Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 11, pp 83-89, November, 2026

The Ethical Challenges of AI Generated Political Content in Digital Media

Zhang Hanyu[1*], Mohd Haizra Hashim[1]

[1] INTI International University

**Corresponding author: Zhang Hanyu; email: I26038797@student.newinti.edu.my*

Second author email: mhaizra.hashim@newinti.edu.my

ABSTRACT :

AI-generated political content can support accessible communication while creating new opportunities for deception and misleading representation. This article examines how ethical responsibility should be allocated when audience detection and disclosure provide incomplete protection. It develops a conceptual ethical analysis informed by selected literature on political deepfakes, public trust and synthetic campaigning, using deontological ethics, utilitarianism and social responsibility theory. The discussion distinguishes the credibility of individual messages from broader uncertainty about authentic evidence and separates experimental findings from normative arguments. It also considers why clearly labelled content can still harm the people represented and why technical capacity for personalised persuasion does not establish greater effectiveness. The article argues that responsibilities should reflect control over production and distribution, capacity to prevent harm, and the ability to provide correction. Campaigns, platforms and journalists therefore carry different but connected duties, while audiences retain a reasonable role in checking sources. Transparency should accompany verification, proportionate intervention and identifiable human accountability. The analysis does not establish the prevalence or electoral effects of synthetic media. It offers a practical basis for assessing political AI use while protecting public access to reliable information and legitimate political expression.

Keywords: AI-generated political content; media ethics; political communication; deepfakes; public trust; accountability

1 Introduction

Generative artificial intelligence allows political actors, media organisations and ordinary users to produce or modify text, images, audio and video. In political communication, these tools can support translation, accessibility and the explanation of policy. They can also create realistic material that falsely attributes a statement or action to a person. The ethical difficulty is that the same production tools can serve both legitimate expression and deception. A judgement about a particular use therefore requires attention to what audiences are likely to understand, whose interests are affected and who controls the communication.

This article focuses primarily on synthetic audiovisual political content, including political deepfakes and AI-generated campaign images. A deepfake can make a person appear to say or do something that did not occur. AI-generated political images may also imply that a candidate attended an event or received support that never existed. Text generation and personalised messaging form part of the wider context, but most of the audience evidence discussed here concerns audiovisual material. This distinction matters because findings about a particular video cannot automatically describe every use of AI in campaigning.

Existing research provides reasons for concern without supporting the assumption that deepfakes are always more persuasive than other misinformation. Barari et al. (2025) found political deepfakes to be similarly credible to other forms of fabricated media. In a particular experimental setting involving a scandal associated with participants' own party, authentic political videos could also be mistaken for deepfakes. These findings concern credibility judgements, rather than observed election results. They nevertheless direct ethical attention towards two problems: false information may be accepted, and genuine evidence may become easier to doubt.

The interests at stake extend beyond factual accuracy. De Ruiter (2021) argues that the use of a person's image or voice can be morally problematic when it represents that person in ways to which they would object. Disclosure may reduce deception without fully resolving this concern. Pawelec and Łabuz (2026) similarly question whether describing some campaign applications as "softfakes" understates their risks. An idealised candidate image or artificial testimonial can influence political representation even when it is less technically elaborate than an impersonation video.

The central question is therefore: how should responsibility for AI-generated political content be allocated when disclosure and audience detection provide incomplete safeguards against deception and distrust? The article argues that responsibility should reflect control over production and circulation, capacity to prevent foreseeable harm, and ability to offer correction. This approach connects the duties of campaigns, platforms, journalists and audiences without assuming that they possess equal information or resources. Its purpose is to develop an ethical argument for responsible communication, rather than to measure the scale or electoral effects of political AI use.

2 Review Approach and Analytical Framework

2.1 Approach and scope

This article presents a conceptual ethical analysis informed by selected literature. Sources were purposively retained for their direct relevance to deception and audience autonomy, authenticity and public trust, disclosure and detection, or institutional responsibility; foundational ethical texts supplied the normative framework. Bibliographic details and cited claims were checked using official journal or publisher records, accessible abstracts and relevant full-text passages. This process did not involve an exhaustive database search or reading every publication in full. The article reports no original survey, experiment or content analysis and makes no claim to a systematic or comprehensive review.

The analysis distinguishes different forms of evidence. Experiments test responses under specified conditions, while analyses of online comments illustrate interpretations expressed by particular users. Normative articles offer arguments about what should be permitted or required. These sources cannot be treated as interchangeable proof of political harm. The assessment connects their reported findings and arguments with questions about duties, consequences and institutional responsibility. Its limitations include selective coverage and the contexts and measures of the cited studies. It does not estimate the prevalence or overall effect of synthetic political content or reanalyse the underlying data.

2.2 Three ethical perspectives

A deontological perspective evaluates duties and respect for persons, rather than judging an action solely by the advantage it produces. Kant's (1785/2012) account of duties and persons as ends provides a foundation for asking whether citizens are treated as persons capable of judgement or merely as instruments for a campaign's success. In this article, that question supports duties against deliberate false attribution and deceptive impersonation. It also requires attention to the person represented. A campaign cannot answer every objection to a fabricated statement simply by arguing that the content helped it reach more voters.

Utilitarianism evaluates actions through their consequences for those affected. Mill's (1864/2014) utilitarian account informs the comparison of communication benefits with foreseeable harms. Accessible translation or lower production costs can offer public benefits, but a campaign's private electoral advantage should not be counted as if it were automatically a benefit to everyone. Relevant harms include reputational injury, confusion about evidence and the cost of correction. Their likelihood, distribution and uncertainty matter: a small production benefit may be obtained by imposing substantial risks on people who did not agree to bear them.

Social responsibility theory connects media freedom with obligations towards the public. Peterson's account, presented in Siebert et al. (1956), provides the institutional starting point for examining access to reliable information, verification and accountability. Applying this perspective to platforms is an extension of the present analysis, rather than a claim that a historical press theory directly anticipated contemporary recommendation systems. The central question is how institutions exercising influence over public communication should answer for that influence while protecting legitimate expression.

These perspectives are complementary without being identical. A consequence-based argument may favour a useful translation, while a duty-based assessment still requires accurate attribution and respect for the speaker. Institutional responsibility then identifies who can verify the translation, explain its production and correct an error. The following analysis uses these questions together. Ethical evaluation should identify both a defensible action and an actor capable of carrying it out.

3 Ethical Tensions in Synthetic Political Communication

3.1 Deception and audience autonomy

Political persuasion is not inherently unethical. Campaigns normally seek to influence citizens through arguments, images and emotional appeals. The ethical boundary becomes more difficult when synthetic media hides the basis on which an audience is being asked to judge. A fabricated endorsement presented as a genuine recording gives viewers a false reason to support a candidate. The problem involves both the truth of the message and the audience's opportunity to understand its source.

Deontological reasoning objects to this practice because deliberate misrepresentation undermines informed judgement and attributes speech without a proper basis. De Ruiter (2021) adds that the represented person's interests cannot be reduced to whether an audience was deceived. A label may inform viewers that a voice is synthetic, yet the use could still misrepresent the speaker's identity or wishes. Consent, intention and the nature of the representation must therefore be considered alongside disclosure.

Empirical evidence needs equally careful interpretation. Momeni's (2025) analysis of Instagram comments on a political deepfake illustrates how some users responded to a fabricated endorsement as if it were authentic. These self-selected comments do not establish representative public opinion or a causal effect on voting. Kelani and Lepadatu (2026) examine deepfake propaganda and associated comments during the 2024 United States presidential election. Their content analysis identifies polarised narratives but does not measure whether exposure caused polarisation. Such studies help identify ethical concerns without settling the size of their electoral consequences.

Personalisation raises a related issue. Generative systems can produce messages for different audiences, but technical capacity does not demonstrate unusually powerful persuasion. In two preregistered experiments, Argyle et al. (2025) found attitude changes in response to AI-generated persuasive messages. However, microtargeted customisation and interactive elaboration were not clearly more persuasive than a generic message. This finding challenges the assumption that increasingly tailored content must necessarily produce stronger influence. The ethical concern should focus on the character of the communication, the use of personal information and opportunities for informed judgement, rather than on unproven claims of irresistible persuasion.

Synthetic campaign images also require attention to implication. Pawelec and Łabuz (2026) show why the language of softfakes can minimise serious concerns. An image of fictional supporters may communicate that real people endorse a candidate, even without a written false statement. Ethical assessment should ask what the presentation reasonably invites an audience to believe. Content can communicate a factual impression visually, and a production label does not establish whether that impression is accurate.

Consider a clearly hypothetical comparison. A campaign uses a synthetic voice to translate a candidate's approved speech. Accurate translation with a clear explanation can improve access to political information. By contrast, a synthetic voice used to invent an opponent's statement creates false attribution. The first case still requires checking accuracy and the candidate's agreement; the second raises a direct objection to deception. The example shows how duties constrain the pursuit of communication benefits, while institutional responsibility identifies the need for human review before publication.

3.2 Authenticity and public trust

The consequences of synthetic media extend beyond whether a single false message is believed. Pawelec (2022) examines threats to democratic functions through disinformation, hate speech and damage to the conditions of public communication. The ethical concern is that citizens need some reliable means of deciding which claims and records deserve confidence. Where evidence can be dismissed merely because manipulation is possible, accountability becomes more difficult.

Barari et al. (2025) provide an important distinction between fabricated content appearing credible and authentic content being misclassified. Their finding concerning scandal videos associated with participants' own party should not be generalised into a claim that all audiences routinely reject genuine evidence. It nevertheless illustrates a vulnerability in the evaluation of politically inconvenient information. The possibility of a deepfake may supply a reason for doubt without supplying evidence that a particular recording is false.

Credibility and reputational harm are also distinct. Hameleers et al. (2024a) compare different degrees of deception in political deepfakes in the United States and the Netherlands. Subtle distortions could be more credible, while implausible fabrications could still delegitimise political actors. These results suggest that disbelief does not necessarily remove every negative consequence. An audience may reject the apparent statement yet retain a negative impression or uncertainty about the person depicted. Ethical analysis should therefore consider representation and public understanding in addition to the binary question of belief.

This does not justify predictions of inevitable democratic collapse. Simon et al. (2023) argue that fears about generative AI and misinformation are overstated, drawing principally on evidence from wealthy democracies. Their commentary provides a reason to scrutinise broad claims, rather than an experimental demonstration that all risks are small. A utilitarian assessment should compare expected benefits and harms using evidence appropriate to the context. Possibility alone is insufficient to establish widespread damage, but uncertainty does not excuse deliberate deception or remove a duty to correct foreseeable errors.

Institutional responses can themselves affect trust. If journalists or platforms describe every disputed recording as potentially artificial without explaining the evidence, they may encourage suspicion rather than understanding. A responsible response should make a limited claim: what has been verified, what remains uncertain and why. Protecting public trust does not require telling audiences to accept all official sources. It requires providing defensible reasons for confidence or doubt and correcting mistakes when those reasons change.

3.3 Disclosure and the limits of detection

Detection research offers a mixed picture. Appel and Prietzel (2022) report three preregistered studies in which analytical thinking was associated with successful identification in two studies, while political interest was associated with identification in one. Groh et al. (2024) conducted five preregistered experiments involving 2,215 participants and compared transcripts, audio and video. Audiovisual information could assist detection, but high-quality text-to-speech material was more difficult to identify than recordings using voice actors. These findings support neither complete confidence in individual detection nor the claim that people can never recognise synthetic media.

Audiences also use context. Hameleers et al. (2024b) found that conflict with existing understandings of reality could motivate doubt more than visible technical flaws. Wack and Prochaska's (2026) experiment in Kenya further shows the importance of comment context and partisan identification for audience interpretation. These studies caution against a universal checklist based only on unnatural faces or voices. As production quality and viewing conditions vary, technical cues may be insufficient or unavailable.

Disclosure can help, but different interventions should not be grouped together. Lu and Yuan (2024) studied AI disclaimers and comprehension in political deepfake parodies through two United States experiments. Their results concern particular satirical content and a climate-related issue, rather than proving that every mandatory labelling policy will work. By contrast, Ternovski et al. (2022) examined general warnings about deepfakes and found increased suspicion of genuine videos. Tulin et al. (2025) tested an explanatory video about generative AI among Canadian adults; greater suspicion did not necessarily improve discrimination between true and false news. The latter study concerns an educational intervention, not labels attached to individual posts.

These findings create a practical ethical tension. People should know when realistic political media has been generated or substantially altered, yet broad warnings can contribute to misplaced doubt. Chomanski and Lauwaert (2025) argue against legally requiring AI-generated political content to be labelled. Their position is a normative argument about the likely value of such laws, not experimental proof that all disclosure is ineffective. The present analysis supports context-specific ethical disclosure while recognising that production information cannot by itself establish a political claim's truth.

A useful explanation should identify the substantial alteration, the responsible source and whether the content is intended as a factual record, illustration or satire. It should remain understandable when the material is reshared where possible. Disclosure also has limits: a visible label does not authorise a harmful impersonation or correct a false factual implication. Peters et al. (2025) find differences in attitudes towards deepfakes according to their use and target. This supports attention to context, but does not establish that all political satire is harmless or legally permissible.

4 Shared Responsibilities and Practical Safeguards

4.1 Political actors and campaign organisations

Campaigns have substantial responsibility where they commission, approve and release synthetic content. They can prevent a misleading message before it reaches an audience, and they possess information about its intended meaning and production. Ethical review should therefore occur before publication. A named person should check factual claims, attributed speech, permissions and likely audience interpretation, then remain answerable for corrections. Retaining source material and an approval record would support that responsibility.

Jungherr et al. (2026) found strong disapproval of deceptive campaign uses of AI without a significant corresponding decline in party favourability in their United States studies. This does not show that deception never carries an electoral cost. It does suggest that expressed public disapproval cannot be assumed to provide an effective safeguard in every setting. Campaign standards should consequently prohibit deceptive impersonation and fabricated endorsements rather than depending solely on the possibility of voter punishment.

From a consequential perspective, low production cost is not a sufficient defence when an avoidable fabrication shifts the burden of verification onto citizens and opponents. From a duty-based perspective, accuracy and non-deception remain relevant even when a message proves popular. Social responsibility connects both concerns to a practical requirement: a campaign should be able to explain who approved the material and how an error will be corrected. These are ethical proposals, rather than claims about an existing universal legal obligation.

4.2 Digital platforms

Platforms influence the visibility and circulation of content through recommendation, advertising and moderation. Their responsibility follows from that control, although they do not originate every misleading post. They should provide clear ways to report realistic impersonation and disputed attribution, preserve relevant context where possible, and explain significant moderation decisions. A correction or warning is more useful when it addresses the underlying claim than when it merely announces that AI might have been used.

Intervention should be proportionate to the content and likely harm. Fabricated election instructions and deceptive impersonation justify a stronger response than a clearly contextualised political illustration. Satire and criticism require room to operate, including when they are uncomfortable for

powerful people. At the same time, a claim of satire should not automatically end assessment of false attribution or foreseeable harm. Platform policies need both an identifiable basis for intervention and a route to challenge mistakes.

No single label or detection system can provide complete protection. The ethical requirement is therefore an accountable process combining evidence, context, reporting and correction. Where a platform promotes or sells distribution of political content, it should consider its own contribution to reach rather than transferring all responsibility to the uploader. This follows the article's principle that greater control creates a greater opportunity, and corresponding responsibility, to reduce harm.

4.3 Journalists and media organisations

Journalists have a different role: they can make synthetic material more influential by treating it as news. Verification should precede amplification, particularly where a recording attributes a damaging statement to an identifiable person. Relevant checks include locating the original source, seeking independent confirmation and distinguishing the recording's authenticity from the truth of the associated claim. Technical analysis can assist this process without replacing editorial judgement.

Gutiérrez-Caneda et al. (2024) identify ethical concerns through qualitative interviews with professionals and researchers, including transparency and changes to journalistic roles. These findings describe the perspectives studied rather than all journalists. Raemy et al. (2025) examine deepfakes in relation to journalism's normative commitments. Together, these sources support treating verification and accountability as professional questions rather than merely technical tasks.

Reporting should also avoid making an unsupported allegation more memorable than its correction. A newsroom can explain why disputed material matters without reproducing it uncritically or implying certainty that its evidence does not support. If an initial report proves wrong, the correction should reach the audience through the same relevant channels where feasible. Protecting trust depends partly on explaining and repairing errors, rather than projecting an impossible claim of perfect detection.

4.4 Audiences and education

Audiences retain a reasonable duty to check sources before sharing consequential political claims. However, this duty should reflect the information available to an ordinary user. Ahmed (2023) reports third-person perceptions in which people consider others more vulnerable to deepfake influence and themselves better able to detect it. Such perceived competence is not a measure of actual ability. Education should therefore encourage careful evaluation without promising that confidence or familiarity guarantees accuracy.

Source comparison, contextual checking and an understanding of different forms of manipulation offer a broader basis for judgement than searching for visual mistakes alone. The results of Tulin et al. (2025) also suggest the need to evaluate whether education improves discrimination, rather than merely increasing suspicion. This does not mean that media literacy is futile. It means that institutions should assess what an intervention actually achieves and avoid treating public vigilance as a substitute for their own responsibilities.

5 Discussion and Limitations

The proposed allocation of responsibility addresses a weakness in recommendations that ask everyone to be more transparent or more sceptical without specifying who can act. Campaigns control production decisions; platforms affect reach and context; journalists influence verification and interpretation; audiences decide whether to share. The responsibilities overlap, but they are not equal. An actor's duty should reflect its contribution to a risk, knowledge of that risk and practical capacity to prevent or remedy it. This approach also avoids treating downstream audiences as the final defence against every failure earlier in the communication process.

The three ethical perspectives reveal why a single safeguard is insufficient. Disclosure can respect informed judgement while leaving false attribution or reputational harm unresolved. Removal can reduce circulation while restricting legitimate expression. Educational warnings can raise awareness while encouraging unwarranted doubt. A defensible response therefore combines duties against deception, a context-sensitive assessment of consequences and institutional arrangements for explanation and correction. Where the perspectives conflict, the response should identify the affected interests and explain why a less restrictive alternative is insufficient.

The argument also qualifies the claim that only human use, rather than technology itself, creates the ethical problem. Particular uses remain central, but design, ease of reproduction and distribution arrangements influence what harms are foreseeable and who can prevent them. Langguth et al. (2021) discuss image manipulation as both a technical and social problem. Ethical evaluation should similarly examine the communication process rather than isolating the creator's intention. Benign intent or accurate labelling cannot settle every question about the resulting representation.

Several limitations remain. This is a selective conceptual analysis and does not estimate how common deceptive political AI use is. Credibility, detection, attitudes and party favourability are different outcomes; none can be treated alone as a measure of electoral impact. Evidence from the United States, Europe, Canada and Kenya also reflects particular political and media environments. The international perspective adopted here does not

establish how Malaysian audiences would respond. Differences in language, trust and circulation practices require contextual investigation before empirical conclusions are transferred.

Further research could test whether combined source information, contextual explanation and correction improve understanding without increasing general distrust. It could also examine how responsibilities change when content moves between campaigns, news organisations and platforms or loses its original disclosure. Such research would inform the application of the proposed ethical principles. Until then, uncertainty should encourage careful claims and proportionate safeguards, alongside a clear refusal to present fabricated political evidence as authentic.

6 Conclusion

AI-generated political content presents ethical challenges concerning deception, representation, audience autonomy and the reliability of public evidence. Existing research does not support treating every deepfake as exceptionally persuasive, and particular findings should not be converted into claims about inevitable election effects. The ethical concern nevertheless extends beyond immediate belief: synthetic media can harm the people represented and create uncertainty about authentic information.

Responsibility should therefore follow control, capacity and the ability to provide remedy. Campaigns should prevent deceptive production, platforms should address their role in circulation, journalists should verify before amplification and audiences should exercise reasonable care. Deontological ethics, utilitarianism and social responsibility theory together show why disclosure, verification and accountability must operate as connected safeguards. Responsible political AI use requires both protection against misleading representation and respect for legitimate expression. Its justification depends on how the content is produced, understood, distributed and corrected, with identifiable people and institutions answerable for those decisions.

Generative AI Assistance

AI tools assisted in preparing the initial draft. OpenAI Codex subsequently supported manuscript restructuring, development of the ethical analysis, language revision, and checks of bibliographic records and cited claims against available journal and publisher material. No original research data were generated or analysed.

REFERENCES

- Ahmed, S. (2023). Examining public perception and cognitive biases in the presumed influence of deepfakes threat: Empirical evidence of third person perception from three studies. *Asian Journal of Communication, 33*(3), 308–331. https://doi.org/10.1080/01292986.2023.2194886
- Appel, M., & Prietzel, F. (2022). The detection of political deepfakes. *Journal of Computer-Mediated Communication, 27*(4), Article zmac008. https://doi.org/10.1093/jcmc/zmac008
- Argyle, L. P., Busby, E. C., Gubler, J. R., Lyman, A., Olcott, J., Pond, J., & Wingate, D. (2025). Testing theories of political persuasion using AI. *Proceedings of the National Academy of Sciences, 122*(18), Article e2412815122. https://doi.org/10.1073/pnas.2412815122
- Barari, S., Lucas, C., & Munger, K. (2025). Political deepfakes are as credible as other fake media and (sometimes) real media. *The Journal of Politics, 87*(2), 510–526. https://doi.org/10.1086/732990
- Chomanski, B., & Lauwaert, L. (2025). Automated propaganda: Labeling AI-generated political content should not be required by law. *Journal of Applied Philosophy, 42*(3), 994–1015. https://doi.org/10.1111/japp.70002
- de Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology, 34*(4), 1311–1332. https://doi.org/10.1007/s13347-021-00459-2
- Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. (2024). Human detection of political speech deepfakes across transcripts, audio, and video. *Nature Communications, 15*, Article 7629. https://doi.org/10.1038/s41467-024-51998-z
- Gutiérrez-Caneda, B., Lindén, C.-G., & Vázquez-Herrero, J. (2024). Ethics and journalistic challenges in the age of artificial intelligence: Talking with professionals and experts. *Frontiers in Communication, 9*, Article 1465178. https://doi.org/10.3389/fcomm.2024.1465178
- Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2024a). Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. *Computers in Human Behavior, 152*, Article 108096. https://doi.org/10.1016/j.chb.2023.108096

- Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2024b). They would never say anything like this! Reasons to doubt political deepfakes. *European Journal of Communication, 39*(1), 56–70. https://doi.org/10.1177/02673231231184703
- Jungherr, A., Rauchfleisch, A., & Wuttke, A. (2026). Artificial intelligence in election campaigns: Perceptions, penalties, and implications. *Political Communication, 43*(4), 585–606. https://doi.org/10.1080/10584609.2025.2611913
- Kant, I. (2012). *Groundwork of the metaphysics of morals* (M. Gregor & J. Timmermann, Eds. & Trans.; Rev. ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511919978 (Original work published 1785)
- Kelani, Z. A., & Lepadatu, D. (2026). Generative artificial intelligence, misinformation and political polarization: A content analysis of deepfake propaganda during the 2024 US presidential election. *Journal of Information Technology & Politics*. Advance online publication. https://doi.org/10.1080/19331681.2026.2652897
- Langguth, J., Pogorelov, K., Brenner, S., Filkuková, P., & Schroeder, D. T. (2021). Don't trust your eyes: Image manipulation in the age of deepfakes. *Frontiers in Communication, 6*, Article 632317. https://doi.org/10.3389/fcomm.2021.632317
- Lu, H., & Yuan, S. (2024). "I know it's a deepfake": The role of AI disclaimers and comprehension in the processing of deepfake parodies. *Journal of Communication, 74*(5), 359–373. https://doi.org/10.1093/joc/jqae022
- Mill, J. S. (2014). *Utilitarianism* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139923927 (Original second edition published 1864)
- Momeni, M. (2025). Artificial intelligence and political deepfakes: Shaping citizen perceptions through misinformation. *Journal of Creative Communications, 20*(1), 41–56. https://doi.org/10.1177/09732586241277335
- Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society, 1*(2), Article 19. https://doi.org/10.1007/s44206-022-00010-6
- Pawelec, M., & Łabuz, M. (2026). Are softfakes really "soft"? Semantic, ethical, and legal perspectives on deepfakes and AI-generated visualizations in political campaigning. *International Journal of Politics, Culture, and Society*. Advance online publication. https://doi.org/10.1007/s10767-026-09579-z
- Peters, C. S., McGinnis, A. J., Im, S., & Hennessey, H. L. (2025). Satire versus illicit: Policy issues regarding the creation of deepfakes using AI technology. *Behavioral Sciences & the Law, 43*(4), 462–471. https://doi.org/10.1002/bsl.2726
- Raemy, P., Bendahan Bitton, D., Ötting, H., Hoffmann, C. P., Godulla, A., & Puppis, M. (2025). Deepfakes and journalism: Normative considerations and implications. *Journalism Studies, 26*(14), 1724–1744. https://doi.org/10.1080/1461670X.2025.2547300
- Siebert, F. S., Peterson, T., & Schramm, W. (1956). *Four theories of the press: The authoritarian, libertarian, social responsibility, and Soviet communist concepts of what the press should be and do*. University of Illinois Press.
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-127
- Ternovski, J., Kalla, J., & Aronow, P. M. (2022). The negative consequences of informing voters about deepfakes: Evidence from two survey experiments. *Journal of Online Trust and Safety, 1*(2). https://doi.org/10.54501/jots.v1i2.28
- Tulin, M., Pantazi, M., Starke, C., Sivolap, M., & Dobber, T. (2025). Echoes of doubt: Exposure to information about generative AI decreases believability of news. *International Journal of Communication, 19*, 3712–3735. https://ijoc.org/index.php/ijoc/article/view/24772
- Wack, M., & Prochaska, S. (2026). Making sense of AI-generated disinformation: How audience interpretations influence the impact of deepfakes in Kenya. *Social Media + Society, 12*(3), Article 20563051261462092. https://doi.org/10.1177/20563051261462092