Journal of Digital Media, Communication & Policy
Vol. 1, Issue 1, 2026

Vol. 1, Issue 1, pp. 22-34, 2026

Combating Misinformation on Social Media Platforms: Policy Responses and Platform Accountability in Kenya's Digital Media Landscape

Tarei Bett *^1

^1 Graduate, Kenyatta University
*Corresponding author: journals@isrnjournals.org

Abstract

The issue of social media misinformation poses a governance challenge to Kenya in the form of consumer trust, electoral integrity, platform engineering and constitutional rights. This paper discusses the nature of the development of misleading information in the context of the Kenyan digital media and considers the policy responses and the accountability of platforms. It relies on a critical narrative review of scholarly literature, Kenyan legal tools and institutional reports, as opposed to providing original field data. The discussion relates psychological accounts of misinformation to the political economy of digital sites and the multilingual population of Kenya. It claims that making a wide-ranging definition of falsehoods criminal is an unsatisfactory basis on information integrity, especially after the Court of Appeal ruled on false-publication offences in March 2026. A more justifiable solution is a combination of accurate legal safeguards, open platform procedures, responsible advertisement, free research, and easy corrections and long term media literacy. The article suggests an integrated system and allocates roles to governmental agencies, forums, newspaper and televisions and researchers and civil society without sacrificing privacy and civic engagement. Its major finding is that responsibility should be focused on the systems that enhance the negative information and give substantial solutions as well as work on enhancing personal judgment.

Key words: *misinformation, platform accountability, Kenya, digital governance, freedom of expression, social media*

PDF Download | Cite this | Share

Article History

Received: 24 June 2026 | Revised: 24 July 2026 | Accepted: 22 August 2026

1. Introduction

Social media has increased the possibilities by which Kenyans access news, debunk official narratives, coordinate action and commentary on civic matters. False claims, misleading political campaigns and audiovisual content are transported through the same communication infrastructure at an incredibly high speed. The Kenyan policy dilemma is not that limited to the deletion of false posts. It is about the circumstances in which the citizens are exposed to information, the institutions that verify it and the business systems that decide on the visibility. Proper responses should safeguard individuals against consequential

© 2026 The Authors | Published by ISRNJ

deceit but retain the transparency that renders digital participation worthwhile. This paper proposes an accountability model that includes constitutional protections, transparency on platforms, and homegrown public resilience.

Misinformation here is the information that conflict with the best available evidence in a given situation. The disinformation element introduces intentional misinformation, yet with the post alone, intent would usually be hard to verify. Malinformation is the use of true information in malicious manner such as the revelation of personal information. Wardle and Derakhshan (2017) differentiate these categories to avoid the situation when policy considers all information problems as the same. Vraga and Bode (2020) also highlight the significance of knowledge and evidence in identifying misinformation. Practically speaking, a fake announcement of an election, a false report that will be fixed and a provocative opinion will need various evaluations and solutions. A good frame-work should not eliminate such differences.

The term fake news is less practical as it may be used to refer to fake news or a political attack on honest reporting. Tandoc et al. (2018) record the heterogeneity of practices that fall under the term expression, such as fabrication, manipulation and satire. Policymaking based on this elastic tag endangers to mix up deceit with condemnation. This article, therefore, dwells on verifiable claims, orchestrated deceptive actions and foreseeable harms. It considers the information integrity as the capacity of an open communication space to promote credible knowledge and informed involvement as opposed to full adherence to official utterances. Information published by the government, announcements of the platform, and professional reporting can be scrutinized.

Kenya offers an important location due to the fact that participation in digital space has been in interaction with the existing political disparities and institutional mistrust. Nyabala (2018) contextualizes online communication in the broader history of politics in the country, demonstrating that technology cannot be viewed in isolation of power. This view can be used to understand why misinformation can be convincing in the presence of corrective information. Social relations, political loyalty and experience of the public institutions are some of the ways that audiences judge messages. A fact-check on an untrusted organisation can have minimal impact, whereas a message without a substantiation by an acquaintance can seem true. The right action thus goes beyond merely providing more facts to the enhancement of the credibility and availability of the information environment.

The article takes a critical narrative review and policy analysis. It is based on peer reviewed research on misinformation, moderation, literacy and digital political communication; scholarly books on platforms and Kenyan politics; pertinent legal literature and reports on organisations studying information disorder. The sources were chosen based on their conceptual input, applicability to Kenya or relevance to platform governance. Kenyan cases serve as contextual evidence, and international studies are informative about the potential mechanisms and interventions. The article does not purport to be a comprehensive systematic review, neither does it estimate the prevalence of misinformation in the country nor does it present interviews. Its recommendations are analytical proposals based on the reviewed literature and principles of the constitution.

Evidence and inference are of particular importance in this area. Experimental results in other countries do not necessarily determine what will be effective with Kenyan audiences. Similarly, a harmful campaign documented is an indication of failure that needs to be investigated but cannot indicate that all users saw or thought it. Misinformation research also often uses headline judgments or sharing intentions rather than voting behaviour, health decision or real violence. The interpretation of the article is informed by these limits. Policy needs to react to plausible risks without believing hypothetical causal assertions. Reach, belief, consequences of behaviour and institutional damage should be evaluated individually.

The analysis is categorised into five major sections. The second, after this introduction, looks at the functioning of misinformation in the digital space in Kenya. The third considers the legal and institutional reaction, such as the constitutional boundaries of comprehensive false-information crimes. The fourth takes into account platform responsibilities and practical accountability mechanisms. The fifth suggests a comprehensive policy framework and ends with the implementation and research priorities. Collectively, these sections answer a key question: in what ways can Kenya minimize malicious deceit without focusing too much power over popular communication on either government agencies or technological firms in the private sector?

2 Misinformation in Kenya's Digital Media Landscape

This is because misinformation is being spread across a networked media environment as opposed to solitary applications. False information can start in a chat room, be shown in a YouTube video, be discussed by an influencer and then be discussed on the radio. Screenshots are able to strip away source data, and truncated videos hide context required to understand something. This movement between the old and new communication channels can be explained with the help of Chadwick (2017) account of the hybrid media systems. In the case of Kenya, this means that platform-specific interventions should be aware of the broader dissemination of claims. Deleting a single upload can leave copied content, recollected stories and offline conversation intact.

There are various explanations in the psychological literature as to why false information withstands criticism. Repeating something that has been previously said may be more believable, although repetition provides no additional proof. As shown by Fazio et al. (2015), people are not always safeguarded by knowledge against this effect. Information that fits a predetermined interpretation can also be promoted by political or social relevance. The mechanisms cannot be interpreted as the sign that Kenyan audiences are particularly prone. Instead, they pinpoint generic vulnerabilities which may be combined with local conditions. The systems of verification should take into consideration repetition and social endorsement rather than the belief that users scrutinise all claims with a conscious action of checking evidence.

Sharing and belief are connected yet different results. A person can tell a questionable tale to show solidarity, amuse others or caution them, even without thinking that it is true. The article by Pennycook and Rand (2021) describes how attention and reasoning failures are related to misinformation engagement. This difference alters the design of interventions: enhancing factual knowledge might not be sufficient to resolve impulsive redistribution. An immediate message encouraging accuracy may disrupt the habitual sharing, but a long correction might be too long. Both source evaluation and responsible communication, such as the principle that uncertainty is a reason to wait before forwarding, can thus be taught through Kenyan literacy initiatives.

Digital distribution is also more inclined to material that draws instant attention. Vosoughi et al. (2018) discovered that the false news propagated more and faster than true news in the Twitter dataset they examined. Their results cannot be translated into a Kenyan prevalence estimate or supposedly represent all platforms. They, however, explain why the use of organic circulation can put at a disadvantage deliberate checking. The separate diffusion of moral-emotional language with political networks is related by Brady et al. (2017). When rewards on striking content are provided by users and recommendation systems, misleading claims are viewed before contextual reporting can develop. Governance should thus look into amplification as well as accuracy of individual messages.

These dynamics are especially consequential in an electoral environment in Kenya. In an investigation on TikTok by Madung (2022), the author has recorded deceptive and incendiary political content about the 2022 elections. Mozilla Foundation (2022) then announced that a number of platforms failed during the election period. These inquiries develop tangible responsibility inquiries regarding execution, neighborhood

setting and the correlation amid what is vowed and seen. They fail to prove the fact that online misinformation dictated the outcome of elections. To be more cautious, it is necessary to independently monitor electoral information systems since even in the cases where the direct behavioural impact of a failed claim of procedural fault or a crafted political narrative remains unknown, a lack of confidence can be eroded.

Some disinformation is characterised by organised production as opposed to normal errors. Madung and Obilo (2021) examined operations of paid influence with respect to Kenyan public debate. Where campaigns entail co-ordinated posting, hidden sponsorship or impersonation, corrections do not impinge on their organisational motives. Actors can keep making new claims at a higher rate than single stories can be verified. The study of political communication also associates the issue of digital manipulation with the overall transformation of campaigning and the influence of media (Bennett and Livingston, 2018). The policy should then look into funding, contracting and distribution networks in Kenya. Sponsor and campaign organiser accountability is an addition to platforms offering paid reach or rewarding misleading engagement.

Additional challenges are brought about by language and culture. Kenyan communication comprises of English, Kiswahili, Sheng and others, mixed up in one post. Significance may rely on a proverb, ethnic allusion, ironic saying or a locally familiar event. Classification can be trained to rely on other linguistic environments, and thus fail to detect threats or misunderstand benign speech. Moderation studies also point out that contextual judgment is still the focus even with the use of technical systems by the companies (Gillespie, 2018). In the case of Kenya, sufficient enforcement necessitates language skills and contextual escalation pathways. Translation is a handy tool but cannot be used to replace reviewers who can decode coded meanings and political relations.

Misinformation is also overlapping with the lack of equal access to trustworthy options. Verification can be tedious due to expensive connectivity, inadequate time and skewed digital capabilities. An individual who gets a compressed video via a known group might not have access to the original source or searchable archive. The conditions are to be studied empirically instead of being ascribed to the audiences without evidence. They however imply a principle of policy: people must have access to reliable information in a form that can be practically utilized. Online fact-checking can be supplemented by short audio explanations, low-data corrections, community radio, and any other kind of available institutional notices. Part of it is infrastructure and inclusion issue of information quality.

Elections are not the only consequences. Misleading health claims can hamper proper treatment, falsified job offers can exploit jobseekers and financial impersonation can lead to economic loss. These types demand varying professional skills and response. Lazer et al. (2018) suggest the multidisciplinary analysis of misinformation, and Ecker et al. (2022) provide an overview of the cognitive and social factors that render false beliefs hard to update. The response that Kenya should take must also be along differentiated lines. Claims in the area of health authority, consumer protection institutions and election institutions can be verified by the respective authorities, but none of them should become the common judge of all disputed public communication.

Trust as a state of misinformation uptake is also a consequence that could be ruined through fraudulent campaigns. Researchers, however, warn that the amount of misinformation that people consume should not be overestimated. In their American study of Facebook, Guess et al. (2019) identified concentrated sharing within specific groups, whereas in their study of misinformation as part of a wider media diet, Allen et al. (2020) did not. Such results stimulate proportional Kenyan study: high-visibility cases must not represent the measurement. The level of exposure and damage should be taken into account, and priorities of intervention must be based on it. Exaggerating the issue may be an inadvertent way to foster a lack of trust in journalism and justify overly restrictive policies.

Lastly, synthetic media heightens the requirement to verify but not all misleading images are technical advancements. Vaccari and Chadwick (2020) relate political deepfakes to the uncertainty and trust, highlighting the dangers beyond persuasion. Fabricated audio, altered video and impersonation must be dealt with in Kenyan policy but ordinary editing, satire and artistic expression should still be appreciated. Provenance tools can aid in proving origin or modification, however, lack of this information cannot be used to prove falsity. The broader difficulty is that of contextual verification. Recording a real image with a fake caption can be more deceptive than a synthetic image that is clearly fake, hence enforcement should also take into account the claim made.

3 Policy Responses and Constitutional Constraints

The Constitution of Kenya has defined the beginning of misinformation governance. Article 33 guarantees the right to the freedom of expression but in exclusion of certain categories including incitement and hate speech; Article 34 upholds the freedom of the media; Article 35 covers the right to access information and Article 31 involves protection of privacy (Constitution of Kenya, 2010). These rights have to be taken in unison. The people should be insured against harm done specifically to them and they ought to have an opportunity to challenge those in authority. Article 24 demands the justification of the restrictions of rights. Policy should in turn have obvious rules, valid ends and reasonable restrictions instead of an overall assumption that speech that is controversial is illegal.

In the Computer Misuse and Cybercrimes Act, 2018, the offences related to false publications and publication of false information were introduced and are provided in sections 22 and 23. Those clauses brought about constitutional litigation. The Court of Appeal on 6 March 2026 held both sections to be unconstitutional as being too broad in Bloggers Association of Kenya (BAKE) v. Attorney General and 6 others (2026). It is a material policy development: when discussing the response of Kenya it must not be treated as a mere repetition of the original offences, which should form a ground on which they are executed now. It also explains why the intention to fight misinformation does not eliminate the precision when the criminal responsibility is suggested.

The meaning of that ruling goes beyond the language of two clauses. Widespread false-information crimes have the potential to render confusion among journalists, common users and people trying to narrate swiftly emerging events. A prosecution that is not followed by a conviction can be very costly and not encourage involvement. This article views the ruling as in favor of more selectively targeted governance, and not in favor of all misleading communications. Other illegal actions are not out of law and constitutional protection. Platform obligations and civil remedies, privacy protections and high-specific responses to harmful behavior provide separate paths that should be assessed separately.

Hate speech and misinformation are similar, but they should not be merged into one. In the National Cohesion and Integration Act, 2008 in section 13, issues of threatening, abusive or insulting behaviour related to ethnic hatred are stated. The assertion of a lie will incite hostility, yet not all such statements meet the legal criteria of hate speech. Similarly, the use of accurate information can be harmful incitement. The conduct, the provision and the evidentiary threshold applicable must be determined by the regulators and investigators. When the ideas are confused, there is a danger of over-restraint of grave dangers, and overrestraint of mere mistakes. These differences can be explained through well-developed examples that can be given by the public.

An additional policy dimension is data protection. The Data Protection Act, 2019 governs the processing of personal-data and creates the responsibilities that are applicable to profiling and targeted communication. The misinformation interventions in themselves have the potential to produce privacy risks, when it comes to intrusive monitoring or gathering of sensitive political information. The interest of a platform in determining coordinated campaigns does not imply that it can unlimitedly retain personal data. Similarly,

investigators must have the right protection in exploring vulnerable populations. Aggregated records, restricted access and definite retention times can be utilized in privacy-conscious analysis. The implementation of data protection may also deal with the elements of the deceptive targeting, yet it does not necessarily determine whether a certain statement is true.

Preventive response that is underutilized is access to information. Transparency is based on the Access to Information Act, 2016, and timely disclosure will decrease the uncertainty of the public decisions. Independent verification is easier in an institution where clear procedures, source documents and corrections have been published. The proposed policies in this article are that agencies in charge of elections, health and public finance should have available clarification channels and clarify any change in evidence. This communication should be differentiated with partisan messaging. An official source is in authority as long as it is within its competency though its assertions can be evaluated against records and independent reporting. Verification is enhanced by transparency, since it can be scrutinized, rather than trust being required.

Co-ordination of institutions is required since issues of information cut across mandates. The regulation of communications, data protection, election administration, media standards and criminal investigation are performed in the framework of various legal rules. An integrated process can enhance the process of sharing referrals and information without bringing all functions under one regulator. Institutions that are open to the public should have written guidelines as to which organ is approached with which complaint and what redress can be provided. The independence is particularly relevant when the contentious matter revolves around government actions. The management of oversight should disallow the officials to make informal requests to moderation that would influence the moderation quietly. Such interaction would be more visible with public disclosure of legally permissible requests.

International rules are useful principles which do not form part of Kenyan laws. UNESCO (2023) promotes governance that is human rights-centred, transparency and engagement, and the African Commission on Human and Peoples Rights (2019) places the rights of expression and information in the African context. These tools can enlighten norms and citizen engagement. They however need to be implemented in a way that takes into consideration the Kenyan administrative capacity and constitutional arrangements. The imported regulatory frameworks can be costly resources that are not sustainable by the small institutions. An effective framework must first set up the necessary responsibilities, show enforcement ability and increase the requirements as technical capacity and independent control increase.

One of the comparative examples of duties related to platform processes and systemic risks can be found in the European Union Digital Services Act (European Parliament and Council of the European Union, 2022). Kenya can also emulate its transparency and accountability focus without wholesaling its institutional framework. The policy question of interest is what obligations can be evaluated and effectively implemented at the local level. A Kenyan framework would emphasise effective complaint systems, disclosure of advertisements and risk reporting over comprehensive certification systems. Public involvement should be used to develop any reform, and draw the line between the proposed obligations and the duties that are already enforced by the current Kenyan law.

The state should also be subject to public accountability. Misinformation fighting agencies can themselves make false statements, withhold material evidence or label criticism a deception. A reliable system must have a method of correction, disclosure of sources and scrutiny of the published communication. Literacy programmes funded by the government must instruct how to verify any claims, even those of officials. The judicial system, the press and civil society organisations must be able to investigate enforcement without reprisals. Such symmetry is part of the legitimacy of the policy of misinformation: citizens are more willing to accept intervention that is carefully restricted when such institutions are subjected to similar calls to accuracy, explanation and accountability.

4 Platform Accountability and Practical Responsibilities

Platforms have power in terms of ranking, advertising, recommendation and moderation. What they are in charge of can not be evaluated solely by the fact that a user made a misleading content in the first place. Gillespie (2018) focuses on the concept of moderation as an element of platform governance, and Gorwa (2019) explains the complicated nature of relations between companies, states, and other participants. Accountability in Kenya should enquire what a service was aware of a risk, how it was designed to distribute, whether it was reasonable in how it responded or not. This is because a company can collapse due to uneven application, or insufficient local resources or advertising methods despite being unable to ensure the accuracy of all messages.

The meaning of assessment should be backed by transparency reports. The number of posts erased around the world cannot tell us whether Kenyan users were promptly safeguarded or whether they selected content in local languages. Report received, decisions made, appeals resolved and major policy categories would be useful reporting. It would also describe data constraints and measurements alterations. The article suggests country-level reporting where possible, with the help of language-sensitive analysis and privacy protection. These are suggested standards of accountability, not purported to be provided by all the current platforms. Scientists should be capable comparing reporting periods without confusing any change in counting techniques with any change in real injury.

The number of removals is a poor performance indicator. An increased number can indicate improved detection, more malignant action or extended enforcement that unjustifiably impacts lawful speech. On the other hand, a small number can be an indicator of prevention or neglect. Better evaluation takes into account the approximate prevalence of the violating content, the exposure prior to the action, the response time, the reversal rates and the ramifications of being a law abiding participant. Platforms might host assessments carefully sampled with estimates of uncertainty and methodology descriptions. Claims of improvements would then be tested by independent reviewers. This strategy makes the focus of the discussion on the metrics of publicity to the evidence of the experiences of the users and the allocation of the harm.

The moderation materials need to be in line with Kenyan linguistic and political contexts. The reviewers should be properly trained, having access to the contextual knowledge and escalation mechanisms in case of uncertain cases. According to Roberts (2019), the labour behind the commercial moderation is highlighted, and the workers conditions form the part of the accountability question. Hasty and uninformed critics can make unequal decisions and viewing harmful content can create significant personal expenses. Platforms ought then to post valuable information on the capacity of relevance but not to name the workers. Quality is maintained with the assistance of labour protections, workable workloads, and professional support. Local experience should also be safeguarded against political influence and measured by objective, transparent criteria.

Technical automation may assist in detecting duplication of material, suspicious coordination and possibly fraudulent claims, but cannot substitute contextual judgment. Gorwa et al. (2020) state the political and practical constraints of automated moderation. Kenyan systems have to be evaluated with regard to languages, and types of content, such as speech, captions and pictures. False positives and false negatives should be tested. The device that successfully identifies some dangers but constantly silences the lawful minority-language discourse is unsuccessful. Consequential decisions should still be reviewed by humans, especially when a restriction of an account may have an impact on journalism, livelihoods or reach to an established audience.

Various risks should be matched with various interventions. Platforms can employ source context, sharing prompts, distribution restrictions, labels, removal or account action. Proper decision is determined by the evidence clarity, potential harm extent and relevant policy. Pennycook et al. (2021) present evidence that

the accuracy-focused approach can enhance sharing decisions, whereas Roozenbeek et al. (2022) show that inoculation strategies have potential. These results imply that there are other possible alternatives to deletion. The Kenyan trials should consider how such tools can be comprehended and used in local languages, and how their advantages can be maintained once the immediate intervention is over.

Caution signs should be carefully designed. An unexplained assertion of falsity is not as helpful as a label describing the controversial assertion, and giving some evidence in support of it. According to Pennycook et al. (2020), there is an implied-truth risk: users might assume that unlabeled content has been authenticated. Platforms must clarify that lack of a warning does not mean that it is accurate. Appeals are also essential since decision-making on fact-checks and enforcement may be flawed. Infected users require clear explanations, availability of the applicable rule and a means of human reconsideration. Whenever possible, an erroneous restriction should be promptly restored and clarified.

Advertising should be put under greater responsibility since the platforms not only determine paid distribution but also earn income on it. Proposed Kenyan standards must include sponsor identification, available advertising archives and disclosure of the relevant information on targeting and safeguard of personal data. Paid dissemination should be investigated by platforms before fraudulent impersonation and hidden political sponsorship. Allcott and Gentzkow (2017) relate fabricated news to incentives in the broader media market, which explains the importance of economic arrangements. An algorithm that eliminates the misleading regular posts and promotes reach to misleading campaigns is in itself inconsistent. The hosting company should be subject to equal scrutiny as the advertisers, intermediaries and political sponsors.

Company claims would be more testable with access to independent research. The proper access may involve the aggregate exposure data, advertising archives and secure datasets to approved research. It must not publicize intimate messages and identities at random. The ethics of research, security measures and well defined intent are necessary. Scientists should also be left to publish negative results without being influenced by business. Van Dijck et al. (2018) relate platform systems to publicly meaningful values, thereby supporting the notion of external criticism. Universities and civil society groups can play a role in Kenya but sustainable access should not rely solely on the temporary corporate relationships.

Privacy messaging has other limitations compared to recommendation feeds. End-to-end encryption secures confidential communication, and should not be compromised just to allow routine monitoring. Other possible reactions are user controlled reporting, mass forwarding restrictions, verification services and accessible information annexed to self-voluntary submitted claims. The tools have to be evaluated since willful actors can change. Where they are credible, community leaders can aid the dissemination of corrections, but should not become informal surveillance agents. Responsibility must acknowledge the design of any service but demand that privacy protection and reasonable prevention can be sought simultaneously.

Finally, platforms must also post crisis procedures during elections and other risky times. These procedures may designate escalation contacts, risk assessment localised to location and temporary measures to be reviewed. They should not be cloudy in the political intervention and should be open to subsequent independent evaluation. Emergency measures must have purpose, reasons, and expiry reasons and conditions. Major governance should go on between election periods since paid influence networks and fake financial content do not vanish during the periods between elections. Sustainable accountability involves the perpetuation of institutional responsibility and not a high profile commitment and little report on what was really done by companies.

5 Conclusion and an Integrated Policy Framework.

Kenya requires an overlay structure that unites legal accuracy, accountable design and social fortitude. The first one is the common nomenclature that will differentiate between the claims that are erroneous, the acts of deliberate misrepresentation, the unlawful incitement, the cases of impersonation and of legal disagreement. All categories must be related to an appropriate response and standard of evidence. Criminal enforcement falls under well-defined law and constitutional protection; platform policy can deal with other harms with balanced contractual regulations. Uncertainty can be clarified in public without causing punishment. This model of distinction minimizes the urge to consider all contentious information as a single issue and simplifies responsibility to evaluate.

The second priority is procedural accountability which is enforceable. The possible standards of reporting, handling of complaints, disclosure of advertisements and independent evaluation should be determined by the public. A regulator or coordinating mechanism needs to possess a strictly narrow mandate, technical ability and immunity against political guidance. Its rulings must be rationale and revocable. Smaller services must have responsibilities that are proportional to their size and risk and larger platforms must have greater assessment and mitigation capacity. An authoritative framework would also issue its own enforcement history, such as appeals that have been won and errors that are recognized, and thus oversight is not another veil of information management.

The third priority is the continued assistance to independent verification. The newsrooms and fact-checkers require access to source material, investigative abilities and practical funding plans. Grant structure must safeguard editorial freedom and not to be paid based on favored political findings. Corrections are required to recognize the claim, determine what evidence demonstrates and clarify pertinent uncertainty. Lewandowsky et al. (2012) also note that a good correction design is significant. In the case of Kenya, partnership with broadcasters and community organisations can further reach out to audiences that already subscribe to the fact-checking websites. Verification must be an ongoing civic duty and there should be explicit correction policies to the verifiers themselves.

Media literacy is the fourth priority, but it must be practical and measurable. The source tracing, comparison of independent reporting, identification of sponsored content and respectful correction in the social relationships should be taught in education programmes. Guess and colleagues demonstrate that discernment can be enhanced with short literacy interventions in certain contexts, and that a lot of differences exist across people over time (Guess et al., 2020). Wineburg and McGrew (2019) endorse the practice of lateral reading in the form of verification. The Kenyan programmes must be adapted to these insights by locally testing them, providing easily comprehensible examples and learning opportunities. Success implies a better ability to discriminate between credible and non-credible information, as opposed to unselective distrusting of all sources.

The implementation process should be carried over in distinct stages. The first step can be the mapping of the institutional requirements, consultation of the concerned groups and setting of baseline indicators of exposure, literacy and complaint management. Another step can pilot reporting standards, advertising archives and locally modified educational materials. Implementation should be expanded after the evaluation and the reasons as to why an intervention should be retained, revised or abandoned published. The issue of funding should be clear enough that programmes are not reliant on election gimmicks. The resources can be contributed by public institutions, platforms and independent organisations, however, the financial relationships should be revealed and should not predetermine verification results and research conclusions.

The assessment is to be based on quantitative and qualitative evidence. Belief accuracy and trust can be measured with surveys, experiments can be ethically designed to test prompts and corrections, platform data

can provide estimates of exposure and response times, and interviews can provide insight into barriers to reporting or appeal. Disaggregation of indicators should be done where necessary and legally, and language and accessibility needs. Wrongful restriction and decreased civic participation should also be captured with measures. Diluting detrimental exposure with avoidance of valid reporting is a heavy price. Transparent evaluation of such trade-offs would require pre-established evaluation questions and independent publication to aid the decision-makers in evaluating the trade-offs.

The research agenda is still high. Kenya requires representative data regarding who is exposed to misleading information, which claims shape behaviour and how information flows across popular sites, messaging and offline dialogue. Research must look at variations across counties, language difference and the sustainability of literacy interventions without necessarily thinking about homogeneous impacts. Researchers ought to differentiate exposure and belief as well as belief and action. Data can be offered by partnerships, but they need to be governed independently and have privacy. The narrative review presented in this article highlights mechanisms and policy directions; it does not provide any causal impact on Kenya or replace the original empirical research.

An effective fight against misinformation in Kenya eventually involves credible institutions and responsible communication systems. The constitutional structure and the 2026 court of appeals decision give grounds to oppose inaccurate treatment of falsehood, and reported malfunctions of the platform should be considered stricter in terms of amplification and execution. The most justifiable policy is a combination of specific safeguarding of illegal damage and open company procedures, external checking, responsible advertising and open learning. Governments need to publish and fix, platforms need to clarify and correct, and citizens need instruments of judgment at their disposal. This is a combination that can reinforce information integrity and maintain the freedom to investigate, disagree and engage in the digital public sphere in Kenya.

References

Access to Information Act, No. 31 of 2016 (Kenya). https://new.kenyalaw.org/akn/ke/act/2016/31

African Commission on Human and Peoples' Rights. (2019). *Declaration of principles on freedom of expression and access to information in Africa*. https://achpr.au.int/en/node/902

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

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances, 6*(14), Article eaay3539. https://doi.org/10.1126/sciadv.aay3539

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication, 33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Bloggers Association of Kenya (BAKE) v. Attorney General & 6 Others, Civil Appeal 197 of 2020, [2026] KECA 430 (KLR) (Kenya). https://blog.bake.co.ke/wp-content/uploads/2026/03/Cybercrimes-Appeal-Case-Judgement-2026.pdf

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

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

Computer Misuse and Cybercrimes Act, No. 5 of 2018 (Kenya). https://new.kenyalaw.org/akn/ke/act/2018/5

Constitution of Kenya, 2010. https://new.kenyalaw.org/akn/ke/act/2010/constitution

Data Protection Act, No. 24 of 2019 (Kenya). https://new.kenyalaw.org/akn/ke/act/2019/24

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*, 13–29. https://doi.org/10.1038/s44159-021-00006-y

European Parliament & Council of the European Union. (2022). Regulation (EU) 2022/2065 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act). *Official Journal of the European Union, L 277*, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gorwa, R. (2019). *What is platform governance? Information, Communication & Society, 22(6), 854–871*. https://doi.org/10.1080/1369118X.2019.1573914

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society, 7*(1). https://doi.org/10.1177/2053951719897945

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances, 5*(1), Article eaau4586. https://doi.org/10.1126/sciadv.aau4586

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

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

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Madung, O. (2022). *From dance app to political mercenary: How disinformation on TikTok gaslights political tensions in Kenya*. Mozilla Foundation. https://www.mozillafoundation.org/en/blog/new-research-disinformation-on-tiktok-gaslights-political-tensions-ahead-of-kenyas-2022-elections/

Madung, O., & Obilo, B. (2021). *Inside the shadowy world of disinformation-for-hire in Kenya*. Mozilla Foundation. https://www.mozillafoundation.org/en/blog/fellow-research-inside-the-shadowy-world-of-disinformation-for-hire-in-kenya/

Mozilla Foundation. (2022, November 3). *Fellow research: Platforms failed to curb misinformation after Kenyan election*. https://www.mozillafoundation.org/en/blog/fellow-research-platforms-failed-to-curb-misinformation-after-kenyan-election/

National Cohesion and Integration Act, No. 12 of 2008 (Kenya). https://new.kenyalaw.org/akn/ke/act/2008/12

Nyabola, N. (2018). *Digital democracy, analogue politics: How the internet era is transforming politics in Kenya*. Zed Books.

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science, 66*(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478

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

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

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

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), Article eabo6254. https://doi.org/10.1126/sciadv.abo6254

Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

UNESCO. (2023). *Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multistakeholder approach*. https://unesdoc.unesco.org/ark:/48223/pf0000387339

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1). https://doi.org/10.1177/2056305120903408

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001

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

Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication, 37*(1), 136–144. https://doi.org/10.1080/10584609.2020.1716500

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

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record, 121*(11), 1–40. https://doi.org/10.1177/016146811912101102