Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 24 July 2026
Revised 20 Aug 2026
Accepted 07 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (22s): 20–31

KEYWORDS:
Artificial intelligence, AI-enabled scams, Indian news media, corpus-assisted discourse analysis, media framing, deepfakes, digital fraud, technological risk.

Framing AI-Enabled Scams in Indian News Media: A Corpus-Assisted Discourse Analysis

Anil Kumar, Jaskirat Singh & Dr. Palwinder Singh Bhatia

Assistant Professor, Department of Mass Communication,
Guru Nanak Dev University, Amritsar, India
Email: jaskirats300@gmail.com
Email: drpalwindersinghbhatia@gmail.com
Email: anilramkali.jind@gmail.com

ABSTRACT: The rapid proliferation of artificial intelligence (AI) and generative AI technologies has significantly expanded the channels through which digital fraud and deception occur. These include voice cloning, deepfake impersonation, AI-driven digital identities, phishing schemes, investment fraud, and various automated scam operations. This research investigates the construction of AI-facilitated scams within Indian news discourse by applying corpus-assisted discourse analysis (CADA) to a curated dataset of 234 news reports published across 37 media outlets between July 17, 2023, and September 28, 2026. The dataset contains 216 English-language news reports, alongside descriptive sub corpora of 15 Hindi-language and 3 Gujarati-language reports. Utilizing corpus-linguistic techniques integrated with qualitative discourse analysis, the study evaluates lexical distribution, recurring phraseology, framing techniques, social-actor representation, grammatical agency, victimization narratives, technological risk, and institutional countermeasures.

The preliminary empirical evaluation focuses on headline fields and metadata. It details corpus composition, temporal distribution, source coverage, headline terminology, and initial framing patterns, distinguishing these findings from deeper full-text corpus analyses that rely on body text. The headline audit indicates that terms such as "AI scam," "deepfake," and "deepfake scam" serve as the primary lexical markers. Secondary headline patterns frequently emphasize financial loss, victim vulnerability, preventative measures, institutional actions, and emerging threats, while positive framing highlighting innovation or technological benefit remains minimal. These observations demonstrate that news coverage primarily frames AI around the intersection of digital deception, criminal activity, and societal risk. Concurrently, AI is occasionally portrayed as a mechanism for fraud detection and defense. This work provides a transparent analytical framework for evaluating AI-driven crime narratives within Indian journalism and underscores why headline analyses must be supplemented with comprehensive full-text investigations of grammatical agency, collocations, social actors, and discourse frames.

1. INTRODUCTION

Artificial intelligence (AI) and generative AI are reshaping the digital communication landscape while creating novel avenues for cyber-enabled fraud and deceptive practices. Contemporary AI systems can produce remarkably authentic text, audio, images, and video content. Malicious actors leverage these tools to impersonate individuals, alter evidence, and execute automated interactions at unprecedented speeds and scales. From a criminological standpoint, AI should be evaluated not merely as a technological innovation, but as a critical catalyst for modified and novel forms of criminal behavior (Hayward & Maas, 2021; King et al., 2020).

Deepfake technologies and synthetic media represent critical challenges within this emerging risk landscape due to their ability to replicate recognizable identities, voices, and faces with high fidelity. Criminals integrate these capabilities into traditional fraud schemes, such as financial deception, identity theft, social engineering, and corporate fraud. De Rancourt-Raymond and Smaili (2023) highlight how deepfake systems exploit established interpersonal trust by fabricating plausible voices, identities, and digital communications. Furthermore, empirical studies indicate that public exposure to deepfakes can undermine confidence in media authenticity, including within non-Western contexts (Ahmed et al., 2023).

Consequently, AI-driven fraud should not be interpreted strictly as an isolated cybersecurity or technical challenge; it represents a broader communicative and discursive phenomenon. News outlets play a central role in translating complex technological risks for the public through headline choices, lexical selections, source attribution, and descriptions of victims, perpetrators, and institutional interventions. Media scholarship demonstrates that news reporting does not merely report technical milestones; it establishes interpretive frames through which audiences conceptualize the broader societal impacts of AI (Nguyen & Hekman, 2024). Recent literature further confirms that journalistic coverage of AI varies significantly across national settings, incorporating competing narratives of risk, governance, economic opportunity, and technological progress (Ittafaq et al., 2025; Jhan & Kuriakose, 2026).

Media framing gains heightened importance when AI is contextualized alongside criminality, financial loss, fear, and victimization. News narratives can depict AI as an instrument manipulated by human offenders, an amplifying force that expands fraud efficiency, a source of digital unpredictability, or a defensive instrument deployed for threat detection and prevention. These varying representations shape how public discourse assigns responsibility among perpetrators, victims, technology developers, online platforms, banking institutions, regulatory bodies, and state authorities. Research in discourse analysis emphasizes the importance of evaluating how language selection attributes agency to technical systems versus human actors (Nguyen & Wischniewski, 2026; Tandoc et al., 2025).

Although scholarly interest in deepfakes, AI crime, misinformation, and media framing has grown, the linguistic representation of AI-enabled scams in Indian news coverage remains under-researched. Existing technical research prioritizes detection algorithms and synthetic content generation; criminological scholarship focuses on criminal applications and threat vectors; and media studies tend to evaluate broad AI narratives. This study bridges these domains using a corpus-assisted discourse framework to systematically examine lexical choices, phraseological patterns, framing techniques, victim representations, social actor roles, agency, and institutional responses within Indian press reporting.

Accordingly, this research investigates how Indian news media discursively construct AI-facilitated scams and how relations between AI platforms, perpetrators, victims, and regulatory entities are established. The corpus-assisted methodology combines quantitative measurements of linguistic frequency with qualitative discourse analysis to reveal the social functions of these recurring media patterns.

2. REVIEW OF LITERATURE AND THEORETICAL FRAMEWORK

2.1 AI, Crime, and Fraud

Scholarship at the intersection of AI and criminology highlights that artificial intelligence introduces both societal advancements and significant harms. King et al. (2020) outline key potential criminal applications of AI—such as identity manipulation, automated deception, and target exploitation—while evaluating questions of regulatory oversight, legal liability, and prevention strategies. Similarly, Hayward and Maas (2021) stress that criminological inquiry must address how digital infrastructure transforms traditional criminal techniques while giving rise to novel forms of cyber offense.

Studies focusing on deepfake-driven fraud illustrate how synthetic media integrates into established scam operations. De Rancourt-Raymond and Smaili (2023) show that synthetic media enhances deceptive practices by reproducing recognizable voices and behavioral cues, thereby exploiting systemic trust. This perspective is vital for discourse analysis, as news reports on AI fraud simultaneously conceptualize technology as a tool, perpetrators as deliberate actors, victims as vulnerable targets, and institutions as mechanisms of protection and governance.

2.2 Deepfakes, Synthetic Media, and Deception

Deepfake research has expanded beyond technical detection mechanisms to address broader legal, social, and policy implications. In a systematic review, Abbas and Taichahgh (2024) identify ongoing vulnerabilities involving content creation,

verification limits, misinformation spread, and governance frameworks. Sandoval et al. (2024) emphasize the challenges synthetic media presents to criminal justice systems, including evidentiary ambiguity, identity verification hurdles, and declining public trust in official records. These insights demonstrate that deepfakes function as complex social and communicative challenges rather than isolated engineering issues.

Media studies further demonstrate that deepfake reporting frequently intersects with discussions surrounding misinformation, political manipulation, eroded public trust, and digital authenticity (Gosse & Burkell, 2020; Wahl-Jorgensen & Carlson, 2021). These dynamics directly inform the study of AI-enabled scam coverage, as financial fraud similarly relies on manipulating authenticity and interpersonal trust. Analyzing the linguistic structure of scam reporting clarifies how media coverage contributes to broader cultural narratives regarding technological insecurity.

2.3 Media Framing of Artificial Intelligence

Framing theory offers a useful conceptual lens for examining how journalistic practices highlight specific dimensions of an issue while minimizing others. Analyzing 3,098 AI-focused news articles, Nguyen and Hekman (2024) show that media narratives feature diverse interpretive perspectives, with critical and risk-oriented frames increasing over time. Their findings connect news coverage of AI with cybercrime, surveillance, data privacy concerns, and information manipulation.

Ittafaq et al. (2025) broaden this scope by analyzing 38,787 news items across 12 countries, showing that AI reporting is influenced by national media systems and localized thematic priorities. Examining 25,595 headlines, Jhan and Kuriakose (2026) classify primary media themes into economic development, social transformation, technological innovation, and political governance. These studies confirm the utility of computational corpus methods while underscoring the necessity of qualitative contextual interpretation.

Additional research demonstrates that media narratives can depict AI as both an asset and a threat. Hendrickx and Van Coppenolle (2026) analyze how mainstream journalism frames generative AI through contrasting opportunity and risk lenses. Studies examining agency in media discourse note that journalistic phrasing frequently attributes autonomous, human-like actions to technological systems (Nguyen & Wischniewski, 2026). These insights reinforce the importance of examining grammatical agency to understand the interaction between human actors and technical infrastructure.

2.4 Corpus-Assisted Discourse Analysis

Corpus-assisted discourse analysis (CADA) integrates quantitative linguistic data with qualitative text interpretation. Baker et al. (2008) demonstrate how frequency distributions, key terms, collocational networks, and concordance lines reveal dominant discourse patterns that can subsequently be contextualized through detailed textual analysis. This approach is particularly suited for evaluating news reporting on AI scams, moving beyond broad topic classification to analyze how linguistic structures shape relationships between technologies, perpetrators, victims, and state authorities.

The present study utilizes corpus tools as an empirical foundation rather than equating numerical frequency directly with discursive meaning. Quantitative metrics reveal patterns of recurrence and collocational association, concordance checks provide contextual evidence, and qualitative analysis evaluates how these elements construct frames and social identities.

2.5 Indian Media Context

Investigating news coverage within the Indian media environment is essential, as news framing is influenced by distinct institutional, linguistic, and socio-cultural factors. Chen and Liu (2024) show the efficacy of corpus-based approaches in comparing press representations between India and China. Kumar (2022) highlights the necessity of methodological transparency in Indian fact-checking practices, while recent journalism studies examine how Indian news outlets construct stories surrounding organized crime and international events (Verma et al., 2026; Kumar et al., 2026). These contributions emphasize that the Indian news media must be analyzed as a multi-faceted domain rather than a uniform entity.

The multi-language structure of Indian media further demands rigorous corpus design. Although this dataset includes English, Hindi, and Gujarati materials, the structural disparity in sample sizes requires that cross-linguistic analysis remain strictly descriptive unless significantly larger, balanced subcorpora are developed.

2.6 Research Gap

This study addresses three primary research gaps. First, while technical research on AI deception emphasizes detection systems and generative algorithms, the linguistic construction of these risks in journalism receives less analytical focus. Second, media framing research often approaches AI at a macro level, whereas AI-facilitated scams involve a distinct convergence of criminal activity, technology, financial loss, victim vulnerability, and regulatory intervention. Third, corpus-assisted studies focusing on these dynamics within Indian news media remain limited. This research fills these gaps by pairing corpus-linguistic techniques with discourse analysis centered on framing, agency, and social actor representations.

3. AIM, RESEARCH QUESTIONS, AND OBJECTIVES

3.1 Aim

This study examines the linguistic and discursive framing of AI-enabled scams in Indian news coverage, focusing on framing strategies, lexical patterns, phraseological associations, grammatical agency, victim representations, technological threats, and institutional responses.

3.2 Research Questions

RQ1. What key lexical choices and phraseological patterns characterize Indian news coverage of AI-enabled scams?

RQ2. Which keywords, collocations, clusters, and semantic networks are deployed to discursively construct AI-enabled scams in Indian media?

RQ3. How are victims, perpetrators, AI systems, state institutions, digital platforms, and other social actors portrayed within the discourse?

RQ4. How are grammatical agencies and legal/moral responsibility linguistically distributed among human actors, AI technologies, regulatory bodies, and digital platforms?

3.3 Research Objectives

RO1. Identify the principal lexical and phraseological patterns used in Indian news reporting on AI-enabled scams.

RO2. Examine keywords, collocations, clusters and semantic associations associated with AI, scams, fraud, deepfakes, voice cloning, victims, perpetrators and related concepts.

RO3. Analyze the discursive representation of victims, perpetrators, AI technologies, institutions, digital platforms and other relevant social actors.

RO4. Examine how grammatical and lexical choices distribute agency, causality, responsibility and accountability among human actors, AI technologies, institutions and platforms.

4. METHODOLOGY

4.1 Research Design

The study adopts a corpus-assisted discourse analysis (CADA) design combining quantitative corpus-linguistic procedures with qualitative discourse analysis. Corpus procedures are used to identify recurrent lexical and phraseological patterns, while qualitative analysis is used to interpret how these patterns construct crime, risk, victimhood, agency, responsibility and institutional response. The approach is descriptive and interpretive rather than experimental.

4.2 Corpus Construction

The purpose-built corpus contains 234 news articles published between 17 July 2023 and 28 September 2026 and drawn from 37 Indian media outlets. The corpus contains 216 English-language articles, 15 Hindi-language articles and 3 Gujarati-language articles. English constitutes 92.3% of the corpus and is therefore designated as the primary inferential corpus. Hindi and Gujarati are treated as separate descriptive subcorpora to avoid indiscriminate mixing of languages with different lexical, morphological and syntactic properties.

Articles were selected for relevance to AI-related scams, fraud, deception, impersonation, deepfakes, voice cloning or related criminal activities involving AI. The metadata record includes article identifier, indexing/publication date, language, media outlet, URL and headline. The audit found 234 unique article IDs and 234 unique URLs. Two repeated headline pairs represent separately published items and were retained as individual records.

Because the current workbook contains article metadata and headlines rather than the complete body text of all 234 articles, the present Results section distinguishes the completed metadata/headline audit from full-text corpus analyses that require the article-body corpus. This distinction is essential for methodological transparency.

4.3 Data Cleaning and Analytical Unit

The article is the principal sampling unit. For the completed metadata audit, the headline is the principal textual unit. For the planned full-text analysis, the article body will constitute the corpus text, with words and lexical bundles treated as quantitative units and clauses or sentences treated as units for qualitative analysis. Full-text preprocessing should remove navigation material, advertisements, duplicated boilerplate, URLs and HTML artefacts while retaining headlines, body text and attributable quotations.

4.4 Corpus-Linguistic Analysis

AntConc 4.4.2 is designated as the principal corpus-analysis application for word frequency, keyword analysis, collocations, clusters/n-grams, concordances and concordance plots. Raw frequency should be supplemented by normalized frequency per 10,000 words where corpus sizes differ across periods or subcorpora. Keyword analysis requires a clearly specified reference corpus, while collocation analysis requires a reported span/window, minimum frequency and association statistic. These parameters should be documented before final full-text inference.

A critical distinction is maintained between article frequency and token frequency. Article frequency indicates the number of articles containing a form, whereas token frequency indicates the number of occurrences across the corpus. The distinction prevents a term repeated many times in one article from being treated as equivalent to a term distributed across many articles.

4.5 Discourse and Frame Coding

The qualitative framework consists of seven overlapping frames: Crime/Fraud; Threat/Risk; Victimization; Financial; Cybersecurity; Institutional Response; and Innovation/Protection. Frames are not mutually exclusive because a single news report may simultaneously construct an event as criminal, financially harmful, technologically sophisticated and institutionally challenging.

Table-1: Discourse and Frame Coding

Frame	**Operational focus**
Crime/Fraud	AI-enabled deception as criminal activity
Threat/Risk	Danger, warning, vulnerability and uncertainty
Victimization	Deception, vulnerability, loss and victim experience
Financial	Monetary loss, investment, banking and payments
Cybersecurity	Digital protection, detection and cybercrime prevention
Institutional Response	Police, government, banks, regulators and platforms
Innovation/Protection	AI as a tool for prevention, detection or security

4.6 Social Actors and Agency

The analysis distinguishes human, institutional and technological actors. Human actors include scammers, victims, police, officials and experts; institutional actors include governments, banks, regulators, technology companies and platforms; technological actors include AI systems, generative AI, deepfake technologies, voice-cloning systems and algorithms. Agency is examined through active and passive constructions, material processes, causative constructions, nominalizations and actor–

verb associations. This enables the study to examine whether AI is represented as an instrument used by people, an enabling force, an autonomous actor, or a protective technology.

4.7 Temporal Analysis

Temporal comparison is conducted at the level of normalized linguistic frequencies rather than raw article counts where full-text data permit. The period 2026 is incomplete, ending on 28 September, and therefore should not be treated as a complete annual observation. The current headline audit provides descriptive temporal indicators that can guide subsequent full-text analysis but cannot independently establish statistically significant discourse change.

4.8 Ethics

The study uses publicly accessible news content and does not involve direct participation by human subjects. Sources are acknowledged, copyrighted material is not reproduced beyond what is necessary for scholarly analysis, and private individuals can be anonymized where detailed reporting creates unnecessary exposure. The study analyzes media representations rather than determining the factual guilt or innocence of individuals named in news reports.

5. RESULTS

5.1 Corpus Integrity and Composition

Table-2: Corpus Integrity and Composition

Corpus characteristic	Number	Percentage
Total articles	234	100.0%
English	216	92.3%
Hindi	15	6.4%
Gujarati	3	1.3%
Media outlets	37	—
Unique article IDs	234	100.0%
Unique URLs	234	100.0%
Unique cleaned headlines	232	99.1%

The metadata audit establishes a corpus of 234 articles with 234 unique identifiers and URLs. The cleaned headline field contains 232 unique headlines, with two repeated headline pairs published by different outlets. The corpus is therefore internally coherent at the metadata level, although the uneven distribution across outlets and languages limits its interpretation as an equally weighted representation of Indian journalism.

5.2 Temporal Distribution

Table-3: Temporal Distribution

Year	Articles	Percentage
2023	35	15.0%
2024	47	20.1%
2025	69	29.5%
2026*	83	35.5%
Total	234	100.0%

*2026 covers 1 January to 28 September only. The corpus contains 152 articles from 2025–2026, representing 65.0% of the total dataset. The concentration of records in later years should therefore be considered when interpreting temporal patterns.

5.3 Media Outlet Distribution

Table-4: Media Outlet Distribution

Outlet	Articles
indiatimes.com	61
mid-day.com	18
news18.com	15
hindustantimes.com	14
livemint.com	13
freepressjournal.in	12
thehindu.com	9
financialexpress.com	9
ndtv.com	8
indianexpress.com	8
india.com	7
newindianexpress.com	7
deccanherald.com	5
aajtak.in	5
newsx.com	4
dnaindia.com	4
thelogicalindian.com	4
medianama.com	3
deccanchronicle.com	3
sandesh.com	3
firstpost.com, thequint.com, telegraphindia.com, amarujala.com, boomlive.in	2 each
12 other outlets	1 each

The six highest-volume outlets contribute 133 of the 234 articles (56.8%), while the twelve highest-volume outlets contribute 181 (77.4%). This concentration means that outlet-specific editorial and indexing practices may influence aggregate patterns.

5.4 Headline-Level Lexical Patterns

The following results are headline-level article counts rather than full-text token frequencies. They should therefore be interpreted as a corpus audit rather than as the final corpus-linguistic findings.

Table-5: Headline-Level Lexical Patterns

Search expression	Articles	Percentage
AI scam	35	15.0%
deepfake	20	8.5%
deepfake scam	10	4.3%
cybercrime	3	1.3%
investment scam	6	2.6%
romance scam	5	2.1%
scammer	5	2.1%
voice cloning	5	2.1%
victim	4	1.7%
phishing	2	0.9%
AI fraud	1	0.4%
digital arrest	1	0.4%
AI-enabled fraud	0	0.0%
AI impersonation	0	0.0%
deepfake fraud	0	0.0%
synthetic media	0	0.0%
generative AI	0	0.0%

The current workbook returns three headlines containing the exact expression "cybercrime." The earlier manuscript draft reported six; the workbook has been treated as the current empirical source and the value has therefore been corrected to three. This discrepancy should be noted in the project's data-cleaning record.

5.5 Temporal Headline Patterns

Table-6: Temporal Headline Patterns

Year	AI scam	AI fraud	Deepfake scam	Voice cloning	Digital arrest
2023	6	1	3	1	0
2024	7	0	1	3	0
2025	12	0	4	0	0
2026	10	0	2	1	1

The headline evidence indicates that "AI scam" occurs in all four years and reaches its highest annual count in 2025. "Deepfake scam" is visible throughout the period and becomes more prominent in the later years. "Digital arrest" appears in 2026, while voice-cloning terminology appears in 2023–2024 and again in 2026. These are descriptive patterns and should not be interpreted as statistically demonstrated changes without normalized full-text analysis.

5.6 Preliminary Headline Framing Indicators

Table-7: Preliminary Headline Framing Indicators

Indicator	Articles	Percentage
Crime/Fraud	234	100.0%
AI/Technology	234	100.0%
Victim/Deception/Loss	60	25.6%
Prevention/Solution	43	18.4%
Institutional Response	40	17.1%
Threat/Risk/Warning	26	11.1%
Innovation/Benefit	9	3.8%

Crime/Fraud and AI/Technology occur in all 234 headlines. Because the corpus was deliberately constructed around AI- and scam-related coverage, their universal presence is partly a consequence of the sampling design and cannot independently demonstrate that Indian news media overwhelmingly frame AI negatively. More informative secondary indicators are victimization/loss, prevention/solution, institutional response, threat/risk and innovation/benefit.

Victimization, deception or loss appears in approximately one-quarter of headlines. Institutional response is also visible through references to police, authorities, banks, investigations, arrests, warnings and technological interventions. Prevention and protection occur in 18.4% of headlines, indicating that AI is not represented exclusively as a source of danger. Explicit innovation/benefit framing is comparatively rare at 3.8%. These headline-level findings provide a basis for the planned full-text analysis but should not be treated as a substitute for contextual discourse evidence.

6. DISCUSSION

The empirical review indicates that the sampled Indian news coverage frames AI-driven fraud primarily at the intersection of technological advancement and criminal deception. The prevalence of terms such as "AI scam," "deepfake," and "deepfake scam" within headlines highlights a specific linguistic framework that renders emergent technologies intelligible as mechanisms of fraud. This finding aligns with wider scholarship indicating that news media frequently associate artificial intelligence with cybercrime, information disorders, and broader technological threats (Nguyen & Hekman, 2024).

Simultaneously, the inclusion of institutional countermeasures and prevention indicators complicates a strictly threat-oriented interpretation. News titles also spotlight police interventions, public alerts, technological detection tools, and safety protocols. This observation resonates with existing framing studies demonstrating that media discourse often portrays AI through a dual lens of risks and opportunities rather than a single monolithic narrative (Jhan & Kuriakose, 2026; Hendrickx & Van Coppenolle, 2026).

The depiction of victims and institutions carries significant weight for the subsequent stages of discourse analysis. Headlines may portray the victim as an unsuspecting target, the fraudster as a deliberate agent, AI as an enabling mechanism, or an authority as a respondent. These framing choices directly influence how legal and moral responsibility is distributed. Consequently, future agency analysis of full-text articles should investigate whether journalists attribute causality primarily to human offenders, technological systems, digital platforms, or institutional shortcomings. This approach complements recent research emphasizing threat actors, solutions, and agentic representations of AI in news narratives (Ahmad et al., 2026; Nguyen & Wischnewski, 2026).

The multilingual composition of the dataset also holds theoretical implications. Although the corpus represents Indian journalism, 92.3% of the documents are in English. Therefore, these conclusions primarily reflect English-language digital news reporting within the selected sample. Hindi and Gujarati items should be treated as supplementary descriptive data rather than evidence of comprehensive coverage across India's diverse multilingual media ecosystem.

Furthermore, the dense concentration of articles in 2025–2026 should not be automatically misconstrued as proof of a linear escalation in media attention. The later period contains a higher volume of entries, and the 2026 timeframe remains incomplete. Variations in search indexing, shifts in online archiving practices, outlet availability, and the fast-evolving AI-scam news cycle can all impact corpus distribution. Normalized full-text evaluations are therefore necessary before drawing definitive conclusions about temporal trends.

7. LIMITATIONS

Purpose-Built Dataset: The corpus is specifically compiled and does not comprise the entire population of Indian media coverage regarding AI-enabled fraud. Consequently, the findings reflect the characteristics of this specific sample.

Uneven Outlet Representation: Six publishers account for 56.8% of the dataset, and twelve account for 77.4%, introducing the possibility that outlet-specific editorial policies may skew aggregate patterns.

Linguistic Imbalance: English articles represent 92.3% of the dataset, whereas Hindi and Gujarati files are considerably smaller. Cross-linguistic generalizations are therefore invalid without expanded, comparable sub corpora.

Incomplete 2026 Observation Window: The data collection concludes on September 28, 2026, rendering the final year partial. Annual comparisons involving 2026 must remain cautious and descriptive.

Reliance on Metadata and Headlines: The current empirical findings derive from headline text and metadata because the full body text for all 234 articles was unavailable in the working file. Advanced measures such as full-text token frequencies, keyword metrics, collocations, concordances, and grammatical agency statistics cannot be asserted until the full texts are processed.

Methodological Dependencies: Corpus-driven conclusions rely heavily on initial search parameters, preprocessing steps, stop lists, lemmatization choices, reference corpus selection, statistical thresholds, and coding criteria. Documenting these choices is essential for maintaining reproducibility.

Exclusion of Audience Reception: This study examines media texts rather than audience interpretation. It cannot determine how readers perceive or react to AI-enabled scams. Complementary audience research via surveys, interviews, or platform analytics would provide valuable context.

Dynamic Technology Landscape: The AI and fraud environment shifts rapidly. These findings characterize the specified study period and should not be viewed as a permanent representation of AI-driven crime.

8. CONCLUSION

This research establishes a corpus-assisted framework to investigate how Indian print and digital journalism construct AI-facilitated scams as technological, criminal, communicative, and social hazards. The compiled dataset includes 234 articles sourced from 37 media outlets between July 2023 and September 2026, with English-language reporting serving as the primary analytical corpus.

The headline analysis demonstrates that AI-enabled scams are predominantly framed through the terminology of artificial intelligence, fraud, deepfakes, and criminal deception. Victimization, institutional interventions, risk management, and prevention form recurring secondary dimensions. Crucially, the headline data do not support a simplistic binary where AI is portrayed solely as a hazard; protective technologies and preventative measures are also visible. Nevertheless, the corpus-selection criteria mean that the universal presence of AI and fraud indicators cannot independently confirm a purely negative overarching media narrative.

The primary methodological contribution of this work lies in linking quantitative corpus evidence with qualitative discourse interpretation. While frequency and phraseology highlight recurrent linguistic patterns, concordance checks and qualitative evaluation reveal how actors, responsibility, victimhood, and technological agency are discursively built in context. This methodology paves the way for a deeper understanding of emerging AI-associated risks within Indian journalism.

Ultimately, this manuscript serves as an empirically grounded corpus audit and headline-level discourse analysis, awaiting the integration of the full article texts. Once all 234 full-text files are incorporated, the study can expand these initial findings through normalized token frequencies, reference-based keyword analyses, statistically rigorous collocations, clusters,

concordances, and systematic framing and agency coding. Such expanded analysis will establish a robust empirical foundation for answering the core research questions concerning the linguistic formulation of AI-enabled deception in the press.

REFERENCES

[1] Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. Expert Systems with Applications, 252, 124260. https://doi.org/10.1016/j.eswa.2024.124260

[2] Ahmad, T., Arif, R., Ittefaq, M., & Park, S. (2026). Artificial intelligence, deepfakes, and wars: How global news media frame actors, threats, and solutions in 19 countries. Media, War & Conflict. https://doi.org/10.1177/17506352261424826

[3] Ahmed, S., Wang, Y., & Bee, A. W. T. (2023). Adjusting news accuracy perceptions after deepfakes exposure: Evidence from a non-Western context. Telematics and Informatics, 84, 102047.

[4] Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. Discourse & Society, 19(3), 273–306. https://doi.org/10.1177/0957926508088962

[5] Boudana, S., & Segev, E. (2026). Fake news makes the news: Definitions and framing of fake news in mainstream media. Journalism Practice, 20(2), 679–698. https://doi.org/10.1080/17512786.2024.2379898

[6] Chen, C., & Liu, R. (2024). "China-India border clash": Construction of event newsworthiness and discrepancies in Chinese and Indian news discourse. SAGE Open, 14(4). https://doi.org/10.1177/21582440241301472

[7] De Rancourt-Raymond, A., & Smaili, N. (2023). The unethical use of deepfakes. Journal of Financial Crime, 30(4), 1066–1077. https://doi.org/10.1108/JFC-04-2022-0090

[8] Gosse, C., & Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. Critical Studies in Media Communication, 37(5), 497–511. https://doi.org/10.1080/15295036.2020.1832697

[9] Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. Crime, Media, Culture, 17(2), 209–230. https://doi.org/10.1177/1741659020917434

[10] Hendrickx, J., & Van Coppenolle, I. (2026). A tool, a threat, or both? How quality news media frame generative AI. Emerging Media, 4(2), 292–307. https://doi.org/10.1177/27523543261422926

[11] Ittefaq, M., Zain, A., Arif, R., Ala-Uddin, M., Ahmad, T., & Iqbal, A. (2025). Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and trends across 12 countries. Telematics and Informatics, 96, 102223. https://doi.org/10.1016/j.tele.2024.102223

[12] Jhan, P. J., & Kuriakose, R. (2026). How is artificial intelligence framed in online news media? A data-driven analysis of news trends, sentiment, and media framing. Quality & Quantity, 60(5), 15345–15364. https://doi.org/10.1007/s11135-026-02783-1

[13] Jiang, Y., Wen, X., & Qian, X. (2025). Impact of gain-loss framing on online scam susceptibility: The role of scam frames, warning frames, and risk perception. Behaviour & Information Technology, 44(9), 1846–1859. https://doi.org/10.1080/0144929X.2024.2378883

[14] King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. Science and Engineering Ethics, 26, 89–120. https://doi.org/10.1007/s11948-018-00081-0

[15] Kumar, A. (2022). Fact-checking methodology and its transparency: What Indian fact-checking websites have to say? Journalism Practice, 18(6), 1461–1480. https://doi.org/10.1080/17512786.2022.2098520

[16] Kumar, N., Kumar, A., Redhu, P., & Kumar, K. (2026). Prime time international news on Indian television: Frames, valence, and geopolitical visibility. Journalism. Advance online publication. https://doi.org/10.1177/14648849261447545

[17] Nguyen, D., & Hekman, E. (2024). The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. AI & Society, 39(2), 437–451. https://doi.org/10.1007/s00146-022-01511-1

[18] Nguyen, D., & Wischnewski, M. (2026). Generative AI, large language models, and their agentic framing in news media. AI & Society. https://doi.org/10.1007/s00146-026-03070-1

[19] Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: A systematic review. Crime Science, 13, Article 41. https://doi.org/10.1186/s40163-024-00239-1

[20] Tandoc, E. C., Seet, S., Chan, V. X., & Wong, P. J. O. (2025). Exploring AI identity: The media framing of communicative artificial intelligence in Singapore's news sites. Public Understanding of Science, 34(7), 852–867. https://doi.org/10.1177/09636625251317970

[21] Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing fearful futures: Journalistic discourses on deepfakes. Journalism Practice, 15(6), 803–820. https://doi.org/10.1080/17512786.2021.1908838

[22] Verma, S., Yadav, G., & Ahmed, Y. (2026). Constructing organised crime in India: Media framing, criminal celebrity, and the Lawrence Bishnoi Gang. Communication Research and Practice. Advance online publication. https://doi.org/10.1080/22041451.2026.2708573