JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Social Media and Contemporary Indian Democracy: Political Mobilisation, Misinformation, and Electoral Trust

[1]Dr. Sandeep Kumar Singh
[1]Department of Law
Sidhi, India

Abstract: Social media connects political organisation, public debate, and interpersonal communication in India, but its democratic consequences cannot be inferred from popularity or isolated instances of deception. This critical narrative review examines how political mobilisation and misinformation intersect with electoral trust. It synthesises Indian empirical studies, comparative intervention research, political communication theory, and selected Election Commission documents available by 5 November 2026. Sources were purposively selected for relevance, identifiable methods, and applicability; the review does not claim systematic completeness or estimate a pooled causal effect. The synthesis identifies three connected problems: mobilisation occurs within hybrid communication networks; the effects of misinformation depend on interpretation, identity, and social context; and electoral confidence concerns distinct procedures and institutions. Evidence on media literacy and correction varies across samples, interventions, and follow-up periods, limiting universal policy prescriptions. The proposed framework separates exposure, belief, sharing, participation, and institutional judgement, and treats language, prior trust, and institutional responsiveness as conditioning factors. It supports multilingual verification, transparent campaign provenance, proportionate responses, and inspectable electoral communication, while preserving legitimate scrutiny and privacy. The article contributes a conditional explanation and a testable research agenda rather than a national prevalence estimate or an assessment of party-specific responsibility. Democratic resilience depends on opportunities to participate alongside credible evidence, accountable communication, and meaningful avenues for remedy.

IndexTerms **- Indian democracy; social media; political mobilisation; misinformation; electoral trust; political communication.**

1 Introduction

Social media matters to Indian democracy because political information moves through relationships as well as institutions. A campaign speech can become a short video, enter a messaging group, receive a personal endorsement, and reappear in a television discussion. Each transition changes the audience, contextual cues, and perceived credibility of the message. Research on India's hybrid campaigning environment makes it inappropriate to describe digital platforms as a separate sphere replacing established political organization [1, 2]. The relevant question is how these connected channels distribute opportunities to participate, influence the credibility of competing claims, and shape confidence in electoral procedures.

The subject also demands separation between democratic participation and democratic quality. More opportunities to express political preferences may coexist with intimidation, misleading information, and unequal visibility. An increase in campaign sharing does not establish improvement in deliberation, just as the presence of false posts does not establish that an election outcome was determined by misinformation. Comparative scholarship finds varied relationships between digital media and democratic outcomes rather than one universal effect [3, 4]. A useful analysis therefore specifies the outcome under discussion, the population observed, and the mechanism connecting communication to that outcome before judging whether the consequences are beneficial or harmful.

Electoral trust provides a particularly important but frequently underspecified outcome. Citizens may trust the secrecy of their own ballot while doubting campaign fairness, or accept the count while criticizing political advertising. Support for a winning party can also colour assessments of institutional performance. Building on electoral integrity scholarship, this paper treats confidence as a set of judgements about identifiable stages and institutions rather than a single expression of satisfaction [5]. Such a definition protects legitimate scrutiny: confidence supported by evidence is democratically preferable to unconditional reassurance, and criticism supported by documentation must be distinguished from unsupported allegations.

The paper asks three questions. How do social media infrastructures support political mobilisation within India's hybrid communication system? Through which mechanisms might misleading political information affect electoral trust? Which responses are justified by available evidence while preserving expression, privacy, and political contestation? Its contribution is an integrative framework that links these questions without assuming a direct progression from exposure to electoral change. It combines India-specific findings with carefully bounded comparative evidence and distinguishes established observations from

proposed relationships. The framework contributes to political communication scholarship by connecting mobilisation and information disorder to distinct objects of electoral trust, and by specifying the evidence needed to test these connections.

2 Review approach and evidentiary boundaries

This is a critical narrative review with a focused documentary update, not a systematic review or meta-analysis. Searches were conducted on 5 November 2026 through web-based scholarly discovery and targeted searches of publishers, university repositories, and official electoral sources. Search combinations included India, social media, WhatsApp, political mobilisation, misinformation, electoral trust, media literacy, fact-checking, and AI-generated campaigning. Foundational theoretical works were retained where necessary to explain mechanisms. The resulting bibliography contains 28 purposively selected references published between 2006 and 2026. This number describes the final reference list; it is not a database screening total or a measure of evidence completeness.

Selection prioritized direct relevance to India, identifiable research designs, traceable publication details, and a clear role in answering the research questions. Comparative studies were included to explain mechanisms or assess interventions when India-specific evidence was limited. Primary journal and publisher records were used to check bibliographic information; accessible abstracts, publisher summaries, author repositories, and official documents supported the bounded synthesis. Access was not uniform, so the review does not imply that every supplementary appendix or full text was audited. It excludes unsupported social media anecdotes and does not estimate national misinformation prevalence, party-specific responsibility, or the electoral impact of particular platforms.

Evidence is differentiated by function. Conceptual scholarship supplies interpretive frameworks; observational studies describe distributions and associations; experiments assess specific interventions under defined conditions; official communications document institutional positions and announced responses. These categories are not interchangeable. An administrative account of action against content does not independently measure the accuracy of every intervention, while an experiment improving headline discernment does not demonstrate increased trust in elections. Tables 1 and 2 specify constructs and selected Indian evidence. Figure 4 presents the permissible inferential steps. No pooled effect size, original participant sample, reliability coefficient, or statistical significance test is reported because no new dataset was collected.

The analysis proceeded by organizing source-supported observations around mobilisation, information evaluation, trust, and governance, then identifying connections and unresolved assumptions. This interpretive organization was performed for the present review; it was not independent duplicate coding under a preregistered protocol. Contradictory or limited findings were retained rather than treated as exceptions to a predetermined argument. Publication bias, English-language discovery, uneven access, and rapidly changing platform features remain material limitations. The article consequently advances a conditional account and a research agenda, not a definitive assessment of the overall effect of social media on Indian democracy.

Table 1. Concepts and proposed measurement distinctions

Construct	Working distinction	Possible indicator
Mobilisation	Coordination or activation of participation	Event attendance; volunteering; baseline-adjusted participation
Exposure	Opportunity to encounter a claim	Consent-observed receipt; recalled exposure with uncertainty
Belief	Acceptance of a factual proposition	Accuracy judgement and confidence
Sharing	Distribution of material to others	Observed forwarding or clearly labelled intention
Misinformation	False claim; deceptive intent not established	Claim-level verification against traceable evidence
Disinformation	Deliberately deceptive communication	Evidence of intent or coordinated deceptive production
Electoral trust	Confidence in specified procedures or institutions	Separate campaign, administration, count, and remedy items

Definitions informed by [5, 6]; indicators are author-proposed, not validated measures.

3 Conceptual framework

The hybrid media system explains why digital communication cannot be understood solely through platform features. Political influence depends on interactions among journalism, television, party organization, public events, and online circulation [1]. In the Indian context, this directs attention to the brokers who translate a national message into a local language, connect an online audience to a meeting, or give a circulating clip the authority of a familiar organization. Figure 1 maps these reciprocal relationships. Its arrows represent possible communication routes, not measured traffic volumes. The analytical unit is the relationship between institutions and networks, rather than the isolated post.

Connective action adds an account of personalized participation. Individuals can associate a public cause with their own experience and circulate a customised message without entering a formal organisation [7]. That possibility does not mean organised actors disappear. For this review, personal expression and party coordination are treated as potentially complementary: citizens may adapt prepared content, endorse a campaign independently, or reject centrally issued messages. Empirical research must establish the degree of coordination rather than infer spontaneity from an ordinary-looking account. The same caution applies to apparently uniform messaging, which can reflect shared preferences without necessarily proving a centrally managed campaign.

A deliberative perspective evaluates whether communication helps citizens exchange reasons, challenge powerful actors, and revise judgements. Habermas identifies the importance of media independence and connections between public communication and responsive civil society [8]. Applied here, this is a normative standard rather than an empirical claim that online discussion ordinarily meets those conditions. Participation matters partly because it can make experiences and grievances visible; however, intimidation and exclusion can prevent formally accessible forums from functioning as inclusive arenas. The framework therefore asks who can speak safely, whose claims receive attention, and whether a meaningful response is possible.

Information disorder requires equally careful definition. Following Wardle and Derakhshan, misinformation concerns false information without necessarily establishing an intention to deceive, whereas disinformation involves deliberate deception; harmful deployment of genuine information raises a different category of concern [6]. Intent is often difficult to determine from a message alone. This review consequently uses misinformation when deception cannot be established, and reserves stronger descriptions for evidence about production or coordination. Opinion, satire, criticism, and factual error should not be merged into one prohibited category. Table 1 sets out these distinctions and proposes observable indicators for future research.

Figure 2 brings these perspectives together through two conditional pathways. Communication can reduce coordination costs, encourage efficacy, and support participation. It can also circulate decontextualised claims, activate identity-based responses, and

generate uncertainty about institutions. The proposed relationships are conditioned by language, prior beliefs, social relationships, platform design, and the availability of credible correction; moderation has not been statistically tested in this review. Neither is inevitable, and they may operate simultaneously for the same person. Electoral trust is positioned downstream of interpretation and institutional experience, not immediately downstream of exposure. This distinction is central to the article's argument and prevents communication metrics from being mistaken for democratic outcomes.

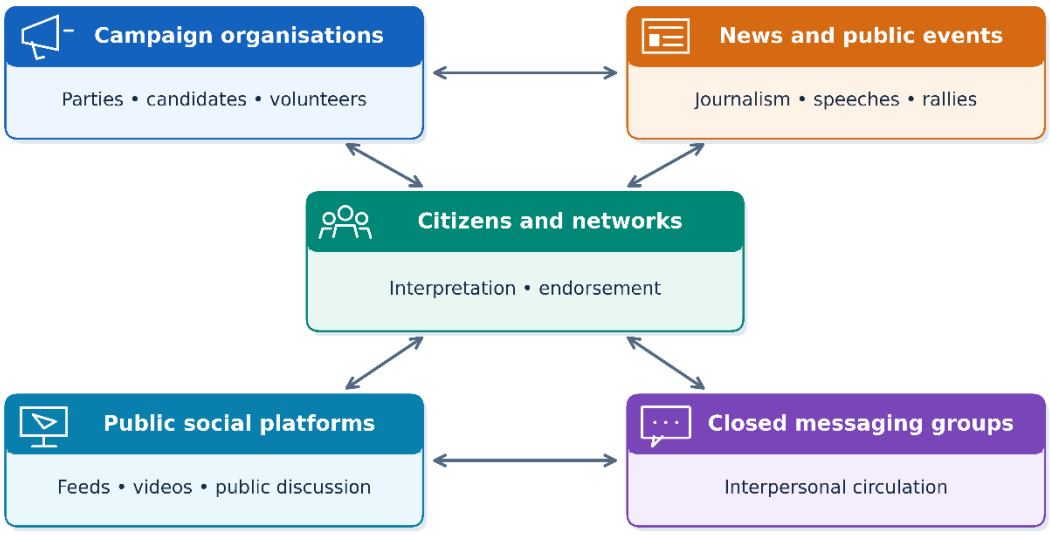

Figure 1. Hybrid political communication in India.
Original conceptual map informed by [1, 2, 7, 9]. Bidirectional arrows denote possible exchanges, not measured traffic. Colours distinguish actors and communication settings.

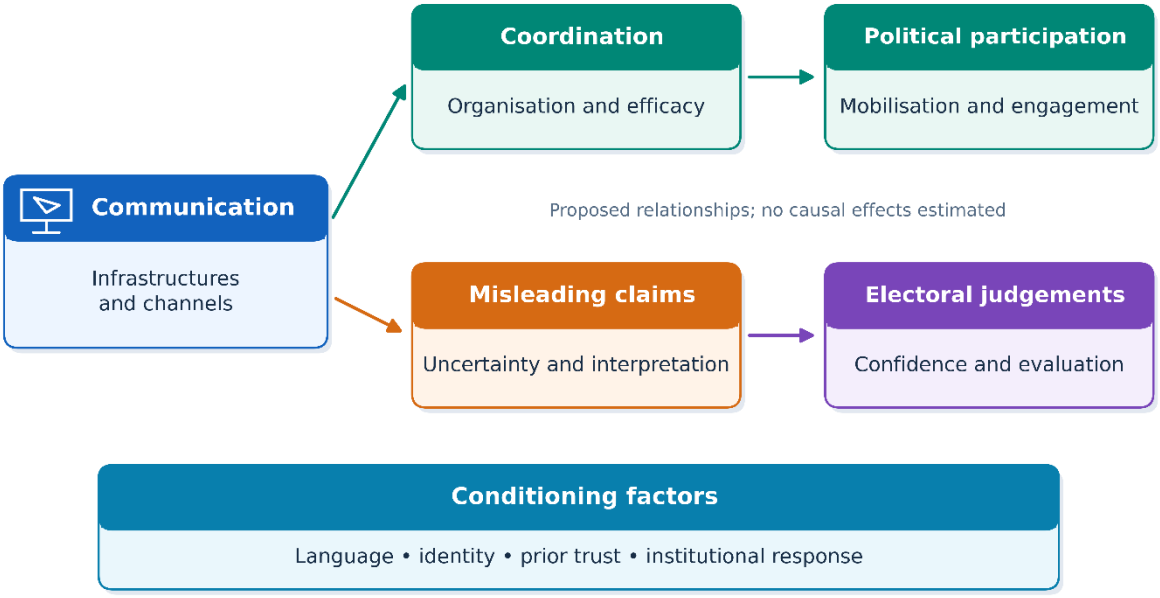

Figure 2. Conditional pathways connecting communication infrastructures with participation and electoral judgements.
Original synthesis informed by [3, 4, 5, 6]. The lower panel identifies conditioning factors; arrows are proposed relationships, not estimated effects.

4 Political mobilisation and unequal participation

Social media can support mobilisation by making events, policy demands, and campaign materials easier to circulate. It can help organisers identify sympathetic audiences and allow citizens to connect distant issues with local concerns. These are plausible coordination functions rather than proof of additional turnout. Boulianne's meta-analysis establishes a broad research basis for examining relationships between social media use and participation, while also illustrating the limits of treating observational associations as causal effects [3]. India's campaign scholarship similarly places information sharing alongside party contact and established media exposure [2]. The implication is that digital mobilisation should be studied as part of a larger organisational process.

Political communication also works through familiarity. Pal's analysis of Narendra Modi's political tweeting examines how apparently ordinary messages and selective interactions contributed to public presentation [10]. The wider analytical lesson is that political relevance is not confined to explicit requests for votes. Greetings, cultural references, and everyday commentary can help sustain audience attention between major campaign moments. This review does not generalise that particular case to every leader or infer electoral persuasion from follower growth. Instead, it identifies personalisation as a mechanism that future comparative studies could examine across parties, regions, and stages of the electoral cycle.

Mobilisation can also be embedded in cultural practice and ideological identification. Udupa's research on enterprise Hindutva examines the relationship between social media affordances, nationalism, and everyday participation in urban India [11]. Its significance for this framework is the need to examine the labour, aspirations, and social meanings behind online activity. A volunteer may experience content production as recognition, belonging, or public service rather than merely campaign work. Understanding these motivations requires contextual methods. It would be misleading to assume that all participants share the same organisational role, or to treat findings concerning one political formation as a representative portrait of every Indian digital public.

The distribution of benefits remains an empirical question. Producing persuasive videos, maintaining language-specific pages, purchasing advertising, and answering criticism require resources. A low cost of opening an account does not establish equality in the capacity to be heard. Similarly, nominal access can conceal constraints associated with literacy, shared devices, time, disability, or exposure to harassment. These are proposed dimensions of analysis, not population estimates supplied by this review. A defensible evaluation would compare opportunities to enter debate with the ability to sustain participation and obtain responses, while identifying which groups are absent from the observable data.

Private and public channels also create different participation environments. Public posts are visible to journalists and opponents, whereas communication in closed groups can depend more strongly on interpersonal relationships. Chauchard and Garimella's study of partisan WhatsApp groups cautions against imagining such spaces as exclusively political: much observed material was social, religious, or entertaining [9]. Political messages may therefore enter an ongoing relationship rather than a dedicated debate. This does not establish that social content causes persuasion. It suggests that research should examine how ordinary interaction supplies context for political communication, and why counting only explicitly partisan posts may miss important features of group life.

Table 2. Selected Indian evidence and limits of inference

Source	Design or evidence	Contribution and boundary
Neyazi et al. [2]	Study of 2014 election campaigning	Connects digital and established campaigning; does not establish a general platform effect.
Pal [10]	Case analysis of political tweeting	Explains public presentation; cannot infer votes from follower growth.
Udupa [11]	Contextual study of urban digital nationalism	Examines participation as social practice; context-specific.
Chauchard and Garimella [9]	533 closed WhatsApp threads in Uttar Pradesh; approximately 36,000 images coded over nine months	Finds varied everyday content; sampled groups do not represent all Indian messaging.
Guess et al. [12]	Media literacy experiments including India	Effects differ by sample and persistence; discernment is not electoral trust.
Badrinathan [13]	Indian field experiment; N = 1,224	No significant average improvement; intervention and population specific.
Neyazi et al. [14]	Cross-sectional survey of English-language internet users; n = 1,013	Associations in a bounded sample; cross-sectional causality unresolved.
Shahi and Hale [15]	580 unique claims from 451 tipline users; English, Hindi, and Telugu	Documents submitted claims and responses; self-selected verification demand.

Qualitative synthesis of selected sources, not a complete evidence inventory or pooled comparison.

5 Misinformation and contemporary political communication

The contemporary challenge includes familiar forms of distortion alongside synthetic media. An old photograph presented as a current event, a genuine speech edited to remove context, and a fabricated voice recording involve different production processes and verification requirements. Their political implications depend on timing, audience, and the claim being made. Research on the spread of true and false news demonstrates why diffusion deserves attention, but findings from historical Twitter data do not establish the rate or consequences of misinformation in India [16]. This review therefore treats circulation speed as a potential risk mechanism requiring local measurement rather than a universal property of every platform or claim.

Interpersonal endorsement may complicate evaluation because users encounter a message through a person they know rather than through its original publisher. The receiver may judge the sender's sincerity while lacking information about the underlying evidence. Yet familiarity should not automatically be described as gullibility: trusted relationships can also carry corrections and warnings. Arun's discussion of WhatsApp, rumours, and lynchings directs attention to the wider political and institutional setting of harmful communication [17]. The appropriate analytical question is how a communication channel interacts with existing conditions. A platform-only explanation risks overlooking responsibility, local conflict, and failures to prevent or address violence.

Qualitative evidence also challenges the assumption that misinformation is simply a shortage of technical skill. The LSE WhatsApp vigilantes report links the reception and circulation of harmful rumours to social prejudice and group relations [18]. Its contribution is contextual understanding rather than an estimate of how frequently all Indian users believe false claims. The distinction matters for intervention design. A user who lacks source-checking skills may need accessible instruction; a user who circulates material to express hostility presents a different problem. The proposed framework therefore separates inability to verify, failure to attend to accuracy, identity protection, and deliberate harmful distribution.

Political psychology offers reasons to distinguish belief from sharing. Pennycook and Rand review research on analytical thinking, attention, and other influences on misinformation judgement [19]. People may share an item to signal affiliation or provoke discussion without expressing complete confidence in its truth. Conversely, a person may privately accept a claim without forwarding it. Content diffusion, expressed belief, and behavioural response consequently require different measures. International work by Allcott and Gentzkow also illustrates the analytical distance between encountering false news and estimating its electoral consequences [20]. Neither engagement counts nor retrospective recollections alone resolve that distance.

Generative AI introduces an additional authentication problem but should not displace attention from ordinary misleading content. Garimella and Chauchard's account of research in India highlights both the presence of AI-generated political material and difficulties in assessing it reliably at scale [21]. Highly visible examples cannot establish national prevalence, average persuasiveness, or a decisive electoral effect. The framework distinguishes the production of synthetic material from its deceptive use: an openly labelled illustration is analytically different from an impersonation presented as an authentic statement. Figure 3

accordingly distinguishes procedural confidence and opportunities for remedy as separate trust objects, each requiring evidence suited to the particular claim.

The documentary update shows institutional attention to these issues. The Election Commission's May 2024 communication instructed parties on responsible social media use [22]. Its January 2025 advisory addressed prominent identification of AI-generated campaign material [23], and an April 2026 communication described action concerning unlawful or misleading online content during assembly elections in Assam, Kerala, Tamil Nadu, Puducherry, and West Bengal [24]. These sources establish announced expectations and administrative responses; they do not independently demonstrate consistent enforcement or increased voter trust. The contemporary significance is therefore a need for evaluation: researchers should examine which actions occurred, how reasons were communicated, and whether people could contest mistakes without assuming that official intervention automatically achieved its intended democratic effects.

6 Electoral trust and the interpretation of evidence

Electoral trust should be disaggregated into at least four objects: campaign fairness, election administration, counting and results, and opportunities for remedy. Figure 3 presents this proposed measurement framework. A citizen might believe polling officials behaved competently while considering campaign finance or political advertising unfair. Another might regard a particular result as credible but doubt the accessibility of complaint procedures. A single question asking whether elections are trustworthy would obscure these differences. Norris's electoral integrity framework supports attention to the full electoral cycle [5]; the four-part structure used here is an analytical adaptation for the social media problem rather than a validated national scale.

The most direct informational risk concerns false procedural instructions. A misleading claim about where, when, or how to vote could affect participation without changing party preference. A different risk concerns allegations that undermine confidence in administration or counting. Such allegations cannot be classified solely by whether they criticise an institution: their evidentiary basis must be assessed. Democratic scrutiny may expose real problems, and trustworthy institutions should be able to answer documented questions. The relevant outcome is therefore calibrated confidence, meaning judgements responsive to verifiable performance and credible evidence, rather than maximum confidence regardless of the circumstances.

Existing survey evidence underscores the complexity of these relationships. Neyazi, Kalogeropoulos, and Nielsen examine misinformation concerns and news participation among English-language internet users in India [14]. Their cross-sectional design identifies associations within that sample, not the direction of causation or national prevalence. Politically attentive people may encounter more contested claims and report greater concern, while concern itself may encourage further information seeking. Such reciprocal possibilities make it inappropriate to interpret anxiety about misinformation as proof of actual exposure or as a direct measure of electoral distrust. Table 2 keeps sample boundaries visible alongside each study's substantive contribution.

Partisan interpretation creates another identification challenge. If supporters of losing parties report lower trust after an election, that pattern may reflect reactions to the outcome, exposure to allegations, prior institutional assessments, or combinations of these processes. Measuring platform use only after the result cannot separate these explanations. Repeated observations before campaigning, during voting, and after counting would provide a stronger basis for analysis. Researchers should also distinguish criticism of a government from confidence in election administrators. Combining these objects into a single political trust score could produce an apparent platform effect that actually reflects changes in approval of incumbent leaders.

The framework therefore proposes three expectations for future testing. First, procedural misinformation should have its clearest effects on procedural knowledge rather than automatically on party preference. Second, a correction should be more likely to support confidence when recipients can inspect its supporting evidence and understand the procedure involved. Third, institutional confidence should vary with prior trust and perceived fairness of the response. These are hypotheses generated by the synthesis, not findings established by it. Figure 4 identifies the sequence of evidence required to move from observing a message to making a defensible claim about an institutional or electoral outcome.

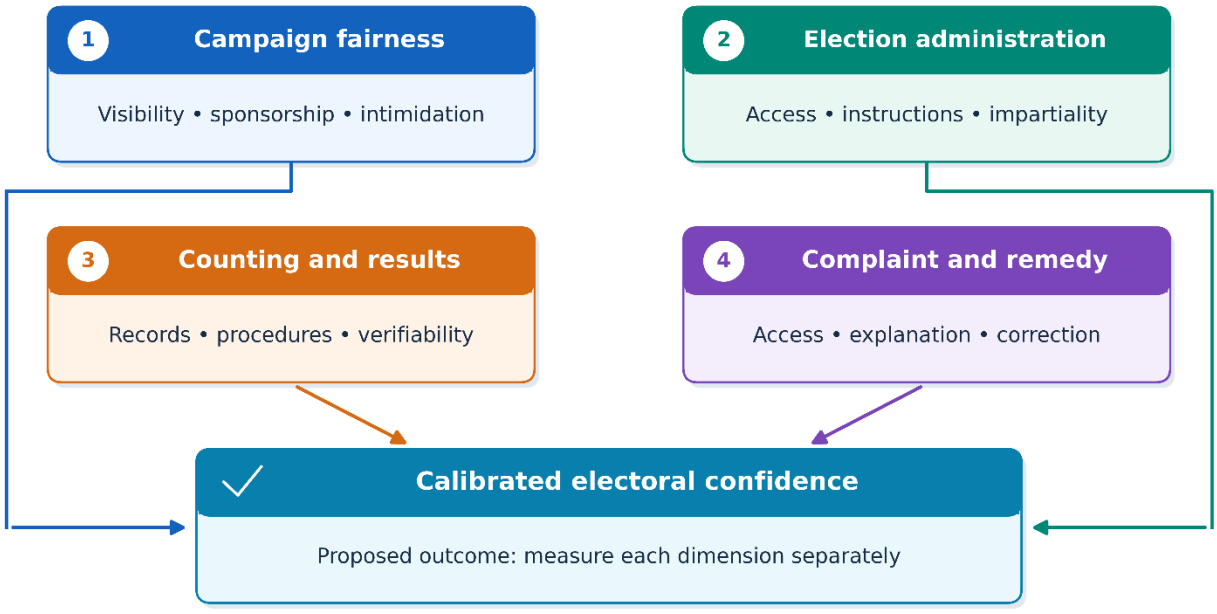

Figure 3. Four proposed dimensions of electoral trust.

Original adaptation informed by [5]. The dimensions contribute separately to the proposed outcome; colour does not indicate measured importance. The framework requires empirical validation.

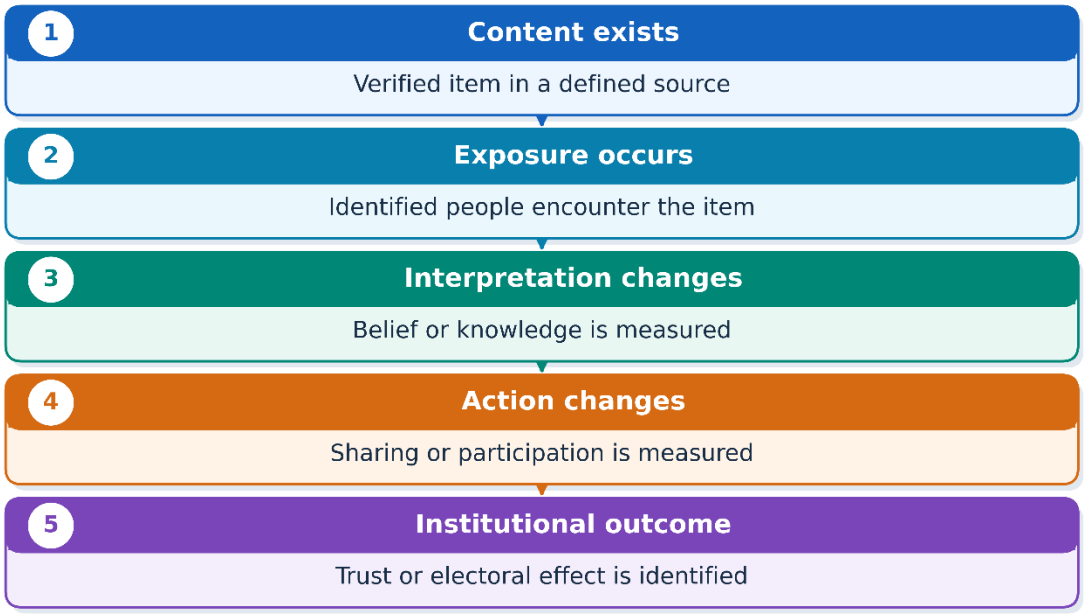

Figure 4. Evidence requirements for claims about social media effects.
Original inference framework informed by [4, 12, 13, 16, 19, 20, 25, 26, 27, 28]. Each stage requires additional evidence; a later outcome cannot be inferred from an earlier observation alone.

7 What intervention evidence does and does not show

Media literacy is a plausible response but not a uniform treatment. Guess and colleagues tested a short intervention in the United States and India and reported a 17.5% improvement in discernment in the highly educated Indian online sample, with no measurable effect in the rural sample and no persistence of the Indian online effect at follow-up [12]. The percentage is a source-reported improvement, not a percentage-point change or a new estimate. The finding supports attention to audience and implementation rather than a blanket conclusion that instruction solves misinformation. Badrinathan's Indian field experiment found no significant average improvement in misinformation identification from its intervention and identified partisan heterogeneity [13]. Taken together, these studies make programme design, population, and follow-up central to evaluation rather than peripheral qualifications.

Fact-checking offers a different mechanism: it addresses the accuracy of identifiable claims. Porter and Wood's experiments provide evidence that corrections can improve factual accuracy across several national settings [25]. Those settings did not include India, so the study cannot establish an Indian electoral effect. Nevertheless, it supplies a reason to evaluate corrections rather than dismiss them as inherently ineffective. Nyhan's review likewise cautions against simplistic claims about misperceptions and correction [28]. For the present framework, the relevant questions are whether a correction reaches the exposed audience, changes the targeted belief, persists, and affects a meaningful behaviour or institutional judgement.

Accuracy prompts and inoculation address different points in the communication process. Pennycook and colleagues examine interventions that shift attention toward accuracy before sharing [26]. Roozenbeek and colleagues test psychological inoculation against manipulative communication techniques [27]. Both approaches offer mechanisms worth investigating in India, but translation alone does not establish equivalence across languages or political contexts. A programme could improve recognition of a manipulation technique without correcting a specific procedural rumour. It could also increase general suspicion. Table 3 therefore pairs each intervention with a proposed outcome and a limitation, rather than rank treatments using incomparable effect measures.

Delivery conditions are especially important on encrypted platforms. Shahi and Hale's study of multilingual WhatsApp tiplines examines submitted claims and fact-checking responses during the 2021 Indian assembly elections [15]. Such material helps reveal verification demand and workflow, but tipline users are self-selected and do not represent all message recipients. A verified response also has limited value if it arrives after the relevant decision. Evaluation should therefore record the interval between submission, verification, response, and onward communication. Researchers should assess understandable explanations and correction reach alongside speed, since an immediate but poorly supported answer can damage credibility.

Interventions should be evaluated against both targeted benefits and possible costs. Removing content may reduce visibility, while mistaken removal can restrict legitimate speech. A warning label may help users interpret synthetic material, but absence of a label must not be treated as proof of authenticity. Forwarding friction may slow harmful distribution while also impeding urgent legitimate messages. These are design tradeoffs requiring measurement rather than assumptions that any constraint improves democracy. Table 4 presents a proposed risk-response matrix. Its priorities are normative recommendations based on the framework, not an empirically optimised ranking of policy instruments.

Table 3. Intervention mechanisms and evaluation requirements

Intervention	Evidence base	Outcome to test	Main boundary
Media literacy	Indian experiments [12, 13]	True–false discernment and persistence	Sample and delivery differences
Fact-checking	Comparative experiments [25]	Targeted belief accuracy	Not an Indian electoral effect
Accuracy prompts	International experiments [26]	Quality of subsequent sharing	Local transfer untested here
Inoculation	International experiments [27]	Recognition of manipulation	Technique recognition is not trust
Multilingual tiplines	Indian tipline study [15]	Time, comprehension, correction reach	Self-selection and incomplete coverage
Synthetic-media labels	ECI advisory [23]	Label recognition; source attribution	Policy expectation, not effect evidence

Outcomes are proposed evaluation targets. No cross-study ranking or common effect size is implied.

8 Governance and institutional communication

A useful governance strategy begins with the claim and the potential harm. False voting instructions require a different response from disputed interpretations of economic performance or political satire. Verification should identify the factual proposition, supporting evidence, relevant jurisdiction, and urgency before assigning a response. Where a matter remains uncertain, communication should state that uncertainty rather than substitute institutional authority for demonstration. Figure 5 presents a proposed response cycle with a review route. Its purpose is to make the reasons for action inspectable and correctable, including when the original classification proves mistaken.

Campaign transparency should identify who sponsors, produces, or materially alters political communication. Synthetic-media labelling is relevant here, but provenance and factual accuracy remain distinct. A clearly identified campaign can still make a misleading claim, and authentic footage can still be presented out of context. The Election Commission's 2025 advisory supplies a concrete documentary reference for examining disclosure expectations [23]. Evaluation should ask whether labels are prominent, understandable, persistent across reposting, and applied comparably across political actors. This article proposes those criteria; it does not claim that a comprehensive audit against them has already been performed.

Institutional communication should make electoral procedures understandable before a controversy occurs. Explanations can specify the relevant stage, responsible authority, records available for inspection, and route for raising a concern. When correcting a false allegation, the response should link to the applicable evidence and distinguish a local incident from a national inference. The Commission's 2024 and 2026 communications demonstrate that social media responsibility and rapid action are explicit institutional concerns [22, 24]. They do not remove the need for independent assessment of accuracy, proportionality, and accessibility. Publicly reported action should be evaluated through those criteria rather than through content-removal totals alone.

Platforms and civil society have complementary responsibilities. Platforms can provide accessible reporting, meaningful explanations for moderation, and information about political advertising. Journalists, researchers, and verification organisations can investigate claims, identify recurrent narratives, and translate procedural explanations. Independence matters because credibility can weaken when an organisation is perceived to protect one political side. Appropriate safeguards include published correction policies, transparent methods, and acknowledgement of unresolved evidence. These recommendations follow from the article's accountability framework; they should not be read as findings that any particular Indian organisation currently satisfies or violates all of these standards.

Privacy should remain part of democratic resilience. Encrypted messaging protects personal, journalistic, and political communication, including communication by vulnerable groups. The difficulty of observing misinformation in private channels does not by itself justify unrestricted monitoring. Consent-based data donation, privacy-preserving matching of verified claims, and aggregated reporting offer research possibilities, although each requires ethical review and careful protection of third parties. Governance should assess the narrowness and necessity of data collection alongside its expected value. The objective is a response that addresses demonstrable harm while retaining the practical conditions for legitimate organising, disagreement, and confidential communication.

Table 4. Proposed responses to distinct information risks

Risk	Proposed response	Accountability safeguard
False voting instructions	Verify urgently; distribute correct local procedure	Named source, timestamp, accessible languages
Misleading edited footage	Restore context and link original material	Explain edit; preserve legitimate commentary
Deceptive impersonation	Authenticate source; label or act proportionately	Evidence record; review and appeal
Unsupported counting allegation	Publish relevant procedural evidence	Allow documented complaints and scrutiny
Opaque campaign sponsorship	Disclose sponsor and material alteration	Comparable rules across political actors
Erroneous moderation	Review, correct, and explain decision	Accessible remedy and error reporting

Author-proposed policy matrix informed by [5, 6, 22, 23, 24]; it is not a statement of applicable legal duties.

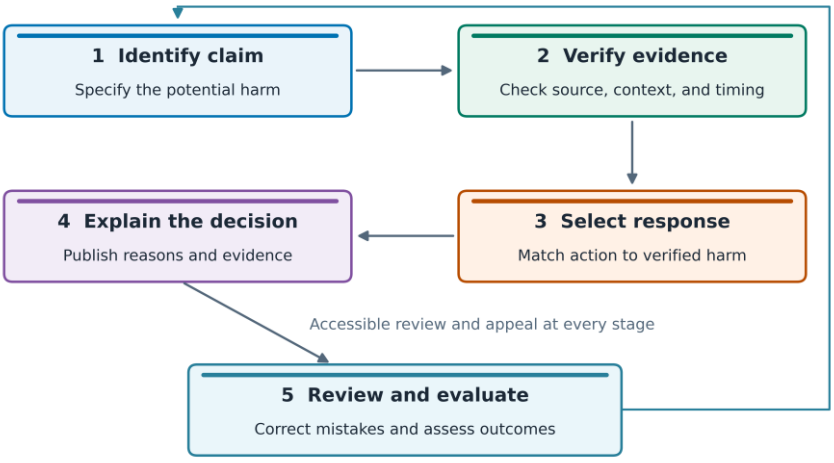

Figure 5. Accountable response to misleading political information.

Original policy framework informed by [5, 6, 22, 23, 24]. The return path represents review and revision. This is a proposed workflow, not a statement of legal requirements.

9 Discussion and research agenda

The synthesis supports a conditional account of social media and Indian democracy. Mobilisation and misinformation are connected through the same communication infrastructures but should not be treated as equivalent activities. A network that distributes inaccurate material may also provide political information, community support, and corrective messages. Likewise, an increase in expressed distrust may reflect manipulation, informed criticism, or improved visibility of genuine concerns. The analytical task is to establish which process is operating. The paper's contribution is to organise those possibilities around distinct

outcomes and mechanisms, making assumptions visible before researchers or policymakers attribute democratic change to a platform.

This framework also clarifies the difference between descriptive visibility and causal influence. A large number of views shows potential attention under a platform's measurement rules; it does not establish unique exposure, attentive processing, belief change, or additional votes. Observational content studies remain valuable for identifying narratives and communication patterns, but their findings should be linked cautiously to population behaviour. Lorenz-Spreen and colleagues' comparative review reinforces the importance of research design when evaluating digital media and democracy [4]. Figure 4 is therefore an inference guide, not a ranking that dismisses qualitative or descriptive evidence, which can reveal mechanisms that experiments fail to capture.

Future Indian research should begin with explicit population definitions. A study of public political groups, an English-language online sample, and a rural community each addresses a different population. National claims require a defensible sampling strategy, coverage of relevant languages, and clear treatment of citizens who use social media infrequently or through others. Researchers should document recruitment, missingness, and attrition, and report the limitations of platform access. Table 5 proposes research questions and designs without inventing sample sizes. Power calculations should follow the selected outcome and design rather than a generic expectation that a large participant count guarantees validity.

Longitudinal designs would strengthen the study of trust. Repeated measurement can establish baseline procedural knowledge, prior institutional confidence, campaign exposure, and reactions after results. Where ethical and feasible, randomised corrections or prompts can test bounded mechanisms within such panels. Outcomes should include factual accuracy, confidence calibration, actual or consent-observed sharing, and distinct trust dimensions. Preregistered analysis and transparent reporting would reduce selective emphasis on favourable findings. Researchers should avoid promising measurement of vote choice or private behaviour where data access cannot support it, and should protect respondents from political identification through released records.

Multilingual research requires more than literal translation. The same phrase can carry different implications across regions, and a manipulated video may combine spoken language, on-screen text, gesture, and culturally specific references. Annotation should therefore preserve context and involve competent language users. Researchers should distinguish disagreement about political interpretation from uncertainty about factual accuracy and report adjudication procedures. Automated classification can assist discovery but should be checked against appropriately labelled material. These are methodological proposals arising from the review's identified gaps, not evidence that a single classifier or annotation scheme already provides reliable coverage of India's communication environment.

Finally, governance itself should become an object of empirical evaluation. Researchers can compare response times, explanation quality, error correction, and accessibility across languages and electoral stages. Case studies should include justified interventions, contested decisions, and acknowledged mistakes rather than select only successful examples. Public documentation can be combined with interviews and consent-based user research to assess whether people understand institutional explanations. Such work would connect operational performance with democratic outcomes without equating administrative activity with legitimacy. The central standard is whether communication and oversight help citizens make informed, contestable judgements about political claims and electoral performance.

An additional priority is to distinguish persuasion from mobilisation of existing supporters. If a campaign message prompts an already committed voter to attend a meeting, the observable behaviour may change without any change in political preference. If it prompts a previously disengaged citizen to seek information, the consequence differs again. Research should therefore measure baseline engagement and specify which transition it seeks to explain. This would help separate recruitment, reinforcement, and conversion, and make comparisons between public video platforms and interpersonal messaging more informative than a single measure of political engagement.

The temporal relationship between verification and political action also deserves closer study. A correction delivered before a scheduled vote, after a campaign event, or following declaration of a result may address the same factual claim while serving different practical purposes. Researchers should record when the claim first became visible, when respondents encountered it, and whether a relevant decision had already occurred. Such timing measures would help explain why an apparently accurate intervention sometimes has limited behavioural consequences. They would also discourage evaluations that judge success solely by the number of published corrections.

The framework has implications for how journal articles communicate uncertainty. Authors should identify whether a claim concerns the existence of misleading material, its frequency in a sample, individual susceptibility, or institutional consequences. These are separate contributions, and none needs to be inflated to matter. For example, identifying a recurrent false procedural message can justify a targeted information response even without establishing changed turnout. Conversely, showing improved recognition of a false headline does not warrant a claim that electoral integrity has been restored. Precision strengthens practical relevance because responses can then match the evidence actually available.

Comparative work across Indian states could examine how the same communication practice operates under different party competition, language environments, and organisational conditions. Such comparisons should justify case selection and avoid treating one state as the national default. Differences in observed misinformation may reflect sampling access or verification capacity rather than underlying prevalence. A useful design would combine common measures with locally informed interpretation and document deviations openly. The objective would be explanation of variation, including null findings, rather than a league table of supposedly more or less digitally vulnerable populations.

Table 5. Research agenda for mobilisation and electoral trust

Research question	Suggested design	Primary outcome
Does exposure precede trust change?	Multilingual pre-election and post-election panel	Dimension-specific trust adjusted for baseline
Which corrections remain effective?	Preregistered randomised study with delayed follow-up	Accuracy, confidence calibration, persistence
How does messaging support mobilisation?	Consented interviews and behavioural observation	Recruitment, reinforcement, participation
Do labels aid authentication?	Language-specific randomised comprehension study	Source recognition without blanket scepticism
Who is absent from available samples?	Coverage audit and targeted recruitment	Representation and nonresponse patterns
Are institutional responses fair and clear?	Independent documentary and user-experience	Explanation quality, error correction, accessibility

	audit	

Proposed studies only. Sampling, ethics approval, and power analysis must precede data collection.

10 Limitations

The review's principal limitation is its purposive and non-exhaustive source base. The bibliography includes heterogeneous methods and publication types, and the available evidence does not support a national effect estimate. English-language discovery and uneven full-text access may underrepresent regional scholarship and contrary findings. Several foundational Indian studies concern earlier elections; they identify mechanisms and methodological lessons but cannot describe the entire 2026 environment. Official communications supply a contemporary documentary update, while their claims about action remain institutional reports. The conceptual figures and proposed indicators should accordingly be treated as tools for organising enquiry, not validated models or empirical estimates.

The review also cannot determine the relative contribution of social media compared with economic conditions, party organisation, interpersonal networks, or established media to particular electoral outcomes. Nor can it establish whether misinformation is increasing nationally or compare political parties' production of misleading material. Those questions require datasets and identification strategies beyond this article. The intervention discussion is similarly bounded: success on a short-term knowledge task does not establish durable changes in electoral confidence. These limits do not make the evidence unusable; they specify which conclusions are justified and identify the additional research necessary for stronger claims.

11 Conclusion

Social media's democratic significance in India depends on how communication infrastructures interact with political organisation, social relationships, and verifiable institutional performance. The available literature supports neither automatic celebration of digital participation nor a universal account of electoral manipulation. This review distinguishes mobilisation, misinformation, and electoral trust while explaining how they may become connected. Its proposed framework makes language, identity, prior confidence, and institutional responsiveness central conditions. Democratic resilience is best pursued through accessible verification, transparent campaigning, proportionate responses, and meaningful opportunities to challenge both misleading claims and institutional decisions. Future research should test these relationships directly and report the boundaries of its evidence.

References

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

[2] Neyazi, T. A., Kumar, A., & Semetko, H. A. (2016). Campaigns, digital media, and mobilization in India. The International Journal of Press/Politics, 21(3), 398–416. https://doi.org/10.1177/1940161216645336

[3] Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. Information, Communication & Society, 18(5), 524–538. https://doi.org/10.1080/1369118X.2015.1008542

[4] Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. Nature Human Behaviour, 7(1), 74–101. https://doi.org/10.1038/s41562-022-01460-1

[5] Norris, P. (2014). Why electoral integrity matters. Cambridge University Press. https://www.cambridge.org/core/books/why-electoral-integrity-matters/3B5036F5F21D99425CE5BB1C54002D82

[6] Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making (Report DGI(2017)09). Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

[7] Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. Information, Communication & Society, 15(5), 739–768. https://doi.org/10.1080/1369118X.2012.670661

[8] Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. Communication Theory, 16(4), 411–426. https://doi.org/10.1111/j.1468-2885.2006.00280.x

[9] Chauchard, S., & Garimella, K. (2022). What circulates on partisan WhatsApp in India? Insights from an unusual dataset. Journal of Quantitative Description: Digital Media, 2. https://doi.org/10.51685/jqd.2022.006

[10] Pal, J. (2015). Banalities turned viral: Narendra Modi and the political tweet. Television & New Media, 16(4), 378–387. https://doi.org/10.1177/1527476415573956

[11] Udupa, S. (2018). Enterprise Hindutva and social media in urban India. Contemporary South Asia, 26(4), 453–467. https://doi.org/10.1080/09584935.2018.1545007

[12] 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. Proceedings of the National Academy of Sciences, 117(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

[13] Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. American Political Science Review, 115(4), 1325–1341. https://doi.org/10.1017/S0003055421000459

[14] Neyazi, T. A., Kalogeropoulos, A., & Nielsen, R. K. (2021). Misinformation concerns and online news participation among internet users in India. Social Media + Society, 7(2). https://doi.org/10.1177/20563051211009013

[15] Shahi, G. K., & Hale, S. A. (2025). WhatsApp tiplines and multilingual claims in the 2021 Indian assembly elections. Online Social Networks and Media, 49, 100323. https://doi.org/10.1016/j.osnem.2025.100323

[16] 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

[17] Arun, C. (2019). On WhatsApp, rumours, and lynchings. Economic and Political Weekly, 54(6), 30–35. https://www.epw.in/journal/2019/6/insight/whatsapp-rumours-and-lynchings.html

[18] Banaji, S., Bhat, R., Agarwal, A., Passanha, N., & Sadhana Pravin, M. (2019). WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. London School of Economics and Political Science. https://www.lse.ac.uk/media-and-communications/assets/documents/research/projects/WhatsApp-Misinformation-Report.pdf

[19] Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

[20] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

[21] Garimella, K., & Chauchard, S. (2024). How prevalent is AI misinformation? What our studies in India show so far. Nature, 630(8015), 32–34. https://doi.org/10.1038/d41586-024-01588-2

[22] Election Commission of India. (2024, May 6). ECI directs responsible and ethical use of social media platforms by political parties and their representatives [Press release]. Press Information Bureau. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2019760
[23] Election Commission of India. (2025, January 16). Advisory for labelling synthetic/AI generated content used by political parties in election campaigning (No. 491/Social media-SOP/2025/Communication) [Advisory]. Official advisory
[24] Election Commission of India. (2026, April 19). General elections and bye-elections 2026: ECI action on unlawful social media content [Press release]. Press Information Bureau. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2253528
[25] Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. Proceedings of the National Academy of Sciences, 118(37), e2104235118. https://doi.org/10.1073/pnas.2104235118
[26] Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. Nature, 592, 590–595. https://doi.org/10.1038/s41586-021-03344-2
[27] Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. Science Advances, 8(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254
[28] Nyhan, B. (2020). Facts and myths about misperceptions. Journal of Economic Perspectives, 34(3), 220–236. https://doi.org/10.1257/jep.34.3.220