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

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 11, pp 107-110, November, 2026

Political Deepfakes and the Burden of Verification in Malaysia

***Wang Xunzhi*[1*], *Mohd Haizra Hashim*[1]**

[1] *INTI International University*
**Correspondence to Wang Xunzhi at i26039083@student.newinti.edu.my*
Mohd Haizra Hashim, mhaizra.hashim@newinti.edu.my

ABSTRACT :

A convincing political forgery can be taken as evidence, and the possibility of forgery can make a genuine recording easier to dispute. This article examines who should bear the work of verifying disputed political media in Malaysia. It draws on published studies of deepfake detection, public attitudes and video campaigning, then applies deontological ethics, consequentialism and virtue ethics to an illustrative case. A viewer has two questions to resolve. Is the recording authentic, and does it support the claim attached to it? A label can tell viewers how material was made, but it cannot answer both questions. Campaigns are best placed to disclose alterations and retain source files. Platforms and newsrooms can keep that information visible, investigate disputed claims and correct errors in the languages in which those claims circulate. These duties should reflect each actor's access to evidence and leave room for legitimate satire and criticism.

Keywords: political deepfakes, media ethics, verification, political communication, Malaysia

1 Introduction

Suppose a short recording circulates during a Malaysian election campaign. A candidate appears to say that a district will lose funding if it votes for a rival. The post gives no source, and the candidate's office says the voice was cloned. Both the accusation and the denial depend on facts that a listener cannot settle from the recording alone. Even if the voice is genuine, the candidate may have been quoting someone else. Establishing who spoke and establishing what the remarks mean are separate tasks.

Video already plays a prominent role in Malaysian campaigns. In his account of the 2022 general election, Tapsell (2023) describes how campaign teams used livestreams and short videos, with TikTok becoming an important channel. Video gave campaigns a way to reach viewers directly and also carried polarising messages about race and religion. That campaign provides a setting for considering how a disputed recording might travel through existing channels.

This article uses the term authenticity burden for the work of checking a recording's origin and establishing the context in which it should be interpreted. A person scrolling through a feed may only have a clipped version and a caption. Campaign teams may have source recordings, platforms may have earlier uploads, and reporters may be able to contact witnesses. The central ethical question is how to distribute the work of verification when access to those resources is uneven.

2 Scope and Approach

This article uses a narrative, literature-based approach. Its sources address digital campaigning in Malaysia, the detection of synthetic political speech, attitudes towards deepfakes and responses to disputed evidence. The discussion applies deontological ethics, consequentialism and virtue ethics to the same hypothetical recording so that the obligations, possible effects and professional judgements can be compared. The case is illustrative and the analysis draws on published work.

The focus is on audio and video that a viewer might take for a record of something a political figure actually said or did. Generative tools can also help translate a speech or illustrate a policy. A synthetic translation becomes ethically relevant here if, for example, it makes a candidate seem to have spoken words in a language they did not use, or alters the substance of the original statement.

Intent matters in deciding who is responsible. Wardle and Derakhshan (2017) distinguish misinformation, disinformation and malinformation. A supporter who forwards a forgery in good faith spreads misinformation. A campaign that commissions it to harm a rival engages in disinformation.

Releasing genuine but private material to harm someone illustrates malinformation. Those distinctions help separate the conduct of the person who made or commissioned a clip from the conduct of someone who encountered it downstream.

3 What the Evidence Shows

3.1 Detection and Disputed Recordings

Groh et al. (2024) tested how 2,215 participants judged real and fabricated political speeches across five experiments. Participants generally distinguished real from fabricated speeches more accurately when they had audio or visual cues than when they had text alone. The researchers also found that deepfakes using text-to-speech audio were harder to identify than comparable versions voiced by actors. The type and quality of the material mattered. Judging a set of clips in an experiment is a narrower task than deciding whether to trust or share an unsolicited recording in an election campaign.

Uncertainty is a different outcome from mistaking a forgery for a real recording. In a UK experiment, Vaccari and Chadwick (2020) compared responses to shorter versions of a synthetic political video with responses to a full version that revealed how it had been made. The shorter clips left more viewers unsure about the statement's truth. The result matters for verification because a viewer who reserves judgement may still need a reliable account of what the clip shows.

A claim of manipulation can also serve the person whose conduct is under scrutiny. In survey experiments with US adults, Schiff et al. (2025) found that false claims of misinformation could increase support for politicians in hypothetical responses to scandals. That benefit, often called the liar's dividend, appeared primarily when the underlying evidence was a text report. Video evidence was generally harder to discredit with a denial. In the opening example, the candidate's claim of manipulation deserves investigation.

3.2 Attitudes Towards Deepfakes in Malaysia

Neyazi et al. (2024) surveyed participants in Malaysia, Singapore and India about their attitudes towards deepfakes. Among respondents who regarded AI as a risk to democracy, greater trust in government was associated with more favourable attitudes towards deepfakes across the three countries. The survey measured responses to statements about deepfakes. It did not test whether respondents could authenticate a clip or establish that government trust caused the observed attitudes.

A Sabri (2024) reviews impersonation and fabricated statements as risks for Malaysian politicians. The review draws attention to the risk to an impersonated person's reputation. Verification also serves audiences who need to decide whether a claim deserves attention. A voter may recognise a politician's voice yet have no way to retrieve the full speech or check a translation. In such a case, access to the original material matters as much as familiarity with the person depicted.

4 Ethical Responsibilities for a Disputed Clip

The three approaches place different weight on duties, consequences and character (Hursthouse & Pettigrove, 2023). Applied to the same recording, each helps decide what a campaign or newsroom should do before the evidence is clear.

4.1 The Duty to Represent Evidence Honestly

From a deontological perspective, viewers are entitled to judge a candidate using evidence that has been honestly presented. A campaign breaches that duty when it passes off fabricated speech as a genuine record. A misleading caption can breach the same duty when it changes the meaning of an authentic recording. The obligation extends to the accused campaign, which should give reasons for a claim of fabrication and make corroborating material available when it can do so responsibly.

An authorised synthetic translation poses a different problem. A campaign could use a generated voice to convey an approved translation and clearly identify the dubbing. Listeners should be able to reach the original statement and understand which version they are hearing. The disclosure is owed because the choice of voice may change what listeners reasonably infer about the recording.

4.2 The Consequences of Publishing or Waiting

For a newsroom, sharing the disputed clip could alert voters to a potentially important allegation. It could also lend the accusation credibility at a point when there is too little evidence to judge it. Waiting for full authentication carries a cost if the threat is genuine and an election is near. An interim report could explain that the recording is disputed, say what the newsroom has verified and describe which evidence it is seeking. The newsroom can decide separately whether replaying the audio serves a public purpose while its origin is unresolved.

A warning can reduce harm and still leave part of the allegation with viewers. In experiments using fictional scenarios, Clark and Lewandowsky (2026) found that a specific warning reduced the influence of a deepfake video on judgements of guilt, while some influence persisted. This gives a reason to pair a disclosure with corroborating evidence or a correction that addresses the allegation itself.

4.3 The Habits of a Fair Communicator

Virtue ethics directs attention to how communicators act under pressure. A fair reporter asks for the full recording, gives the candidate a meaningful opportunity to respond and uses the same evidentiary standard for rival parties. Practical judgement also matters when writing a headline. Reporting that a campaign disputes the origin of a recording conveys what has happened. Calling the recording a deepfake requires evidence that settles the question. If new evidence changes the account, a responsible publisher corrects the original story and makes the change visible.

An individual supporter who discovers that a clip they shared was misleading can correct it in the same group where it circulated. A campaign that commissioned the clip has a greater obligation because it approved the material and can trace its origin. If a longer recording contains private information, a newsroom can verify it and explain the relevant context without republishing every detail. Fairness requires enough disclosure for the audience to judge the claim and care with information that does not bear on it.

5 Putting Verification into Practice

5.1 Campaigns and Creators

Campaign teams can reduce the work others must do by keeping the original recording, the approved script and a record of substantial edits. Where a generated scene or voice could be taken as real, a notice should name the alteration in plain language. A brief on-screen notice can stay with a clip when it is reposted. An audio disclosure may help when the sound circulates separately. A paid political post should also identify the sponsor so that a viewer can trace who is responsible for the message.

5.2 Newsrooms and Fact Checkers

A useful verification starts by locating the earliest available upload and requesting a longer version or the source file. The earliest upload is a lead in the investigation. It should be checked against other recordings, reporting from the event or witness accounts. A reporter can ask someone familiar with the language to check disputed words or subtitles. A detector result may help direct these inquiries. If the source cannot be established, a published account should say which part of the claim remains unresolved.

A correction should address what the original post led people to believe. If a Malay-language clip acquired a misleading English caption, correcting the source video alone would leave the error in the version some viewers saw. A newsroom could correct the caption in English and distribute the update through the channels that carried that version. Media literacy educators can use the same kind of example to teach students to trace a clip and compare its caption with the underlying speech.

5.3 Platforms and Appeals

Platforms control which notices accompany posts and how a disputed clip is promoted. They can keep provenance information visible on reposts, provide a way to report impersonation and make it possible to find the sponsor of a political advertisement. When evidence is incomplete, a visible notice or temporary pause in paid promotion can allow time for review. A blanket removal rule could also catch obvious parody or documented criticism.

UNESCO (2023) calls for transparent moderation procedures and meaningful avenues for appeal. Applied here, a platform should explain which rule it used and give the publisher a way to contest an erroneous decision. An authenticity dispute should turn on the evidence available for the specific clip. A complaint from a candidate may trigger a review, but the outcome should depend on the same criteria for all campaigns.

6 Limitations and Future Research

The sources reviewed here support an ethical analysis of how disputed media should be handled. They do not establish how often Malaysian voters encounter political deepfakes, which audiences are most affected or whether exposure changes a vote. The hypothetical case also leaves open how these practices would work under the time and resource constraints of a real campaign.

A local study could compare how audiences check the same disputed clip when given an on-screen disclosure, a link to the full recording or both. Using versions in the languages in which a claim circulates would help researchers test whether an intervention reaches its intended audience. The study could distinguish belief in the allegation from uncertainty about the clip and willingness to share it, since each would call for a different response.

7 Conclusion

Political recordings can be forged, taken out of context or wrongly discredited. A viewer should be able to trace where a recording came from and see how it supports the claim attached to it. Campaigns can make that possible by preserving sources and disclosing material changes. Newsrooms and platforms can make verification and corrections easier to follow. Assigning those responsibilities to the people and institutions with access to the relevant evidence helps voters assess a disputed recording without having to resolve its origin on their own.

REFERENCES :

- A Sabri, F. A. (2024). Narrative review on detonation of AI-generated deep fakes and misinformation: A threat to politicians in Malaysia. *e-Journal of Media and Society, 7*(2), 71–88. https://ir.uitm.edu.my/id/eprint/97612/
- Clark, S., & Lewandowsky, S. (2026). The continued influence of AI-generated deepfake videos despite transparency warnings. *Communications Psychology, 4*, Article 13. https://doi.org/10.1038/s44271-025-00381-9
- 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
- Hursthouse, R., & Pettigrove, G. (2023). Virtue ethics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Fall 2023 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/
- Neyazi, T. A., Nadaf, A. H., Tan, K. E., & Schroeder, R. (2024). Does trust in government moderate the perception towards deepfakes? Comparative perspectives from Asia on the risks of AI and misinformation for democracy. *Government Information Quarterly, 41*(4), Article 101980. https://doi.org/10.1016/j.giq.2024.101980
- Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025). The liar's dividend: Can politicians claim misinformation to evade accountability? *American Political Science Review, 119*(1), 71–90. https://doi.org/10.1017/S0003055423001454
- Tapsell, R. (2023). Social media and Malaysia's 2022 election: The growth and impact of video campaigning. *Pacific Affairs, 96*(2), 303–322. https://doi.org/10.5509/2023962303
- 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
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* (Council of Europe Report No. DGI(2017)09). Council of Europe. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html