

Examining the Criminal and Societal Impacts of Fake News: A Contemporary Model of Digital Harm and Information Disorder

Ahmed Al-Harbi

Department of Law and Criminology, College of Legal Studies, Riyadh, Saudi Arabia

Sara Al-Qahtani

Department of Criminology and Criminal Justice, Institute of Social and Legal Research, Jeddah, Saudi Arabia

Received: 06 July 2026; **Accepted:** 03 August 2026; **Published:** 01 September 2026

Abstract: The rapid circulation of misleading, fabricated, and strategically manipulated information has transformed fake news from a communication problem into a multidimensional source of social, political, informational, and potentially criminal harm. This research examines fake news through an integrated framework connecting information disorder, digital harm, conspiracy narratives, moral panic, cyber-enabled misconduct, and regulatory responses. The study adopts a qualitative conceptual research design based exclusively on the supplied literature, combining thematic synthesis and comparative analysis to identify the principal mechanisms through which fake news generates harm. Particular attention is given to the definitional complexity of fake news, its technological amplification, its relationship with conspiracy theories and moral panics, and the difficulties associated with regulating harmful online content. The analysis indicates that fake news cannot be adequately understood as a simple category of false information. Instead, it operates within a broader digital ecosystem in which content production, algorithmic circulation, audience interpretation, political communication, and platform governance interact. The study proposes a Contemporary Digital Harm and Information Disorder Model consisting of five interconnected dimensions: content production, amplification, interpretation, harm generation, and governance response. Findings suggest that effective intervention requires a combination of detection mechanisms, contextual assessment, media literacy, proportionate regulation, and institutional accountability rather than reliance on content removal alone. The study contributes a structured analytical perspective for understanding fake news as both an informational phenomenon and a potential source of societal and criminal harm.

Keywords: Fake News, Information Disorder, Digital Harm, Cybercrime, Conspiracy Theories, Moral Panic, Social Media, Content Regulation, Misinformation, Digital Governance.

1. Introduction:

Digital communication technologies have fundamentally altered how information is produced, distributed, interpreted, and contested. Social networking platforms allow information to move rapidly between individuals, communities, institutions, and political actors. While this connectivity provides

substantial opportunities for communication and knowledge exchange, it also creates conditions in which misleading or fabricated information can reach large audiences with limited verification. The resulting problem is not restricted to inaccurate information itself; it concerns the consequences generated when false or misleading narratives become socially

influential.

A central difficulty in analysing fake news is definitional. Tandoc, Lim and Ling (2017) demonstrate that the term “fake news” encompasses multiple forms and interpretations rather than representing a single uniform phenomenon. This conceptual diversity is important because the consequences of fabricated satire, deliberate deception, propaganda-like content, manipulated information, and misleading reporting may differ considerably. Treating all such forms as identical can therefore produce analytical and regulatory weaknesses.

The criminal dimension of fake news is particularly significant because digital misinformation can contribute to harassment, reputational damage, public disorder, manipulation, fraud-like behaviour, and other forms of social harm. However, the classification of fake news as crime cannot automatically be assumed. Some false information may be harmful without satisfying the legal elements of a criminal offence, while some deliberately deceptive information may intersect with existing legal or regulatory categories. Consequently, the relationship between fake news and crime is better understood as a continuum extending from inaccurate communication to intentional digital harm.

The social consequences are equally important. Singh Grewal (2016) places conspiracy theories within the dynamics of a networked world, illustrating how interconnected communication environments can facilitate the circulation of conspiratorial interpretations. Simons (2019) further demonstrates how information can contribute to moral-panic processes in which particular groups, states, or actors become framed as sources of threat. These perspectives indicate that fake news can become consequential not merely because it is false, but because it can influence perceptions of reality and shape collective responses.

The COVID-19 pandemic provides a particularly relevant context for examining information disorder. UNICRI (2021) addresses the misuse of social media for disinformation, while WHO (2020) illustrates the importance of reliable and timely communication during a public-health crisis. Stanton (2020) similarly discusses the contemporary proliferation of conspiracy

theories. Together, these sources demonstrate how uncertainty can create an environment in which unsupported claims, conspiratorial narratives, and misleading information acquire greater visibility.

Regulatory responses have consequently become an important component of the fake-news debate. Schmitz and Berndt (2018), through their examination of Germany's Network Enforcement Act, illustrate the difficulties associated with using legal mechanisms to address harmful online content. The UK Parliament's Online Safety Bill (2022) represents another regulatory approach toward online safety. These developments raise a fundamental question: how can societies reduce demonstrably harmful digital information while avoiding excessive restrictions on legitimate expression?

This study therefore aims to develop an analytical model for understanding fake news as a contemporary source of criminal and societal harm. Specifically, it examines the conceptual foundations of fake news, identifies mechanisms of digital harm, analyses the role of conspiracy theories and moral panics, evaluates detection and regulatory approaches, and proposes an integrated framework connecting information production with societal consequences and governance responses.

2. LITERATURE REVIEW

2.1 Conceptualising Fake News and Information Disorder

The literature indicates that fake news is conceptually heterogeneous. Tandoc, Lim and Ling (2017) identify a range of scholarly understandings of fake news, demonstrating that the concept cannot be reduced to a single form of false reporting. This definitional diversity has direct methodological implications because identifying harmful content requires consideration of its intent, format, dissemination mechanism, and social context.

The distinction between false information and harmful information is particularly important. An inaccurate claim may have minimal consequences if it receives little attention, whereas an intentionally fabricated narrative can produce substantial harm when widely circulated. Therefore, the analytical significance of fake news lies partly in the interaction between falsity and dissemination.

UNESCO (2018) approaches fake news and disinformation from a journalism and education perspective, highlighting the importance of responsible information production and critical evaluation. This approach complements the typological perspective of Tandoc, Lim and Ling (2017) by shifting attention from classification toward the institutional and educational mechanisms needed to respond to information disorder.

2.2 Fake News, Conspiracy Theories and Networked Communication

The relationship between fake news and conspiracy theories constitutes a major theoretical dimension of the problem. Singh Grewal (2016) examines conspiracy theories within a networked environment, suggesting that digitally connected communication spaces can facilitate the circulation and social reinforcement of conspiratorial interpretations. Vittert (2019) similarly considers the rise of conspiracy theories in the United States.

Stanton (2020) situates conspiracy theories within a contemporary information environment characterised by widespread digital communication. The significance of these studies is that conspiracy narratives can transform isolated false claims into broader explanatory systems. Instead of merely communicating an incorrect fact, a conspiracy narrative may establish relationships between multiple claims and construct a perceived hidden cause.

Such processes increase the potential for social harm because audiences may interpret subsequent information through an established conspiratorial framework. In this context, fake news can operate as an entry point into broader information disorder.

2.3 Moral Panic and Social Harm

The theoretical concept of moral panic provides another useful perspective. Thompson (1998) examines moral panics as processes through which particular behaviours or groups become socially constructed as threats. Simons (2019) applies related concerns to contemporary media representations and political scapegoating.

The relevance to fake news lies in amplification. A fabricated or misleading claim can become socially significant when repeated through influential communication networks and interpreted as evidence

of a wider threat. The resulting harm may extend beyond the original falsehood to include distrust, hostility, stereotyping, and disproportionate social reactions.

Woods (2017), through the examination of antisocial media and crime-watching, further contributes to understanding how digital communication can affect perceptions of crime and social threat. These perspectives collectively support an interpretation of fake news as a social process rather than merely a defective information product.

2.4 Technological Detection and Digital Dissemination

Technological methods have increasingly become relevant to the identification of fake news. Singh, Ghosh and Sonagara (2021) examine multimodal analysis for detecting fake news stories, demonstrating the importance of analysing multiple forms of information rather than relying exclusively on textual characteristics.

This is significant because contemporary digital content can combine written language, images, videos, headlines, metadata, and other communication signals. A detection system capable of evaluating multiple modalities may therefore provide a stronger analytical basis than a text-only approach.

The problem is nevertheless complicated by the dynamic nature of online information. A detection system may identify suspicious characteristics, but determining whether content is genuinely harmful requires contextual and interpretive assessment. Technological detection should therefore be regarded as one component of a broader governance framework.

2.5 Regulation and Institutional Responses

Schmitz and Berndt (2018) demonstrate the complexities involved in regulating social-network content through legal mechanisms. Their discussion of the German Network Enforcement Act raises questions about enforcement effectiveness and the potential limitations of legal interventions.

The UK Parliament (2022) similarly illustrates the development of institutional approaches to online safety. These approaches indicate growing recognition that platforms and digital intermediaries cannot be separated entirely from questions concerning harmful

online communication.

UNICRI (2021) reinforces the need to address disinformation as a broader digital governance problem, particularly in crisis environments. UNESCO (2018), meanwhile, emphasises journalism education and information literacy. The literature therefore suggests that regulation, technological detection, professional journalism, and public education should be viewed as complementary rather than mutually exclusive interventions.

2.6 Research Gap

The principal gap identified in the supplied literature is the absence of a unified model connecting the conceptual nature of fake news with its technological dissemination, social interpretation, criminal potential, and regulatory response. Existing sources address specific components: definition and typology (Tandoc, Lim and Ling, 2017), conspiracy theories (Singh Grewal, 2016; Vittert, 2019; Stanton, 2020), moral panic (Thompson, 1998; Simons, 2019), technological detection (Singh, Ghosh and Sonagara, 2021), and regulation (Schmitz and Berndt, 2018; UK Parliament, 2022).

The present study addresses this gap by synthesising these dimensions into a contemporary digital-harm framework.

3. METHODOLOGY

3.1 Research Design

The study adopts a qualitative conceptual research design based exclusively on the eighteen references supplied for analysis. No external literature is introduced. The methodological objective is not to estimate the prevalence of fake news statistically but to develop an analytical model explaining how fake news may evolve from an information event into a source of social or criminal harm.

The research uses thematic synthesis as its primary analytical strategy. Each reference was examined according to its contribution to one or more analytical dimensions: conceptualisation, dissemination, technological detection, conspiracy formation, moral panic, societal harm, crisis communication, and regulatory intervention.

3.2 Analytical Framework

The proposed **Contemporary Digital Harm and**

Information Disorder Model (CDH-IDM) contains five interconnected stages:

Stage 1: Content Production.

The process begins with the creation or alteration of information. Content may be intentionally fabricated, misleadingly presented, selectively framed, or transformed into a narrative that departs from reliable evidence. The typological complexity identified by Tandoc, Lim and Ling (2017) provides the conceptual foundation for this stage.

Stage 2: Digital Amplification.

Once created, information can be distributed through interconnected digital networks. The importance of this stage arises from the networked environment discussed by Singh Grewal (2016). A claim that would otherwise remain marginal may gain substantial visibility when repeated by multiple users or communities.

Stage 3: Audience Interpretation.

The third stage concerns how audiences interpret information. Conspiracy theories become particularly relevant here because information may be incorporated into an existing explanatory belief system. Vittert (2019) and Stanton (2020) illustrate the continuing relevance of conspiracy narratives in contemporary communication environments.

Stage 4: Harm Generation.

Harm emerges when information influences behaviour, perceptions, institutional trust, public discourse, or relationships between social groups. Thompson (1998) and Simons (2019) provide a theoretical basis for understanding how information can contribute to broader processes of social threat construction and moral panic.

Stage 5: Governance Response.

The final stage involves institutional responses, including detection, moderation, education, regulation, and legal intervention. Schmitz and Berndt (2018), UNESCO (2018), UNICRI (2021), and UK Parliament (2022) collectively demonstrate the range of responses available.

The model can also be interpreted as a functional information-processing architecture. At the input level, content is generated and enters a digital

communication environment. At the processing level, multimodal detection mechanisms can assess textual, visual, and contextual indicators, drawing conceptually on the multimodal detection approach discussed by Singh, Ghosh and Sonagara (2021).

The output of the detection stage should not automatically be interpreted as proof of criminality. Instead, suspicious content should undergo contextual assessment. This distinction is essential because false information, satire, political disagreement, and deliberate harmful deception may have different characteristics.

A hypothetical example illustrates the architecture. Suppose a fabricated claim is published concerning a public institution during a period of uncertainty. The claim is shared across several networked communities, becomes integrated into an existing conspiracy narrative, and eventually generates public distrust. A purely technical detection system might identify linguistic or multimodal anomalies. The CDH-IDM framework additionally asks whether the information was deliberately constructed, how it was amplified, how audiences interpreted it, what harm emerged, and which governance response would be proportionate.

3.4 Analytical Criteria

The model evaluates digital information through five criteria: **falsity or misleadingness, intentionality, amplification, consequence, and governance relevance**. These criteria prevent the analysis from treating all inaccurate information as equally harmful.

Intentionality is especially significant for distinguishing accidental misinformation from deliberate deception. Consequence determines whether an information event remains relatively insignificant or develops into a meaningful source of social harm. Governance relevance concerns whether intervention is justified and, if so, what form it should take.

3.5 Methodological Limitations

The methodology is conceptual rather than empirical. It does not provide statistical measurements of fake-news prevalence, causal estimates of behavioural effects, or experimental validation of the proposed framework. Additionally, the supplied literature covers different contexts and purposes, meaning direct comparison between all sources is not always possible. The model should therefore be considered an

analytical framework requiring future empirical testing.

4. RESULTS

The analysis produced five principal findings. First, fake news is best understood as a heterogeneous information phenomenon rather than a single category of false content. The definitional diversity identified by Tandoc, Lim and Ling (2017) demonstrates that content classification is itself a significant analytical challenge. Consequently, policies based solely on a binary true/false distinction may fail to capture differences in intention, presentation, context, and consequence.

Second, the potential harm of fake news is strongly connected to digital amplification. The networked environment changes the significance of individual pieces of information because claims can move rapidly between communities. The literature concerning conspiracy theories indicates that network connectivity can facilitate repeated exposure and collective reinforcement. This means that the social impact of a false claim cannot be evaluated solely according to its original source.

Third, the analysis demonstrates a substantial relationship between fake news, conspiracy narratives, and moral-panic processes. Singh Grewal (2016), Vittert (2019), Stanton (2020), Thompson (1998), and Simons (2019) collectively indicate that information can become socially consequential when it is incorporated into broader narratives concerning threat, blame, or hidden causation. Fake news therefore has the potential to generate secondary harms extending beyond the original misinformation.

Fourth, technological detection represents an important but incomplete response. Singh, Ghosh and Sonagara (2021) demonstrate the relevance of multimodal analysis to fake-news detection. However, detection alone cannot determine whether a piece of content constitutes criminal conduct or produces substantial social harm. Contextual interpretation, intentionality, dissemination patterns, and consequences remain necessary.

Fifth, regulation presents a fundamental governance trade-off. Schmitz and Berndt (2018) illustrate the complexity of legal intervention on social networks, while the UK Parliament (2022) demonstrates institutional attempts to address online safety. The findings suggest that content removal and legal

enforcement should not be treated as universal solutions. UNESCO (2018) and UNICRI (2021) indicate the importance of complementary educational, journalistic, and institutional strategies.

The resulting CDH-IDM framework therefore identifies a sequential but interactive relationship: **content production → digital amplification → audience interpretation → harm generation → governance response**. The stages are not strictly linear because governance interventions can occur at multiple points. For example, media literacy can influence audience interpretation, while platform moderation can interrupt amplification. The model consequently supports a multi-layered approach to digital harm.

5. DISCUSSION

The findings indicate that the criminal and societal dimensions of fake news should be analytically distinguished but not separated. Criminality concerns the relationship between harmful information and legally recognised prohibited conduct, whereas societal harm includes broader consequences such as distrust, polarisation, reputational damage, fear, and distorted perceptions of social threats. Not every harmful falsehood is necessarily a crime, and not every instance of misinformation produces substantial harm. The CDH-IDM model accommodates this distinction by positioning criminal potential within the broader harm-generation stage.

The conceptual problem identified by Tandoc, Lim and Ling (2017) has significant implications for both researchers and policymakers. If fake news is defined too broadly, legitimate disagreement, satire, error, and political criticism may be incorrectly classified as harmful misinformation. Conversely, if the definition is too narrow, strategically deceptive content may escape appropriate scrutiny. A functional approach that considers content characteristics together with intent, dissemination, and consequences therefore offers greater analytical precision.

The relationship between fake news and conspiracy theories also complicates conventional approaches to content moderation. A single false statement may appear relatively insignificant when considered independently but can acquire greater influence when embedded within a wider explanatory narrative. Singh Grewal (2016), Vittert (2019), and Stanton (2020)

provide grounds for understanding this cumulative process. Consequently, interventions focused exclusively on individual pieces of content may overlook the network and narrative structures through which information becomes influential.

The moral-panic perspective provides another important theoretical insight. Thompson (1998) and Simons (2019) suggest that communication can participate in processes through which threats are constructed and amplified. This creates a governance dilemma: authorities attempting to counter misinformation may themselves contribute to the visibility of the contested narrative. Therefore, intervention must consider not only whether content is false but also how enforcement might influence public attention and interpretation.

Technological detection presents both opportunities and limitations. Multimodal analysis, as discussed by Singh, Ghosh and Sonagara (2021), can improve the identification of suspicious material by examining multiple information characteristics. However, technical systems may struggle with contextual meaning, irony, evolving narratives, and distinctions between harmful deception and legitimate expression. Automated detection should consequently support, rather than completely replace, contextual human assessment.

The regulatory literature highlights a further trade-off between platform responsibility, public safety, and freedom of expression. Schmitz and Berndt (2018) demonstrate that legal obligations imposed on social networks can create practical and normative difficulties. The Online Safety Bill discussed by UK Parliament (2022) illustrates continuing institutional efforts to address these concerns. Regulation should therefore be proportionate, transparent, and sensitive to differences between types and levels of harm.

The proposed model also has practical implications for journalism and public education. UNESCO (2018) emphasises journalism education, while UNICRI (2021) demonstrates the importance of addressing disinformation during periods of heightened uncertainty. These approaches suggest that long-term resilience cannot be achieved through enforcement alone. Audiences require the capacity to evaluate claims, recognise manipulative narratives, and

distinguish evidence-based reporting from unsupported assertions.

The principal limitation of the present study is that the proposed model has not been empirically validated. Future research should operationalise the five dimensions and test the framework across different platforms, information crises, and national regulatory systems. Nevertheless, the synthesis provides a structured basis for examining fake news as a multidimensional digital-harm problem rather than simply as a problem of inaccurate reporting.

6. CONCLUSION

Fake news has evolved into a complex phenomenon situated at the intersection of communication, technology, social psychology, governance, and criminal harm. The literature examined in this study demonstrates that its significance cannot be explained solely by the falsity of individual claims. Rather, harm emerges through the interaction between content production, digital amplification, audience interpretation, social narratives, and institutional responses.

The study's principal contribution is the Contemporary Digital Harm and Information Disorder Model, which integrates five dimensions: content production, amplification, interpretation, harm generation, and governance response. This framework provides a conceptual mechanism for distinguishing the existence of false information from the processes through which it becomes socially or potentially criminally consequential.

The analysis further demonstrates that fake news should not be addressed through a single intervention. Technological detection can identify suspicious content; journalism and media education can strengthen information resilience; platform governance can interrupt harmful amplification; and proportionate legal mechanisms can address conduct that crosses relevant regulatory or criminal boundaries. The literature also indicates that conspiracy theories and moral-panic processes can intensify the effects of misleading information by transforming isolated claims into broader narratives of threat and blame.

Future research should empirically test the proposed framework through cross-platform datasets, content-analysis experiments, multimodal detection systems,

and comparative evaluations of regulatory models. Particular attention should be given to distinguishing misinformation from intentional deception and to measuring how different forms of digital information translate into measurable social harm. Such research could strengthen the theoretical and practical foundations for developing safer, more resilient, and more accountable digital information environments.

REFERENCES

1. Schmitz, S. and Berndt, C.M. (2018) The German Act on Improving Law Enforcement on Social Networks (NetzDG): A blunt sword? SSRN, Available at: <https://ssrn.com/abstract=3306964> (Accessed: 4th December 2022).
2. Silverman, C. and Alexander, L. (2016) How teens in The Balkans are duping Trump supporters with fake news. BuzzFeedNews, 3 November. Available at: <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo> (Accessed: 3rd December 2022).
3. Simons G (2019) The anatomy of a moral panic: Western mainstream media's Russia scapegoat. *Changing Societies & Personalities* 3(3), 189.
4. Singh Grewal, D. (2016) Conspiracy theories in a networked world. *Critical Review*, 28(1), 24.
5. Singh, V., Ghosh, I., Sonagara, D. (2021) Detecting fake news stories via multimodal analysis. *Journal of the Association for Information Science and Technology*, 72(1), 3.
6. Stanton, Z. (2020) You're living in the golden age of conspiracy theories. Politico, Available at: <https://www.politico.com/news/magazine/2020/06/17/conspiracytheories-pandemic-trump-2020-election-coronavirus-326530> (Accessed: 3rd December 2022).
7. Suanpang, P., and Jamjuntr, P. (2021) A comparative study of deep learning methods for forecasting tourism business recovery from the covid 19 pandemic crisis. *Journal of Management Information and Decision Sciences*, 1-10. Available at: <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-1543565> (Accessed: 5th December 2022).

8. Suanpang, P., Netwong, T., Kaewyoung, P., Chunhapataragul, T., Nermson, C. (2022) Innovation of Smart Tourism to Promote Tourism in Suphanburi Province. Bangkok: Suan Dusit University.
9. Suanpang, P., Pothipasa, P., Netwrong, T. (2021) International Journal of Cyber Criminology. 15(1), 143.
10. Tandoc EC, Lim ZW and Ling R (2017) Defining 'fake news': A typology of scholarly definitions. Digital Journalism 6(2): 137.
11. Thompson, K. (1998) Moral Panics, London: Routledge.
12. UK Parliament (2022) Online Safety Bill. Available at: <https://publications.parliament.uk/pa/bills/cbill/58-03/0209/220209.pdf> (Accessed: 25th November 2022).
13. UNESCO (2018) Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training. Available at: <https://en.unesco.org/fightfakenews> (Accessed: 7th December 2022).
14. UNICRI (2021) Stop the virus of disinformation. Available at: <https://unicri.it/sites/default/files/2020-11/SM%20misuse.pdf> (Accessed: 3rd December 2022).
15. Vittert, L. (2019) Are conspiracy theories on the rise in the US? The Conversation. Available at: <https://theconversation.com/are-conspiracy-theories-onthe-rise-in-the-us-121968> (Accessed: 3rd December 2022).
16. WHO (World Health Organization) (2020) Coronavirus disease 2019: Situation Report – 45. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4 (Accessed: 25th November 2022).
17. Woods, M. (2017) Antisocial Media: Crime-watching in the Internet Age – Palgrave Studies in Crime, Media and Culture, Melbourne: Springer International Publishing AG
18. Young J (1971) The Drugtakers: The Social Meaning of Drug Use. Lond