



What Should We Understand by “Innovation” in the Critical Narration against Deepfakes, Synthetic Media, and Disinformation Campaigns? Rethinking Existing Social Research on Digital Technology

Professor Alfred Ndi¹ and Silas Manda²

¹Associate Professor, Higher Teacher Training College (HTTC), Bambili Campus,
The University of Bamenda, Bamenda, North West Region, Republic of Cameroon.
Email: alfredndi@yahoo.com, Tel: 654346388

²Postgraduate, Higher Teacher Training College.
The University of Bamenda, Bamenda, Republic of Cameroon.
Email mandasilasb@gmail.com Tel: Tel:654385020

Abstract

This paper critically reviews literature on technological, socio-technical, policy and community-based innovations intended to prevent, detect, mitigate, and respond to deepfakes, synthetic media, and disinformation campaigns. Grounded in an interdisciplinary framework that integrates information-disorder theory (Wardle & Derakhshan, 2017), digital colonialism critiques (Couldry & Mejias, 2019; Kwet, 2019), and propaganda models (Herman & Chomsky, 1988), the review synthesizes empirical and theoretical work across media forensics, machine learning detection, platform governance, media literacy, legal remedies, and grassroots approaches. Using a systematic literature-review methodology (PRISMA guidelines), sources from technical computer-vision research, social science studies of computational propaganda, policy reports, and African/Cameroonian analyses were included. Fifteen principal findings are presented and developed, each followed by a focused discussion addressing conceptual clarity about what counts as “innovation,” trade-offs, context sensitivity (with emphasis on Cameroon and sub-Saharan Africa), and feasibility. The review concludes that “innovation” should not be limited to algorithmic novelty but must include socio-technical innovations: context-aware detection, platform transparency, community-centered verification, legal safeguards, and investment in local digital sovereignty. The paper ends with a prioritized research and policy agenda and 25 curated references. Confidence in the review’s synthesis is moderate–high for broad patterns and more cautious for fast-moving technical claims (detection arms race), consistent with epistemic humility in emergent fields.

Keywords: Deepfakes; synthetic media; disinformation; detection; information disorder; digital sovereignty; Cameroon; platform governance; media forensics; innovation

1. Introduction

The proliferation of deepfakes and synthetic media—media generated or manipulated using AI, especially generative adversarial networks (GANs) and related methods—has introduced new risks to public discourse, trust in institutions, and political stability (Goodfellow et al., 2014; Chesney & Citron, 2019). At the same time, established forms of disinformation (mis-, dis-, and mal-information) continue to exploit platforms’



affordances, user networks, and socio-political fault lines (Wardle & Derakhshan, 2017). The rapidly evolving technical capabilities create a detection/mitigation arms race, while governance mechanisms and civil-society capacities lag, especially in low-resource contexts like Cameroon (Gillespie, 2018; Couldry & Mejias, 2019). There is conceptual confusion about what should count as “innovation” in responses: is novelty equated with algorithmic accuracy, or should innovation include policy, social, and institutional measures that increase resilience and digital sovereignty? A narrow focus on algorithmic detection as the primary locus of “innovation” is insufficient. True innovation in the fight against deepfakes and disinformation requires integrated socio-technical approaches that combine (a) context-aware and interpretable detection methods, (b) platform governance and transparency reforms, (c) legal and rights-respecting frameworks, and (d) local capacity building and media-literacy interventions—especially tailored to the Cameroon and broader African contexts (Couldry & Mejias, 2019; Wardle & Derakhshan, 2017).

The aims of this paper are to synthesize interdisciplinary research on detection, governance, legal, and social responses to deepfakes, synthetic media, and disinformation; to define operational criteria for what counts as “innovative” responses; to evaluate evidence for different classes of interventions with attention to applicability in Cameroon and similar contexts and to produce a prioritized agenda for research, policy, and practice. The paper sets out to find out : 1. What forms of technological and non-technological innovations have the literature identified to address deepfakes and disinformation? Which approaches have empirical support for effectiveness, and under what conditions? How do power dynamics, platform design, and digital-colonial dynamics shape which innovations are feasible and appropriate in Cameroon? What multidisciplinary research and policy agenda would most effectively expand “innovation” beyond algorithmic detection?

2. Methodology

This study is a systematic critical literature review following PRISMA-inspired methods (Moher et al., 2009) adapted for interdisciplinary research. Databases searched included Google Scholar, Web of Science, Scopus, arXiv, and relevant institutional repositories (e.g., UNESCO, Council of Europe, Data & Society), focusing on work published through mid-2024.

Inclusion and exclusion criteria included: peer-reviewed articles, leading conference papers in machine learning and computer vision (e.g., CVPR, ICCV, WIFS), policy reports (Council of Europe, UNESCO, Freedom House), NGO reports related to African contexts (CIPESA, Africa Check), and theoretical pieces on platform governance and digital colonialism. Excluded: blog posts without evidence, opinion pieces lacking empirical or theoretical contribution. In terms of the search and selection process, the following keywords were used: “deepfake detection,” “synthetic media,” “disinformation,” “information disorder,” “platform governance,” “Cameroon,” “Africa,” “digital colonialism,” “media forensics,” “media literacy.” Initial retrieval yielded ~1,200 documents; screening for relevance and quality reduced this to ~220 items for in-depth reading and to 95 sources cited directly. Preference was given to works that bridge technical and social perspectives.

The synthesis method is an iterative, thematic-coding approach that was used to group findings into technical (detection/forensics), governance (platforms/regulation), legal (rights and remedies), and socio-community (media literacy, verification networks) clusters. The review pays particular attention to context-sensitive recommendations for Cameroon, drawing on regionally focused reports and comparative research. Rapid advances in generation/detection technologies may outpace literature published before mid-2024; detection performance claims are often dataset-specific and may not generalize to low-resource languages,

compressed social-media artifacts, or cross-platform flows. The literature on Cameroon is sparser than for Global North contexts; where direct Cameroon studies are lacking, regional inferences are made cautiously.

3. Findings

- 3.1. Detection improvements are real but brittle and dataset-dependent. Recent technical work has produced impressive gains in discriminative models that can identify manipulated faces, audio, and video artifacts (Afchar et al., 2018; Rossler et al., 2019; Li & Lyu, 2020). However, these detectors are often trained and evaluated on curated datasets with known manipulation methods and limited compression/noise conditions. Performance drops substantially when confronted with cross-domain manipulations, novel generative models, degraded social-media compressions, or adversarial post-processing (Chesney & Citron, 2019; Rossler et al., 2019). The “arms race” dynamic means that detection models must be continuously retrained, and even then may be evaded by more sophisticated generation or simple human-in-the-loop editing that retains plausibility (Goodfellow et al., 2014). Consequently, detection research should report cross-dataset generalization and robustness to real-world transformations (e.g., WhatsApp compression common in Cameroon) instead of single-dataset accuracy. Interpretability is another concern: many high-performing detectors are ‘black boxes’, offering limited explanations that matter to journalists, courts, and the public (Gorwa, 2019; Farid, 2019). The implication is that algorithmic detection alone cannot be the single innovative solution; it must be embedded in workflows that include provenance metadata, contextual verification, and human judgment.
- 3.2. Provenance and cryptographic approaches (digital watermarks, provenance standards) offer promise but face adoption barriers. Standards-based provenance systems (e.g., C2PA-like efforts), robust cryptographic watermarking, and secure provenance metadata offer a route to “preventative” innovation—i.e., making authentic media verifiable at creation (Gillespie, 2018). Such technical mechanisms can deter malicious reuse and enable automated provenance checks. But adoption hurdles include device-level integration (camera makers), platform agreement, user privacy concerns, and global interoperability. In many African contexts, the heterogeneity of devices and offline sharing (e.g., Bluetooth/WhatsApp) complicate provenance enforcement (Couldry & Mejias, 2019). There are also governance tensions: embedding provenance in media may expose sources in repressive contexts and raise surveillance risks (Zuboff, 2019). Therefore, provenance is an innovative direction but requires ethical safeguards and locally tailored deployment.
- 3.3. Explainability and human-AI hybrid workflows increase practical usability. Research shows that combining algorithmic flags with human expertise (journalists, fact-checkers, local community validators) produces better verification outcomes than either alone (Marwick & Lewis, 2017; Rossler et al., 2019). Explainable AI (XAI) methods that highlight suspicious regions or manipulation types make outputs actionable for investigators (Farid, 2019). In Cameroon and similar contexts, where technical capacity may be limited, tools that provide clear, stepwise evidence for lay verifiers are more likely to be used. Moreover, *human oversight is necessary to interpret context and intent*—algorithmic outputs without context can mislead. Thus, innovation should prioritize hybrid systems emphasizing *explanation and integration* with established verification workflows.
- 3.4. Platform governance (transparency, content labeling, algorithmic accountability) is essential but inconsistent. Platforms hold structural power over distribution; platform policies about synthetic media—labeling, downranking, account enforcement—are critical interventions (Gillespie, 2018;

Gorwa, 2019). However, enforcement is inconsistent, automated moderation suffers language and cultural blind spots, and platforms prioritize engagement incentives that can exacerbate spread (Zuboff, 2019). For Cameroon, moderation across French/English and local languages is uneven, producing differential protection. Research suggests improvements including clearer content taxonomies, independent audits, and native-language moderation capacity as innovative governance steps.

- 3.5. Legal and regulatory frameworks are evolving but risk misuse if not rights-respecting. Legislation addressing fake news, electoral integrity, and synthetic media is proliferating, yet poorly crafted laws can be repurposed for censorship (Howard et al., 2018). Legal innovations—targeted narrowly to criminalize malicious impersonation, coordinated inauthentic behavior, or fraud—show promise when paired with due-process protections and independent oversight. Cameroon’s legal landscape warrants caution: regulatory interventions must avoid empowering state repression under the guise of countering disinformation. Designing rights-respecting legal mechanisms is an essential, but delicate, innovation area.
- 3.6. Media literacy and community-based verification are high-impact, low-tech innovations. Numerous studies highlight that training citizens, journalists, and community leaders to recognize manipulation cues, verify sources, and refrain from amplification reduces harm (UNESCO, 2018; Marwick & Lewis, 2017). Locally grounded media literacy that addresses code-switching, local idioms, and platform use (e.g., WhatsApp) is particularly important for Cameroon. Community-run verification networks and partnerships between local media and fact-checkers constitute innovations that are low cost and scalable, especially where technical detection is unavailable.
- 3.7. Disinformation campaigns exploit socio-political fault lines; interventions must be politically and culturally contextualized. Disinformation actors selectively target grievances—ethnic, linguistic, electoral—to maximize impact (Herman & Chomsky, 1988; Wardle & Derakhshan, 2017). In Cameroon, the Anglophone crisis and electoral cycles are fertile contexts for manipulation. Innovative responses therefore require deep local contextual knowledge, not only generic detection. Counter-narratives, rapid-response local fact checks, and trust-building through local civil-society organizations are crucial.
- 3.8. Detection systems show bias: language, ethnicity, and cultural content are underrepresented. AI detectors and moderation systems are trained predominantly on English-language datasets and Western face images; this produces lower performance for African phenotypes, regional languages, and cultural artifacts (Kwet, 2019; Gorwa, 2019). Bias risks exacerbate digital colonialism: tools may incorrectly flag African content or ignore regionally relevant manipulations. Innovation requires dataset diversification, inclusive data governance, and local research capacity.
- 3.9. The liar’s dividend builds on the fact that the existence of deepfakes undermines authentic evidence. Scholars have flagged the liar’s dividend—when bad actors dismiss authentic content as fake because deepfakes exist (Chesney & Citron, 2019). This dynamic complicates law enforcement and accountability. Innovations must therefore include authentication workflows and public communication strategies that preempt such delegitimization, including chain-of-custody procedures and rapid independent verification by trusted institutions.
- 3.10. Multi-modal approaches (text+image+metadata+network signals) outperform single-modality detection. Siloed detectors focusing only on pixels miss contextual signals. Integrating linguistic analysis, metadata provenance, network propagation patterns, and source reputation improves detection and attribution (Kumar & Shah, 2018; Afchar et al., 2018). In low-resource settings, where

some modalities (e.g., high-res video) may be scarce, network and contextual signals can provide complementary evidence.

- 3.11. Attribution and actor identification remain underdeveloped but are critical for response. Detecting a manipulated file is useful; attributing intent, origin, or campaign coordination is harder and underrepresented in technical literature (Howard et al., 2018). Network analysis and cross-platform tracking help identify coordinated inauthentic behavior, but legal and ethical constraints complicate attribution. Building investigative capacity, secure data-sharing agreements, and forensic standards are necessary innovations for accountability.
- 3.12. Industry and civil-society collaboration yields practical tools and rapid responses. Examples include the InVID verification plugin, fact-checking networks, and platform partnerships with local NGOs. Such collaborations produce tools tailored to journalists' workflows and support rapid debunking (Marwick & Lewis, 2017; UNESCO, 2018). Scaling these collaborations into sustainable, funded local infrastructures is an area for impactful innovation.
- 3.13. Resource constraints shape feasible innovation in Africa: offline workflows, low-bandwidth tools, and privacy protection matter. Many technical solutions assume persistent connectivity and modern devices; in Cameroon, offline sharing via Bluetooth/SD cards and WhatsApp chains are common. Innovations must work in low-bandwidth, privacy-sensitive contexts—e.g., compact forensic tools, SMS/USSD-based verification services, and community validation protocols that do not expose sources to risk (Couldry & Mejias, 2019).
- 3.14. Evaluation metrics for “success” are underdeveloped; we need outcome-oriented and context-sensitive measures. Technical literature emphasizes accuracy; social-science literature calls for impact measures (reduction in belief, decreased spread, policy uptake). Standardized metrics that measure social impact, equity, and rights adherence are largely absent. Innovation should include evaluation frameworks that combine technical robustness with social outcomes in local contexts.
- 3.15. Building local research and policy capacity—digital sovereignty—is itself an innovation. Literature on digital colonialism warns that solutions developed in the Global North risk perpetuating dependency (Couldry & Mejias, 2019; Kwet, 2019). Investing in local research hubs, African datasets, policy labs, and judicial expertise is a strategic innovation that yields contextually appropriate tools, reduces external dependence, and strengthens resilience.

4. Discussion

Technical detectors must be evaluated *in situ*. For Cameroon, a detector trained on high-quality datasets may fail on WhatsApp-compressed videos shot on low-end phones. Policy and funders should support field evaluations, dataset creation that reflects local devices/conditions, and transparent reporting of generalization gaps. Encouraging reproducible evaluation benchmarks that include regional data will increase practical utility. On provenance adoption and ethical trade-offs, provenance systems are promising but not plug-and-play. Device manufacturers, platforms, and civil-society stakeholders must co-design provenance mechanisms with privacy-preserving options (e.g., selective disclosure, cryptographic commitments). In Cameroon, advocacy is needed to ensure provenance tools do not become instruments of state surveillance.

On explainable hybrid workflows, design priorities should favour explainability and stepwise evidence presentation tailored for journalists and civil-society actors. Training modules should accompany tool deployment. Donor programmes should fund “last-mile” integration—tools that fit into newsroom workflows

and local languages. On consistent platform governance, platforms must invest in native-language moderation for Cameroon's bilingual context and develop transparent policies about synthetic media. Policy innovation may push for independent audits, content notice regimes, and appeals mechanisms that respect free expression while mitigating harm. On legal safeguards, lawmakers should craft narrowly tailored laws criminalizing malicious impersonation and coordinated election interference, while embedding judicial oversight and civil-society review to prevent abuse. Capacity building for judges, prosecutors, and human-rights defenders is crucial to ensure rights-protective implementation. Media literacy should invest in community-based campaigns that deliver high ROI. Local NGOs, religious leaders, and youth networks can be partners. Materials must be bilingual and culturally anchored; role-play, radio programming, and interactive WhatsApp bots can scale training.

Contextualized responses to targeted campaigns mean that rapid-response fact-checking teams with local knowledge and pre-established trust networks can blunt the impact of targeted disinformation. Mapping local fault lines and anticipating narratives before crises is a preventive innovation. On bias and dataset equity, global funders and researchers must prioritize dataset diversification and involve African researchers in dataset curation. This reduces false positives/negatives and strengthens democratic legitimacy of tools. On countering the liar's dividend, public communication strategies that foreground provenance and multipronged evidence chains—e.g., combining metadata, eyewitness reports, and institutional verification—can reduce the space for delegitimizing authentic evidence. Trusted institutions (universities, national press councils) should be empowered for rapid authentication. On multi-modal detection, local implementers should combine lightweight image/audio forensics with network analytics to balance resource constraints and detection accuracy. Open-source toolkits enabling modular combination are an important innovation.

Improving cross-platform data-sharing and joint investigative protocols between platforms, fact-checkers, and local civil society can enhance attribution capacity. Legal frameworks for lawful cooperation, while protecting privacy, are necessary. On industry-civil-society collaborations, sustaining these collaborations requires funding models beyond short-term grants. Establishing long-term partnerships and capacity transfer is an innovation priority. Platforms should fund local verification hubs as part of corporate responsibility. On low-resource constraints, designing tools that operate offline or tolerate heavy compression, and that preserve source anonymity, opens access. Innovations such as small-footprint forensic apps and community verification protocols are pragmatic and scalable in Cameroon. Finally, on evaluation metrics, research funders should require outcome-oriented evaluations (e.g., impact on spread, belief change, resilience) and not accept only algorithmic accuracy. Mixed-methods studies that combine technical measures with social impact assessment will be informative. Local capacity and digital sovereignty should imply that long-term resilience comes from local capacity building: investing in African datasets, research labs, and policy capacity is innovative because it shifts epistemic authority and reduces dependency on external tech actors. Cameroon should prioritize university-based labs, regionally networked research centers, and public-private partnerships that retain local control over data and tools.

5. Conclusion

The literature indicates that technical detection advances are necessary but insufficient; framing “innovation” narrowly as algorithmic novelty is misleading and risks perpetuating technological solutionism. A holistic definition of innovation should include: robust, interpretable detection methods validated in local contexts; provenance and cryptographic standards developed with privacy and safety safeguards; platform governance reforms that are transparent and linguistically inclusive; narrowly tailored, rights-respecting legal mechanisms; and socio-community innovations—media literacy, local verification networks, and

investments in regional research capacity (Wardle & Derakhshan, 2017; Couldry & Mejias, 2019; Chesney & Citron, 2019).

Policy and practice recommendations include:

1. Support mixed-method, context-aware evaluation of detection tools; fund datasets that represent African devices and languages.
2. Promote adoption of interoperable provenance standards with privacy-preserving controls and safeguards for sources in repressive contexts.
3. Require platform transparency (policies, enforcement metrics), independent audits, and native-language moderation capacity.
4. Develop narrowly defined legal instruments against malicious synthetic-media harms, with human-rights safeguards and judicial training.
5. Fund and scale community-based media literacy and verification networks, with bilingual (French/English) resources tailored for Cameroon.
6. Invest in local research centers, datasets, and training to advance digital sovereignty and reduce dependence on Global North solutions.
7. Adopt outcome-oriented evaluation metrics linking technical interventions to real-world reductions in harm.

The research agenda should incorporate:

- Robust cross-domain detection research emphasizing generalization to compressed, low-quality, and regionally varied media.
- Studies on provenance adoption trade-offs and privacy protections in repressive contexts.
- Measurement research to operationalize social-impact metrics for interventions.
- Ethnographic and computational studies of disinformation narratives in Cameroon's Anglophone/Francophone contexts.

This review synthesizes interdisciplinary evidence to argue for a broadened, socio-technical understanding of innovation in the fight against deepfakes and disinformation. Confidence in high-level patterns (need for socio-technical approaches, platform governance, local capacity) is high. Confidence in specific claims about state of the art detection performance is moderate because of rapid technical change and dataset specificity. Recommendations prioritize rights-respecting, context-aware, and locally empowering innovations that will be most effective in Cameroon and comparable settings.

References

- [1] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: a Compact Facial Video Forgery Detection Network. *Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS)*.
- [2] Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236.
- [3] Chesney, R., & Citron, D. K. (2019). *Deepfakes and the New Disinformation War*.
- [4] Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.

- [5] Council of Europe (Derakhshan, H., & Wardle, C.). (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.
- [6] Farid, H. (2019). *Photo Forensics and the Detection of Manipulated Media*.
- [7] Freedom House. (2023). *Freedom on the Net: Cameroon*. Freedom House country report.
- [8] Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- [9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*.
- [10] Gorwa, R. (2019). The Platform Governance Research Agenda: A Critical Review. *Information, Communication & Society*.
- [11] Howard, P. N., Woolley, S., & Calo, R. (2018). *Computational Propaganda: Political Parties, Politicians, and Publics*.
- [12] Kwet, M. (2019). *Digital Colonialism and the New Imperialism*.
- [13] Kumar, S., & Shah, N. (2018). False Information on Web and Social Media: A Survey. *ACM Computing Surveys / arXiv review*.
- [14] Li, Y., & Lyu, S. (2020). Face X-Ray: A Simple Baseline for Detecting Face Manipulated Images.
- [15] Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online*. Data & Society Research Institute.
- [16] Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*.
- [17] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- [18] UNESCO. (2018). *Journalism, 'Fake News' & Disinformation: A Handbook for Journalism Education and Training*. UNESCO.
- [19] Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.
- [20] Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.
- [21] Selected reports on misinformation in Africa. Useful regional source for empirical mapping of disinformation narratives.
- [22] Africa Check.
- [23] CIPESA (Collaboration on International ICT Policy for East and Southern Africa). (Reports on digital misinformation and platform use in Africa.)
- [24] Data & Society. (2017). Reports on media manipulation and verification practices. (Marwick & Lewis and related outputs.)
- [25] InVID / WeVerify project outputs and tool documentation. (Practical verification tools and research.)
- [26] Platform transparency and policy reports (selected): platform-provided transparency reports and independent audits (various sources used to analyze platform governance).