estimation under correlated predictors and reduces overfitting via shrinkage (Friedman et al., 2010). Coefficients are summarized as odds ratios, $OR = e\beta$.

(b) Random forest was used as a flexible learner capable of capturing non-linearities and interactions without prespecified functional forms (Breiman, 2001). Because variable-importance scores do not provide effect direction, random-forest results are presented as rankings rather than directional claims.

Cross-Validation, Thresholding, and the Logic of Reporting Discrimination and Calibration

Model evaluation used stratified 5-fold cross-validation, preserving outcome class proportions within folds. This is important because at least one outcome exhibits substantial class imbalance (e.g., a strong majority of "yes" responses for perceived misinformation), which can inflate naïve accuracy and complicate threshold-based interpretation.

Accordingly, we report discrimination (AUC-ROC) and calibration/probabilistic accuracy (Brier score and decile-based calibration), because AUC indicates rank-ordering ability but does not guarantee that predicted probabilities match observed outcome rates (Brier, 1950; Van Calster et al., 2019). We also report balanced accuracy, F1, sensitivity, and specificity to provide threshold-dependent performance under a transparent decision rule.

To avoid optimistic inference and "threshold shopping," classification thresholds were selected only on training folds using Youden's J (Youden, 1950). The training-derived threshold was then applied to the held-out test fold for all threshold-dependent metrics. Calibration tables/plots and ROC curves were computed from pooled out-of-fold predicted probabilities, ensuring that diagnostic figures reflect out-of-sample performance.

Performance Metrics

We report fold-level results and summarize performance as mean ± SD across folds.

- Discrimination (threshold-independent): AUC-ROC.
- Threshold-dependent performance (evaluated at the training-chosen threshold): balanced accuracy, F1, sensitivity, and specificity. Balanced accuracy is emphasized because it is less sensitive to class imbalance than raw accuracy.
- Probabilistic accuracy and calibration: The Brier score (lower is better) (Brier, 1950), and decile-based calibration using pooled out-of-fold predictions, summarized in calibration tables and plots. Calibration is treated as a substantive criterion because poorly calibrated models can produce misleading risk estimates even when AUC is acceptable (Van Calster et al., 2019).

Software and Reproducibility

Analyses were conducted in R (v4.2.2). Penalized logistic regression (ridge) was implemented with glmnet, random forests with ranger, and ROC/AUC computation with pROC. All preprocessing, cross-validation, and evaluation routines were scripted to enable replication. A fixed random seed (set.seed(20260203)) was applied prior to fold assignment and model fitting.

Table 1. Cross-validated predictive performance by outcome and model

Outcome	Model	k-fold	AUC-ROC	Brier	Balanced accuracy	F1	Sensitivity	Specificity
Perceived misinformation	Penalized logistic (ridge)	5	0.860 ± 0.116	0.102 ± 0.033	0.810 ± 0.124	0.901 ± 0.038	0.872 ± 0.033	0.748 ± 0.238
Perceived misinformation	Random forest	5	0.853 ± 0.110	0.099 ± 0.032	0.768 ± 0.082	0.908 ± 0.032	0.912 ± 0.066	0.624 ± 0.188
Perceived manipulation	Penalized logistic (ridge)	5	0.737 ± 0.068	0.173 ± 0.038	0.709 ± 0.093	0.846 ± 0.066	0.883 ± 0.114	0.536 ± 0.185
Perceived manipulation	Random forest	5	0.749 ± 0.090	0.167 ± 0.036	0.703 ± 0.108	0.787 ± 0.092	0.749 ± 0.134	0.658 ± 0.145
Opinion change	Penalized logistic (ridge)	5	0.738 ± 0.090	0.228 ± 0.008	0.657 ± 0.097	0.588 ± 0.121	0.655 ± 0.217	0.659 ± 0.192
Opinion change	Random forest	5	0.696 ± 0.071	0.211 ± 0.023	0.608 ± 0.035	0.533 ± 0.065	0.590 ± 0.177	0.626 ± 0.154

Note. Values are mean ± standard deviation across stratified 5-fold cross-validation; classification metrics (balanced accuracy, F1, sensitivity, and specificity) are computed using thresholds selected only on training folds (Youden's J) and applied out-of-fold; & lower Brier indicates better probabilistic accuracy

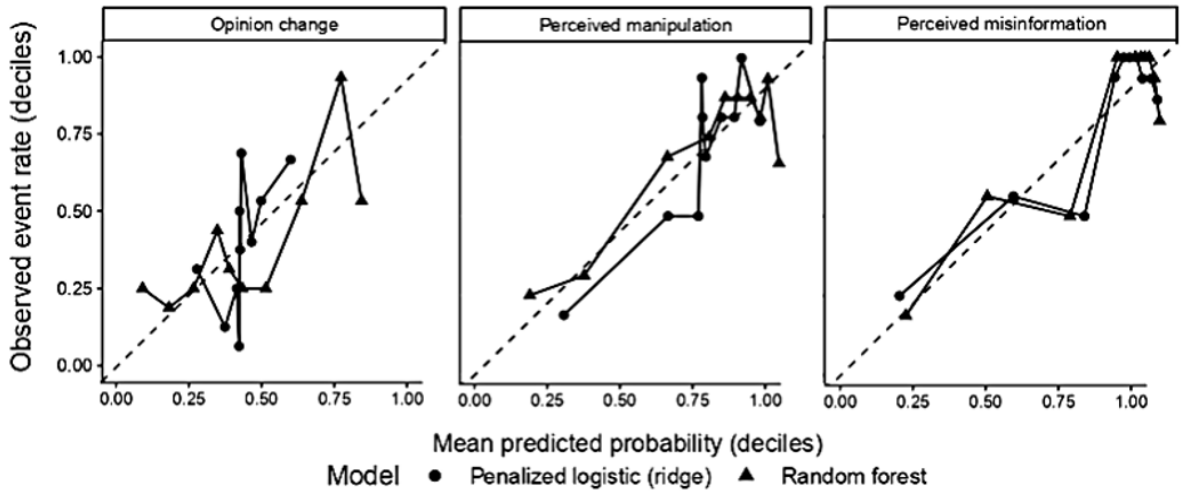

Figure 1. Decile-based calibration plots by outcome and model (pooled out-of-fold predictions) (Note. Points represent observed outcome rates within deciles of predicted probability; the 45° line indicates perfect calibration and curves are computed from pooled out-of-fold predictions to reflect out-of-sample probability accuracy) (created by the authors)

Missingness was handled via complete-case analysis implemented separately for each outcome while requiring complete data on the shared predictor set; consequently, analytic N varies slightly across outcomes (n ≈ 157-158). Derived outputs (cross-validated metrics, coefficient/importance summaries, and calibration/ROC artifacts) are available from the corresponding author upon reasonable request.

RESULTS

Analytic Sample and Outcome Prevalence

After complete-case filtering on each outcome and the shared predictor set, the analytic sample ranged from n = 157 to n = 158 across outcomes and models. Outcome prevalence indicated class imbalance, most notably for perceived misinformation (majority "yes"), whereas perceived manipulation and opinion change were less skewed. Because imbalance can inflate accuracy and complicate threshold-based interpretation, we report discrimination (AUC-ROC) alongside probabilistic accuracy (Brier score) and calibration, and we present threshold-dependent metrics computed under a training-only threshold rule (Youden's J) within each cross-validation fold. Overall predictive performance is summarized in **Table 1**, and visual diagnostics based on pooled out-of-fold predictions are shown in **Figure 1** (calibration) and **Figure 2** (ROC curves).

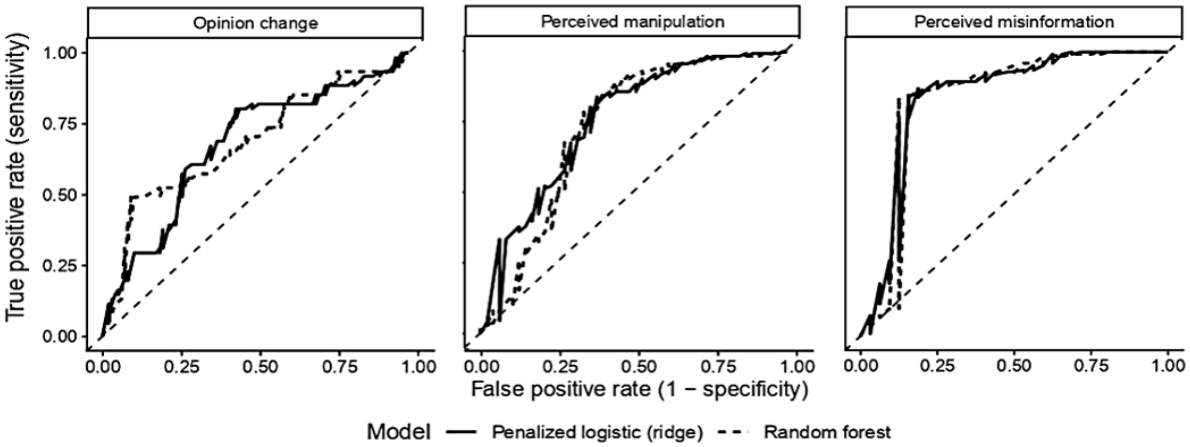

Figure 2. Out-of-fold ROC curves by outcome and model (Note. ROC curves are pooled from cross-validated out-of-fold predictions for ridge-penalized logistic regression and random forest; the diagonal indicates chance-level discrimination) (created by the authors)

Predictive Performance Differed by Outcome

Perceived misinformation. Predictive performance was highest for this outcome. The ridge-penalized logistic model achieved AUC-ROC = 0.860 ± 0.116 with Brier = 0.102 ± 0.033, while the random forest achieved AUC-ROC = 0.853 ± 0.110 with Brier = 0.099 ± 0.032. Threshold-dependent metrics indicated different sensitivity-specificity profiles across models, consistent with thresholds being learned within training folds and applied out-of-fold. Perceived manipulation. Performance was moderate. Ridge yielded AUC-ROC = 0.737 ± 0.068 and Brier = 0.173 ± 0.038, and random forest yielded AUC-ROC = 0.749 ± 0.090 and Brier = 0.167 ± 0.036. Balanced accuracy was broadly similar across model families, indicating comparable out-of-sample classification utility under the same evaluation protocol. Opinion change. This outcome was least predictable from the compact predictor set. Ridge achieved AUC-ROC = 0.738 ± 0.090 with Brier = 0.228 ± 0.008, while random forest achieved AUC-ROC = 0.696 ± 0.071 with Brier = 0.211 ± 0.023 (**Table 1**). Threshold-dependent performance was correspondingly lower, indicating limited separability given the available indicators.

Calibration and Discrimination

Calibration analyses based on pooled out-of-fold predictions indicated that both model families produced usable probability estimates for perceived misinformation, consistent with low Brier scores (~0.10). For perceived manipulation and especially opinion change, higher Brier values (~0.17 and ~0.21-0.23, respectively) aligned with weaker probabilistic accuracy. Decile-based calibration plots are reported in **Figure 1** and pooled out-of-fold ROC curves in **Figure 2**.

Ridge-Penalized Logistic Regression: Top Predictive Signals

To synthesize which indicators were consistently prioritized across model families, we visualize the strongest ridge signals (signed coefficients) alongside random-forest importance rankings, complementing the tabular summaries. For interpretability, we report odds ratios as OR = exp (β). These values are presented descriptively as predictive signals rather than inferential effect estimates. Perceived misinformation. The strongest ridge signals included multiple platforms selected (β ≈ -1.57; OR ≈ 0.208), confidence in meme messages (β ≈ -1.50; OR ≈ 0.224), and selecting Instagram among most-used platforms (β ≈ -1.07; OR ≈ 0.342). Positive-direction signals included emotion: laughter (β ≈ 0.93; OR ≈ 2.53) and selecting X among most-used platforms (β ≈ 0.68; OR ≈ 1.98). Additional signals included multiple emotions selected (β ≈ 0.35; OR ≈ 1.42) and emotion: annoyance (β ≈ 0.41; OR ≈ 1.50) (**Table 2**).

Perceived manipulation. The strongest signals included multiple platforms selected (β ≈ -0.75; OR ≈ 0.47), emotion: reflection (β ≈ -0.51; OR ≈ 0.60), and selecting Instagram (β ≈ -0.48; OR ≈ 0.62). A smaller positive-direction signal was observed for verification (β ≈ 0.22; OR ≈ 1.24). Several other predictors showed moderate magnitudes, consistent with lower overall separability than perceived misinformation. Opinion change. Coefficients were comparatively small in magnitude across predictors (e.g., multi-platform β ≈ -0.10; OR ≈

Table 2. Ridge-penalized logistic regression: top predictors ranked by absolute coefficient magnitude

Outcome	Predictor	Coefficient	OR	Direction
Perceived misinformation	Multiple platforms selected	-1.570	0.208	↓ Lower predicted odds
	Confidence in meme messages (0-2)	-1.500	0.224	
	Selected Instagram among most-used platforms	-1.070	0.342	
	Emotion: laughter	0.928	2.530	↑ Higher predicted odds
	Emotion: indifference	-0.887	0.412	↓ Lower predicted odds
	Selected TikTok among most-used platforms	-0.872	0.418	
	Selected Facebook among most-used platforms	-0.753	0.471	
	Selected X (formerly Twitter) among most-used platforms	0.684	1.980	↑ Higher predicted odds
	Emotion: annoyance	0.408	1.500	
	Multiple emotions selected	0.353	1.420	
Perceived manipulation	Multiple platforms selected	-0.749	0.473	↓ Lower predicted odds
	Emotion: reflection	-0.513	0.599	
	Selected Instagram among most-used platforms	-0.477	0.621	
	Multiple emotions selected	-0.346	0.707	
	Selected TikTok among most-used platforms	-0.340	0.712	
	Verifies meme information	0.218	1.240	↑ Higher predicted odds
	Selected Facebook among most-used platforms	-0.166	0.847	↓ Lower predicted odds
	Confidence in meme messages (0-2)	-0.161	0.852	
	Emotion: indifference	0.154	1.170	↑ Higher predicted odds
	Emotion: laughter	-0.141	0.868	↓ Lower predicted odds
Opinion change	Multiple platforms selected	-0.095	0.909	↓ Lower predicted odds
	Emotion: indifference	-0.094	0.910	
	Emotion: annoyance	-0.092	0.912	
	Selected X (formerly Twitter) among most-used platforms	-0.08	0.923	
	Selected Instagram among most-used platforms	-0.079	0.924	
	Multiple emotions selected	-0.070	0.933	
	Verifies meme information	0.068	1.070	↑ Higher predicted odds
	Emotion: laughter	0.063	1.070	
	Emotion: reflection	-0.061	0.940	↓ Lower predicted odds
	Selected TikTok among most-used platforms	-0.052	0.949	

Note. Coefficients (β) are shown with odds ratios (OR = exp(β)) and direction reflects the sign of β. Rankings summarize magnitude-based predictive signals within a penalized, cross-validated workflow; they should not be interpreted as inferential effect estimates, and confidence intervals are not reported. Ridge coefficients are shrinkage estimates and are reported as magnitude-ranked predictive signals rather than inferential odds ratios; we do not report confidence intervals because estimates arise from penalization and cross-validated evaluation. Coefficients/ORs are ranked by magnitude as predictive signals; they are not inferential effect estimates and do not include CIs due to shrinkage and cross-validated estimation.

0.91; verification β ≈ 0.07; OR ≈ 1.07), consistent with weaker discrimination and higher Brier values for this outcome. Caution (ridge ORs).

Because ridge estimates are penalized and evaluated under cross-validation, the ORs in **Table 2** should be interpreted as magnitude-based rankings of predictive contribution within the ridge model family, not as fully inferential odds ratios with confidence intervals, and not as causal effects. To characterize interpretable predictive signals, we summarize ridge coefficients as odds ratios (OR = exp (β)) and rank predictors by absolute coefficient magnitude (**Table 2**). Because coefficients are penalized and estimated within a cross-validated predictive workflow, these ORs are reported descriptively as magnitude-based predictive rankings, not inferential effect estimates; accordingly, we do not report confidence intervals.

Random Forest: Variable-Importance Rankings

Because random forest does not yield a single set of interpretable coefficients with directionality, we summarize impurity-based variable importance rankings to identify which predictors were most frequently prioritized by the model. Importance values are non-directional and are used for ranking predictive salience. For perceived misinformation, the highest-ranked predictors were multiple platforms selected and confidence, followed by platform-selection indicators (Instagram, Facebook, and TikTok) and multiple emotions selected. For perceived manipulation and opinion change, the model similarly emphasized breadth

Table 3. Selected top random forest variable-importance rankings (non-directional)

Outcome	Rank	Predictor	Importance
Perceived misinformation	1	Multiple platforms selected	9.091
	2	Confidence in meme messages (0-2)	7.781
	3	Selected Instagram among most-used platforms	2.887
	4	Selected Facebook among most-used platforms	2.584
	5	Selected TikTok among most-used platforms	2.411
	6	Multiple emotions selected	1.939
	7	Emotion: reflection	1.323
	8	Emotion: indifference	1.282
	9	Verifies meme information	0.793
	10	Emotion: annoyance	0.617
Perceived manipulation	1	Multiple platforms selected	6.010
	2	Emotion: reflection	2.500

Note. Selected top-ranked predictors are reported for descriptive illustration of relative predictive salience. Importance is impurity-based and does not indicate effect direction. The cross-outcome ranking structure is presented graphically in part B in **Figure 3**.

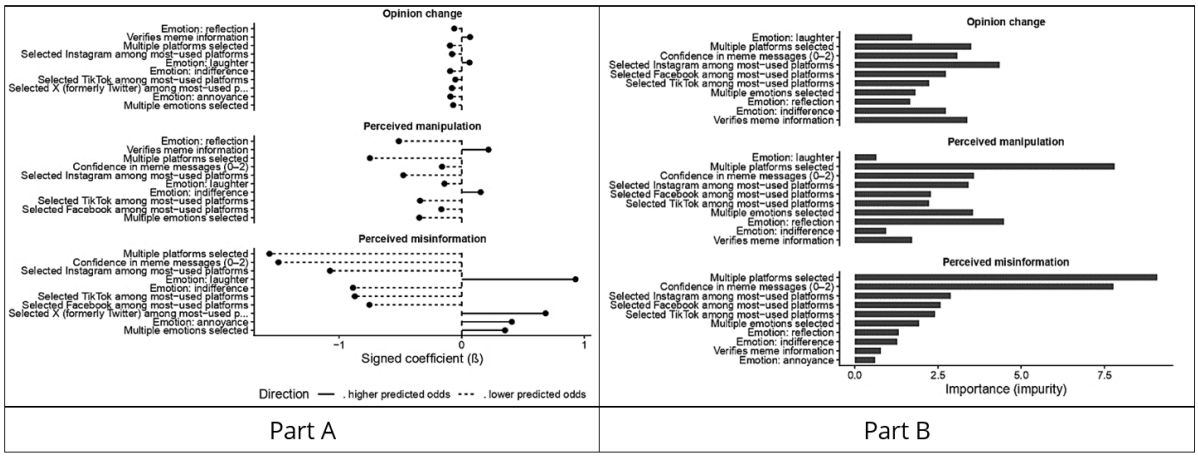

Figure 3. Cross-model synthesis of recurring predictors across outcomes (Note. Ridge: predictors ranked by absolute coefficient magnitude (|β|) (A) & Random forest: predictors ranked by variable importance (B); **Figure 3** highlights convergence in which features repeatedly emerge as salient across outcomes and model families; rankings are descriptive and do not imply causal effects) (created by the authors)

of platform selection and confidence-related indicators, while affective indicators contributed lower but non-trivial importance depending on the outcome. Because random forests do not yield directional coefficients, we report selected impurity-based variable-importance rankings in **Table 3**.

Summary of Patterns

Taken together, perceived misinformation was most predictable from the current indicator set, perceived manipulation was moderately predictable, and self-reported opinion change was least predictable. Across model families, multi-platform selection and confidence in meme messages repeatedly emerged as high-salience predictors (directional within ridge; rank-salient within random forest), while affective indicators contributed to outcome-specific ways. These findings describe out-of-sample predictive structure under a conservative evaluation protocol (training-only threshold selection, and reporting both discrimination and calibration), rather than causal effects of memes on attitudes. A cross-model synthesis of recurring predictors across outcomes is provided in **Figure 3**.

DISCUSSION

This study takes a deliberately conservative, predictive approach to three self-reported outcomes at the intersection of meme culture and informational integrity: perceived misinformation, perceived manipulation, and self-reported opinion change after seeing a meme. Across an interpretable ridge-penalized logistic

regression and a flexible random forest, predictive performance differed sharply by outcome. Perceived misinformation was most predictable (AUC-ROC ≈ 0.85-0.86; Brier ≈ 0.10), perceived manipulation showed moderate predictability (AUC-ROC ≈ 0.74-0.75; Brier ≈ 0.17), and opinion change was least separable with the available predictors (AUC-ROC ≈ 0.70-0.74; Brier ≈ 0.21-0.23). This hierarchy is consistent with the idea that integrity-related perceptions often operate as generalized evaluations of the information environment, while attitudinal change is more contingent on specific exposure episodes, attention, and contextual anchoring (Nyhan, 2020; Pennycook et al., 2020; Vosoughi et al., 2018).

One potentially unexpected pattern is that the self-reported verification indicator showed only weak predictive salience across outcomes. Several non-exclusive explanations are plausible. First, the measure is binary and self-reported, which may compress meaningful variation and be vulnerable to social desirability (respondents may report "verifying" even when verification is infrequent or superficial), thereby reducing discriminative value (Nyhan, 2020). Second, verification is likely episodic and content-dependent: students may verify only when a meme appears consequential, ambiguous, or personally relevant—conditions not captured by a compact, context-free item (Lazer et al., 2018; Wardle & Derakhshan, 2017). Third, verification could operate through indirect pathways (e.g., digital literacy, media trust, political interest, or epistemic vigilance) that were not measured in the current instrument; without these mediators, the direct predictive contribution of a single verification item is likely attenuated (Guess et al., 2020; Nyhan, 2020; Pennycook et al., 2020). Finally, under high base rates for perceived misinformation (where most respondents endorse that memes can misinform), verification may add limited separability beyond more proximal predictors such as confidence in meme messages and multi-platform repertoire breadth (Nyhan, 2020; Van Calster et al., 2019). These considerations suggest that future work should operationalize verification with greater granularity (frequency, triggers, and verification methods) and, where feasible, complement self-reports with behavioral indicators (Guess et al., 2020; Pennycook et al., 2020).

Substantively, the most recurrent signals across models were "breadth" indicators—especially multi-platform selection and confidence in meme messages—for perceived misinformation and manipulation. In ridge models, both appear among the largest-magnitude coefficients; in random forests, they rank among the highest-importance predictors (without directional interpretation). One plausible reading is that multi-platform repertoires correspond to more heterogeneous discursive styles, reputational cues, and cross-platform circulation contexts, while confidence reflects an evaluative stance toward meme content that maps directly onto integrity judgments (Guess et al., 2020; Nyhan, 2020). These patterns should be interpreted as predictive associations in this sample, not causal effects.

Affective indicators contributed to outcome-specific ways, suggesting that emotions function as interpretive cues shaping integrity judgments rather than as direct proxies of exposure (Milner, 2016; Papacharissi, 2015). In our models, laughter and annoyance carried more predictive signal for perceived misinformation, whereas reflection showed stronger salience for perceived manipulation. This pattern is consistent with the idea that emotional framing and vernacular humor help audiences classify meme content as playful satire, persuasive framing, or potentially misleading communication, depending on platform-native engagement logics and interpretive context (Highfield, 2016; Phillips & Milner, 2017; Shifman, 2014).

Methodologically, the paper responds to common reviewer expectations for predictive work in social science by going beyond "AUC-only" reporting. We report discrimination (AUC-ROC) but also probabilistic accuracy and calibration (Brier score; decile-based calibration), plus threshold-dependent metrics (balanced accuracy, F1, sensitivity, specificity) under a training-only threshold rule (Youden's J). This matters because discrimination can appear adequate even when predicted probabilities are miscalibrated, and because threshold choices can materially alter conclusions if selected post hoc on test data (Steyerberg, 2019; Van Calster et al., 2019). For outcomes with class imbalance (especially perceived misinformation), calibration and threshold-aware metrics help prevent overclaiming model utility under skewed base rates (Davis & Goadrich, 2006; Saito & Rehmsmeier, 2015).

Implications

These findings have three implications. Theoretically, they support treating political memes as participatory political talk that travels through affective cues and interpretive norms, where integrity-related perceptions may crystallize more readily than self-reported attitude change (Papacharissi, 2015; Shifman,

2014). Methodologically, they show why communication research that uses prediction should report calibration (Brier/deciles) alongside discrimination and should select classification thresholds on training data only—especially when class imbalance makes accuracy easy to inflate (Steyerberg, 2019; Van Calster et al., 2019). Practically, lightweight survey items can help flag profiles more likely to perceive misinformation or manipulation, informing media literacy and verification-oriented interventions for youth in Global South contexts.

Limitations and Future Research

- Small sample and estimate stability. The modest analytic n limits the stability of cross-validated estimates and increases uncertainty around AUC and calibration; performance may vary materially with different samples or political moments. Replication with larger, non-student Global South samples is needed before making transportability or applied-use claims.
- Measurement and construct coverage. Predictors capture self-reported "most-used" platforms and emotions, not behavioral exposure intensity, meme content features, or algorithmic reach. Adding time-on-platform, frequency of meme encounters, topic/source cues, ideology, media trust, and digital literacy should improve prediction—especially for opinion change (Guess et al., 2020; Nyhan, 2020).
- Design and inference. The cross-sectional, self-report design precludes causal claims. Longitudinal panels, experiments, or mixed designs can test temporal ordering between affect, perceived integrity, and attitudinal outcomes (Gelman et al., 2020; Pennycook et al., 2020).
- External validity and transportability. The Salvadoran university sample advances Global South evidence but may not generalize across age groups, platform ecologies, or political moments. Replications should assess model transportability across contexts and platforms (Persily, 2017; Tucker et al., 2018).
- Context sensitivity beyond student samples. Recent Latin American evidence suggests that interaction with political memes can relate differently to ideological versus affective polarization across platforms, underscoring the need to test transportability in non-student and broader Global South samples (Zumárraga-Espinosa et al., 2025).

CONCLUSION

This article provides a predictive assessment of perceived misinformation, perceived manipulation, and self-reported opinion change in relation to political meme engagement using compact indicators in a Global South university sample. Across ridge regression and random forests, perceived misinformation was most predictable, perceived manipulation showed moderate predictability, and opinion change was least predictable with the available predictors. Multi-platform selection and confidence in meme messages emerged as recurrent signals across model families, while affective reactions contributed additional predictive information consistent with theories of affective publics and participatory political talk (Papacharissi, 2015; Shifman, 2014).

Future research should directly address the limitations revealed by this design. First, improving prediction of self-reported opinion change—the least separable outcome in our models—will likely require behavioral measures of platform use (time-on-platform, frequency of meme encounters, and cross-platform sharing) and content-level features (topic, source cues, and humor style) that capture exposure episodes and contextual anchoring more directly (Nyhan, 2020; Prior, 2007). Second, verification should be measured with greater granularity (how often, under what triggers, and which methods are used) rather than a single binary self-report item, ideally complemented by behavioral or task-based indicators of discernment (Guess et al., 2020; Pennycook et al., 2020). Third, replication beyond student samples—across age groups and diverse Global South settings—will be necessary to evaluate transportability and context sensitivity of calibrated predictive models, particularly given the known dependence of information effects on platform ecologies and political moments (Tucker et al., 2018; Van Calster et al., 2019).

Author contributions: CHHM: conceptualization, formal analysis, investigation, methodology, supervision, writing – original draft writing – review & editing; **CAEM:** formal analysis, investigation, methodology, supervision, writing – original draft writing – review & editing. Both authors approved the final version of the article.

Funding: This research was funded by Universidad Salvadoreña Alberto Masferrer.

Ethics declaration: This study was approved by the Universidad Salvadoreña Alberto Masferrer Ethical Committee on 16 February 2025 with approval number CEI-USAM-2025-016. This study compiled with internationally recognized ethical principles for research involving human participants. Participation was voluntary and preceded by informed consent. To protect privacy, the survey did not collect personally identifiable information (e.g., names, phone numbers, email addresses, or precise place of residence). Data were analyzed in aggregate and stored securely with access restricted to the research team.

AI statement: During the preparation and revision of this work, the authors used ChatGPT (OpenAI) for editorial assistance, including language refinement, clarity, readability, and organization of the manuscript. Following the use of this tool, the authors carefully reviewed and edited the content and took full responsibility for the final version of the article.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Ahmed, S., & Masood, M. (2024). Breaking barriers with memes: How memes bridge political cynicism to online political participation. *Social Media + Society, 10*(2). https://doi.org/10.1177/20563051241261277

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *PNAS, 115*(37), 9216-9221. https://doi.org/10.1073/pnas.1804840115

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*, 1-3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Chadwick, A. (2017). *The hybrid media system: Politics and power* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190696726.001.0001

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233-240). ACM. https://doi.org/10.1145/1143844.1143874

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1-22. https://doi.org/10.18637/jss.v033.i01

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press. https://doi.org/10.1017/9781139161879

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *PNAS, 117*(27), 15536-15545. https://doi.org/10.1073/pnas.1920498117

Halversen, A., & Weeks, B. E. (2023). Memeing politics: Understanding political meme creators, audiences, and consequences on social media. *Social Media + Society, 9*(4). https://doi.org/10.1177/20563051231205588

Highfield, T. (2016). *Social media and everyday politics*. Polity Press.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094-1096. https://doi.org/10.1126/science.aao2998

Masood, M. (2024). Political meme use can lead to political intolerance: Evidence from a 2-wave panel survey. *International Journal of Public Opinion Research, 36*(4), Article edae052. https://doi.org/10.1093/ijpor/edae052

Mihăilescu, M.-G. (2024). Never mess with the "memers": How meme creators are redefining contemporary politics. *Social Media + Society, 10*(4). https://doi.org/10.1177/20563051241296256

Milner, R. M. (2016). *The world made meme: Public conversations and participatory media*. MIT Press. https://doi.org/10.7551/mitpress/9780262034999.001.0001

Mina, A. X. (2019). *Memes to movements: How the world's most viral media is changing social protest and power*. Beacon Press.

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625-632). ACM. https://doi.org/10.1145/1102351.1102430

Nissenbaum, A., & Shifman, L. (2017). Internet memes as contested cultural capital: The case of 4chan's /b/ board. *New Media & Society, 19*(4), 483-501. https://doi.org/10.1177/1461444815609313

Nyhan, B. (2020). Facts and myths about misperceptions. *Annual Review of Political Science, 23*, 303-321.

Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199999736.001.0001

Paulenová, M., & Kluknavská, A. (2025). Scroll, laugh, learn: Political memes as catalysts for active information-seeking. *Popular Communication*. https://doi.org/10.1080/15405702.2025.2571785

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31*(7), 770-780. https://doi.org/10.1177/0956797620939054

Persily, N. (2017). The 2016 U.S. election: Can democracy survive the Internet? *Journal of Democracy, 28*(2), 63-76. https://doi.org/10.1353/jod.2017.0025

Phillips, W., & Milner, R. M. (2017). *The ambivalent Internet: Mischief, oddity, and antagonism online*. Polity Press.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press. https://doi.org/10.1017/CBO9781139878425

Ross, A. S., & Rivers, D. J. (2017). Digital cultures of political participation: Internet memes and the discursive delegitimization of the 2016 U.S. presidential candidates. *Discourse, Context & Media, 16*, 1-11. https://doi.org/10.1016/j.dcm.2017.01.001

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE, 10*(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

Shifman, L. (2014). *Memes in digital culture*. MIT Press. https://doi.org/10.7551/mitpress/9429.001.0001

Spiegelhalter, D. J. (1986). Probabilistic prediction in patient management and clinical trials. *Journal of Clinical Epidemiology, 39*(5), 421-433. https://doi.org/10.1002/sim.4780050506

Steyerberg, E. W. (2019). *Clinical prediction models: A practical approach to development, validation, and updating* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-16399-0

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology, 21*(1), 128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. Hewlett Foundation. https://doi.org/10.2139/ssrn.3144139

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*, Article 230. https://doi.org/10.1186/s12916-019-1466-7

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making, 26*(6), 565-574. https://doi.org/10.1177/0272989X06295361

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe*. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Wiggins, B. E., & Bowers, G. B. (2015). Memes as genre: A structurational analysis of the memescape. *New Media & Society, 17*(11), 1886-1906. https://doi.org/10.1177/1461444814535194

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35. https://doi.org/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3

Zhang, J. (2025). When memes govern: A study of TikTok video memes in the 2024 US presidential election. Party Politics. https://doi.org/10.1177/13540688251380615

Zumárraga-Espinosa, M., AbuJoudeh-Rosero, R., & Carofilis-Cedeño, C. (2025). El doble rostro de los memes: la relación dual entre memes políticos y tipos de polarización en el contexto político de Ecuador [The double face of memes: the dual relationship between political memes and types of polarization in the political context of Ecuador]. *Castalia-Revista de Psicología de la Academia*, (45), 63-85. https://doi.org/10.25074/07198051.45.3022