

Ethical Governance of AI-Generated Political Content on TikTok and YouTube

Han Yajie^{1*}, Mohd Haizra Hashim²

¹Faculty of Education and Liberal Arts, INTI International University, Persiaran Perdana BBN, Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

***Email:** i26044483@student.newinti.edu.my¹

Abstract

AI-generated political videos pose ethical issues related to truthfulness, consent, audience autonomy, and platform responsibility. This review discusses these issues in relation to TikTok and YouTube through a narrative synthesis of research studies published from 2022 until September 2026 and available primary governance documents. Recent research shows that deepfake detection varies with audiovisual characteristics and that warnings may not eliminate reliance on fabricated content. Audits of platforms reveal differences in political content exposure but do not establish the effects of synthetic videos on voting behavior or general radicalization effects. TikTok prohibits political promotion for money and misleading use of synthetic videos; YouTube relies on transparency policy, misinformation guidelines, and complaints about likeness, whereas Google has region-specific advertising policies. The review proposes a six-dimensional responsible synthetic political communication framework that translates existing governance principles into coordinated duties for disclosure, provenance, consent and identity protection, factual integrity, autonomy-preserving targeting, and correction and audit. Its contribution is to connect ethical responsibilities across production, distribution, and remedy; it remains a normative proposal requiring empirical evaluation.

Keywords

Artificial intelligence, Political communication, Media ethics, Deepfakes, Platform governance

Introduction

Political content generated by artificial intelligence (AI) entails fake political speeches, clones of politicians' voices, synthesized visuals, and manipulated videos related to politicians, political campaigns, and policies. The ethics of such content depend on the meaning of that content, its means of dissemination, and the audiences' ability to determine its evidence value. Empirical research on political speeches reveals varying evaluations of their authenticity based on modality

Submission: 22 September 2026; **Acceptance:** 20 October 2026; **Available online:** October 2026



Copyright: © 2026. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the web [site: https://creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)

and generation method (Groh et al., 2024). Qualitative research highlights audience trust and stakeholder collaboration as governance challenges (Leibowicz, 2026). Audits examine political recommendation patterns on TikTok and YouTube (Ibrahim et al., 2026; Jiang et al., 2026). Together, these findings indicate a need to examine governance responsibilities concerning both representation and distribution.

Governance differs between the two platforms. TikTok prohibits paid political advertising, including compensated political branded content produced by creators, while allowing organic political communication that complies with its policies (TikTok, 2026). YouTube demands the disclosure of realistic AI modifications, along with misinformation and identity protections (YouTube Help, n.d.-a, n.d.-b, n.d.-c). Comparative research identifies multiple safeguards beyond labeling (Gao et al., 2026). The four aspects considered include truth and context, autonomy and targeting, harm and identity, as well as accountability. Responsibilities for production, distribution, and correction are considered for Malaysia and ASEAN.

Methodology

A narrative review synthesizes heterogeneous evidence to address an interpretative question rather than exhaustively searching and pooling studies. Narrative review guidelines highlight the importance of having a clearly stated aim, scope, search strategy, and discussion of synthesis (Theile & Beall, 2024). It compares audience experiments, platform audits, policy analyses, and governance instruments with different outcomes.

The targeted search used "political deepfake" or "AI-generated political content," "TikTok," "YouTube," "detection," "persuasion," "trust," "disclosure," and "governance." During revision, supplementary title- and DOI-based web searches targeted Nature Portfolio, ACM Digital Library, SAGE Journals, Taylor & Francis Online, Wiley Online Library, Oxford Academic, AIS eLibrary, and PubMed (16 October 2026). The original search log was unavailable; these platforms document the supplementary verification. Bibliographic details and further sources were checked through publisher records and reference lists. The literature window was January 2022–September 2026; living platform policies were checked on 16 October 2026. Sources had to address synthetic political media, political exposure on either platform, or related governance. Peer-reviewed journal and conference articles and institutional documents were prioritized.

For each study, the synthesis records its design, unit of analysis, exposure setting, outcomes, and inferential limits. Evidence across four ethical topics informed the responsibilities in Table 2. Exposure evidence was separated from attitudes or behaviors; policy documents supplied standards rather than evidence of causal efficacy. The review is selective rather than systematic; therefore, it does not produce pooled effect sizes or formal risk-of-bias scores. Studies were selected for Table 1 if they provided experimental, audit, incident, or comparative policy evidence on political deepfakes or TikTok/YouTube governance; conversational persuasion and stakeholder interviews were used as supporting context.

Results and Discussion

Evidence on Detection, Persuasion and Trust

Detection, credibility, and persuasiveness are separate processes. Groh et al. (2024) carried out five preregistered experiments with 2,215 participants: audiovisual content facilitated better detection than text-based information; however, synthesized speech was more difficult to detect than voice actor audio clips. The research conducted in the Netherlands among 1,244 participants by Hameleers et al. (2022) shows that political deepfakes placed in the simulation of a social media environment were no more persuasive than plain text disinformation. Neither study considers ordinary use of TikTok or YouTube.

Clark and Lewandowsky (2026) found that participants continued to rely on the content of deepfake videos when judging alleged wrongdoing, even after receiving explicit warnings that the videos were fabricated. Across their three preregistered experiments, the scenarios involved fictional political figures and a social media influencer rather than actual electoral outcomes. Additional evidence comes from platform audits: partisan exposure disparities have been found in TikTok recommendations, whereas YouTube studies differentiate between recommendations and user decisions and observe distinctions between long-form videos and Shorts (Brown et al., 2026; Ibrahim et al., 2026; Jiang et al., 2026). Table 1 summarizes the selected empirical and comparative policy evidence, its research limitations, and the corresponding governance implications.

Table 1. Selected studies on synthetic political content and platform governance

Study and exposure setting	Reported finding and limit	Governance implication
Hameleers et al. (2022) Simulated social media; N = 1,244	Deepfakes were no more persuasive than textual disinformation; not a live-platform test.	Distinguish modality effects from platform effects.
Groh et al. (2024) Political-speech experiments; N = 2,215	Audiovisual cues improved discernment; synthetic audio reduced it.	Avoid blanket assumptions about human detection.
Clark and Lewandowsky (2026) Three controlled experiments	Warnings did not eliminate reliance on fabrications; fictional framing limits inference.	Test disclosure with context and remedies.
Ibrahim et al. (2026) TikTok audit; 323 experiments	Partisan exposure asymmetries during the 2024 election; no voting-effect estimate.	Audit distribution as well as uploads.
Brown et al. (2026) YouTube real-user audit; 2020 data	Content-related pathways, without average radicalization pathways in this setting.	Separate user choice from recommendation effects.
Jiang et al. (2026) YouTube long-form and Shorts audit	Recommendation patterns differed by video format; synthetic content was not isolated.	Evaluate Shorts and long-form separately.

Continue Table 1

Gao et al. (2026) Cross-platform policy analysis	Multiple governance strategies, including disclosure and monetization controls; not an effectiveness trial.	Coordinate safeguards across platform functions.
Husna Tamanna (2026) Cross-platform incident analysis; N = 1,498	Topics and communication goals varied across platforms; observational evidence.	Design context-sensitive mitigation; avoid causal claims.

Truth, Deception and Contextual Integrity

Contextual integrity requires enough information to evaluate a political statement. This includes the source, date, material edits, and whether attributed words were actually spoken. This is an operational definition of truthful depiction, not an application of a full privacy theory. TikTok explicitly targets manipulated and fictitious depictions. YouTube limits harmful manipulation and attribution (TikTok, 2025; YouTube Help, n.d.-b).

The disclosure of synthetic production does not mean that the original claim is true. Dependence on falsified claims even after warnings illustrates this distinction (Clark & Lewandowsky, 2026). On this basis, we propose an ethical approach that combines clear disclosure, sufficient source and contextual information, and correction of false claims. We recommend preserving satire and translation when their purpose and representational status are clear to audiences. We further propose that a warning label should not, by itself, make fabricated content ethically acceptable when it violates protections against harm or deceptive attribution. Contextual assessment should also extend to authentic videos, which can be presented in misleading ways.

Autonomy, Manipulation and Political Targeting

Persuasion becomes problematic from an ethical perspective if the audience cannot identify the sponsor, evaluate the claims, and comprehend how personal data are used. Salvi et al. (2025) discovered that when provided with participant information, GPT-4 outperformed human debate partners in the experimental setting. However, the published author correction by Salvi et al. (2026) identifies a calculation error affecting the comparison between personalized and non-personalized GPT-4. The corrected analysis yielded a p value of .0678, reported as .07, rather than .04; this comparison therefore did not reach statistical significance. Although the personalized GPT-4 condition outperformed the human-human baseline, the corrected findings do not establish an additional persuasive advantage attributable specifically to personalization. They support further investigation of AI-assisted persuasion but do not establish its effects in political videos on TikTok or YouTube.

The suggested protective measures include sponsorship disclosure, open access to all material variants of campaign messages, and prohibition of targeting based on sensitive characteristics and vulnerability assessment. Such measures are based on the transparency, privacy, and accountability principles from the regional AI ethics guidelines (ASEAN Secretariat, 2024, 2025). TikTok's existing prohibition on paid political advertising already includes

compensated political branded content produced by creators (TikTok, 2026). We therefore recommend strengthening the detection and enforcement of this prohibition in influencer collaborations, including undisclosed paid arrangements (TikTok, 2026).

Harm, Justice and Identity

Identity-based harm can occur where realistic synthetic media imputes speech or behavior to an individual without consent. TikTok prohibits AI-generated content that uses the likenesses of adult private persons without permission or falsely depicts public figures in prohibited contexts, such as fabricated endorsements (TikTok, 2025). The criteria considered by YouTube in evaluating likeness claims include realism, identifiability, disclosure, and public interest (YouTube Help, n.d.-c). In March 2026, YouTube expanded its likeness-detection pilot to government officials, journalists, and political candidates. However, it emphasized that detecting a likeness does not automatically result in content removal, because removal requests remain subject to policy assessment and consideration of public-interest exceptions (Hanif & Miller, 2026).

We propose that responses to identity-based harm should consider foreseeable harm, the available evidence, and the public-interest value of criticism and satire. We recommend giving particular weight to consent and timely remedies for private persons, while treating public-figure status as insufficient, by itself, to justify deceptive attribution. We also recommend accessible complaint mechanisms with explanations in languages appropriate to the affected communities. These proposals draw on the inclusiveness and human-benefit principles in Malaysia's AI governance guidelines (Ministry of Science, Technology and Innovation [MOSTI], 2024). This review does not identify which Malaysian communities are most affected by the harms examined.

Accountability and Democratic Trust

Accountability links upload decisions, amplification, complaints, and correction. Gao et al. (2026) have identified various ways of governing AI content on social media, and Leibowicz (2026) mentions trust and coordination based on interviews conducted with governance stakeholders. As far as TikTok and YouTube are concerned, the audit should identify the presence of labels upon exposure, resolution of identity complaints, and whether correction reaches viewers rather than relying on rules published as proof of good practice.

Provenance is another complementary way to establish information accountability. Content Credentials link signed creation and editing claims to an asset without guaranteeing the truth of the depicted event (Coalition for Content Provenance and Authenticity [C2PA], 2026). The same caution should be applied to recommendation evidence: modification of input news affected YouTube recommendations and exposure without a change in democratic attitudes of the survey respondents (Yu et al., 2024).

Comparison of TikTok and YouTube Governance

TikTok's advertising policy as of July 2026 prohibits paid political advertisements, sponsored political content, and promotion. In its integrity policies, TikTok requires disclosure of realistic AI-generated or materially altered content and prohibits certain kinds of misinformation

even when labeled (TikTok, 2025, 2026). We identify two governance issues requiring further assessment: the effectiveness of detecting and enforcing policy violations, including undisclosed paid political collaborations, and the governance of recommendation systems that distribute political content.

YouTube requires creators to disclose realistic AI-generated or materially altered content. It may also apply labels automatically to content created with its generative AI tools, content containing C2PA metadata, or content identified as AI-generated or altered by its internal systems (YouTube Help, n.d.-a). Misinformation restrictions and likeness complaint procedures complement disclosure (YouTube Help, n.d.-b, n.d.-c). Google has separate synthetic content policies for election advertisements, which are available in some places (Google Ads Help, n.d.).

This is therefore a comparison of various mixes of prevention, transparency, identity protection, and remedy. It does not establish a general hierarchy of either platform in terms of its ethical superiority. Gao et al. (2026) identify variation across AI-content governance mechanisms, including content moderation, disclosure, ownership, and monetization policies. Their findings support examining how multiple safeguards operate together rather than evaluating labeling in isolation. The YouTube case also illustrates why no single platform can be seen as a single exposure context (Jiang et al., 2026).

Regional and International Governance

The guidelines of UNESCO for digital platforms link freedom of expression to responsible and transparent governance (UNESCO, 2023). Under Article 50 of the EU AI Act, providers have certain machine-readable marking requirements, while deployers of deepfakes are under disclosure requirements. Artistic, fictional, and satirical content is subject to a specific disclosure requirement, but not exempted entirely (European Parliament & Council of the European Union, 2024). The European Commission published a voluntary Code of Practice in June 2026 to support implementation of these transparency requirements, which apply from 2 August 2026 (European Commission, 2026). The AI Act establishes the legal obligations, whereas the Code provides voluntary practical guidance for meeting them.

The 2024 ASEAN guidance and the 2025 extension for generative AI offer an organizational framework for transparency, risk management, and human supervision (ASEAN Secretariat, 2024, 2025). In Malaysia, the guidelines aim at supporting the voluntary adoption of responsible AI principles along with current legislation (MOSTI, 2024). Their applicability is in guiding the procedures of the platforms, campaigns, and independent assessment; they cannot be considered an election-specific enforcement regime.

Responsible Synthetic Political Communication Framework

Table 2 presents the proposed six-dimensional responsible synthetic political communication framework. The disclosure and factual-integrity dimensions distinguish recognizing synthesis from discounting fabricated claims, following Clark and Lewandowsky (2026). Provenance provides the verifiable history of the content (C2PA, 2026), whereas consent and identity protection are meant to handle the harm recognized by the platforms' guidelines.

Autonomy-preserving targeting is based on the privacy and transparency standards per region, while correction and auditing tie in these protective measures with stakeholder coordination recognized by Leibowicz (2026). Its contribution is to adapt existing UNESCO, ASEAN, and Malaysian principles into political-video-specific duties for creators, platforms, and oversight actors across production, distribution, and remedy (ASEAN Secretariat, 2024, 2025; MOSTI, 2024; UNESCO, 2023).

This is a normative synthesis rather than an established or empirically validated governance model, and it has not been tested as an intervention for preventing electoral harm. Label comprehension, provenance retention after posting, complaint resolution time, access to political-message archives, and correction reach can serve as potential indicators. Future research could compare TikTok, YouTube Shorts, long-form YouTube videos, and different language communities.

Table 2. Responsibilities across the synthetic political communication chain

Dimension	Creator or campaign	Platform	Regulator, civil society, or researchers
Conspicuous disclosure	Identify synthetic elements and sponsor.	Display understandable labels during exposure.	Test visibility and comprehension.
Provenance and traceability	Preserve sources and material edits.	Retain and interpret provenance across sharing.	Support interoperable verification.
Consent and identity protection	Obtain required consent; avoid deceptive attribution.	Offer accessible likeness complaints and appeals.	Evaluate equitable access to redress.
Factual and contextual integrity	Verify claims; distinguish reconstruction from record.	Review harmful election claims and misleading context.	Support trusted civic information.
Autonomy-preserving targeting	Identify sponsors; avoid sensitive-trait exploitation.	Explain targeting and preserve message archives.	Scrutinize profiling and campaign variants.
Correction and audit	Correct promptly through original channels.	Link remedies to copies and report outcomes.	Evaluate correction reach and enforcement.

Discussion

Controlled deepfake experiments assess perceptual judgments about a specific individual while recommendation audits assess exposure. The combination does not establish a causal link between deepfake exposure and changes in voting behavior. Brown et al. (2026), for instance, find content-related recommendation paths in their audit without average radicalization paths.

Clark and Lewandowsky (2026) found that participants may accept warnings yet still rely on fabricated content; the task-framing limitation in that study precludes the strong claim that all warnings are ineffective. Leibowicz (2026) also highlights the importance of trust and coordination to the governance system. The findings therefore suggest combining labels with contextual information and complaint mechanisms.

For Malaysia, we propose piloting the six dimensions through a campaign-video checklist: disclose synthetic elements and sponsors in Bahasa Melayu and other relevant community languages; retain source and editing records; document consent or a public-interest justification for identity use; verify claims and context; review targeting for sensitive-trait exploitation; and publish corrections with complaint-resolution records. Platforms, election authorities, journalists, and civil society could use shared incident reports to coordinate responses across ASEAN, assessing label comprehension, provenance retention, complaint resolution time, and correction reach. These are proposed applications of the regional guidance, not empirically demonstrated outcomes (ASEAN Secretariat, 2024, 2025; MOSTI, 2024).

TikTok prohibits paid political influencer content while YouTube's advertising guidelines differ based on jurisdiction. Neither of these alone constitutes evidence that the recommendation algorithms are safe. Direct analysis of deepfake incidents from across both platforms reveals differences in topics and aims of communication, but it does not reveal that platform affordances cause these differences (Husna Tamanna, 2026).

The selective search strategy and heterogeneous evidence limit completeness and causal inference. English-language searching may be underinclusive of local literature. Publication dates differ from data-collection dates: the TikTok audit examined the 2024 election, while Brown et al.'s YouTube audit used 2020 data. These findings do not establish current synthetic-content effects (Brown et al., 2026; Ibrahim et al., 2026). Research priorities for Malaysia and ASEAN include mobile video disclosure experiments, language variation evaluation, and incident audits of exposure, reporting, and correction of content while protecting political critique.

Conclusion

The governance of AI-generated political content on TikTok and YouTube needs to link truthful representation and audience autonomy with responsible distribution and remedy. The proposed governance framework entails requirements of transparency and responsibility regarding production and funding, factual accuracy, and dignity for the creators and political campaigns. Platforms are asked to provide adequate labeling, proportional sanctions, fair evaluation of recommendations, and recourse procedures. This list is based on the findings concerning detection rates, limitation of warnings, and differential exposure by platforms and does not establish electoral effects or validate the framework itself.

Acknowledgements

There are no grant or funding bodies to be acknowledged for preparing this paper.

Conflict of Interest

The authors declare no conflict of interest.

References

- ASEAN Secretariat. (2024). *ASEAN guide on AI governance and ethics*. <https://asean.org/book/asean-guide-on-ai-governance-and-ethics/>
- ASEAN Secretariat. (2025). *Expanded ASEAN guide on AI governance and ethics: Generative AI*. <https://asean.org/wp-content/uploads/2025/01/Expanded-ASEAN-Guide-on-AI-Governance-and-Ethics-Generative-AI.pdf>
- Brown, M. A., Bisbee, J., Lai, A., Bonneau, R., Nagler, J., & Tucker, J. A. (2026). Evaluating echo chambers, rabbit holes, and radicalization pathways on YouTube. *Political Communication*. Advance online publication. <https://doi.org/10.1080/10584609.2026.2671765>
- Clark, S., & Lewandowsky, S. (2026). The continued influence of AI-generated deepfake videos despite transparency warnings. *Communications Psychology*, 4(1), Article 13. <https://doi.org/10.1038/s44271-025-00381-9>
- Coalition for Content Provenance and Authenticity. (2026). *Content Credentials: C2PA technical specification* (Version 2.4). https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html
- European Commission. (2026, June 10). *Commission publishes Code of Practice on marking and labelling AI-generated content*. <https://digital-strategy.ec.europa.eu/en/news/commission-publishes-code-practice-marking-and-labelling-ai-generated-content>
- European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Gao, L., Ahmed, A., Chen, O., Reyl, M., Cheema, Z., Feamster, N., Tan, C., Thomas, K., & Chetty, M. (2026). Governance of AI-generated content: A case study on social media platforms. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (Article 1604, pp. 1–22). Association for Computing Machinery. <https://doi.org/10.1145/3772318.3790415>
- Google Ads Help. (n.d.). *Political content*. Retrieved October 16, 2026, from <https://support.google.com/adspolicy/answer/6014595>
- 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>
- Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2022). You won't believe what they just said! The effects of political deepfakes embedded as vox populi on social media. *Social Media + Society*, 8(3), Article 20563051221116346. <https://doi.org/10.1177/20563051221116346>
- Hanif, A., & Miller, L. (2026, March 10). Expanding likeness detection to civic leaders and journalists. *YouTube Blog*. <https://blog.youtube/news-and-events/expanding-likeness-detection-civic-leaders-journalists/>

- Husna Tamanna, A. (2026). Same frame, different story: A cross-platform analysis of political deepfakes. In *AMCIS 2026 Proceedings*. <https://aisel.aisnet.org/amcis2026/sigadit/sigadit/26/>
- Ibrahim, H., Jang, H. D., Aldahoul, N., Kaufman, A. R., Rahwan, T., & Zaki, Y. (2026). Systematic partisan content skews in TikTok during the 2024 US elections. *Nature*, 654(8120), 1004–1011. <https://doi.org/10.1038/s41586-026-10447-1>
- Jiang, S., Arya, A., Kweon, S., Liang, I., Kumar, D., & Vaccaro, K. (2026). Long story short: Auditing U.S. political polarization in recommendations for long- vs. short-form videos on YouTube. In *Proceedings of the ACM Web Conference 2026* (pp. 6147–6158). Association for Computing Machinery. <https://doi.org/10.1145/3774904.3792277>
- Leibowicz, C. R. (2026). Regulating reality: Exploring synthetic media through multistakeholder AI governance. *Policy & Internet*, 18(3), Article e70049. <https://doi.org/10.1002/poi3.70049>
- Ministry of Science, Technology and Innovation. (2024). *National guidelines on AI governance and ethics*. Government of Malaysia. <https://www.mosti.gov.my/wp-content/uploads/2024/09/NATIONAL-GUIDELINES-OF-AIGE-20241118.pdf>
- Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2025). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 9(8), 1645–1653. <https://doi.org/10.1038/s41562-025-02194-6>
- Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2026). Author correction: On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*. Advance online publication. <https://doi.org/10.1038/s41562-026-02588-0>
- Theile, C. M., & Beall, A. L. (2024). Narrative reviews of the literature: An overview. *Journal of Dental Hygiene*, 98(1), 78–82. <https://pubmed.ncbi.nlm.nih.gov/38346895/>
- TikTok. (2025, August 14). *Community guidelines: Integrity and authenticity*. <https://www.tiktok.com/community-guidelines/en/integrity-authenticity/>
- TikTok. (2026, July). *Politics, governments, and elections*. <https://ads.tiktok.com/resources/help/article/tiktok-ads-policy-politics-government-and-elections>
- UNESCO. (2023). *Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multi-stakeholder approach*. <https://unesdoc.unesco.org/ark:/48223/pf0000387339>
- YouTube Help. (n.d.-a). *Disclosing use of GenAI content*. Retrieved October 16, 2026, from <https://support.google.com/youtube/answer/14328491>
- YouTube Help. (n.d.-b). *Misinformation policies*. Retrieved October 16, 2026, from <https://support.google.com/youtube/answer/10834785>
- YouTube Help. (n.d.-c). *Protecting your identity*. Retrieved October 16, 2026, from <https://support.google.com/youtube/answer/2801895>
- Yu, X., Haroon, M., Menchen-Trevino, E., & Wojcieszak, M. (2024). Nudging recommendation algorithms increases news consumption and diversity on YouTube. *PNAS Nexus*, 3(12), Article pgae518. <https://doi.org/10.1093/pnasnexus/pgae518>