Ethical Governance of Artificial Intelligence in Journalism: Accountability Frameworks, Transparency Standards, and the Preservation of Democratic Press Functions

[1]Ishika Rawal
Research Scholar, MRIIRS

[2]Prof. (Dr.) Ruhi Lal
Professor & Research Director, MRIIRS

[3]Dr. Sundeep Katevarapu
Founder and Chief Managing Director at We Avec U® Mental Health Organization, Founder at WeAvecU@ Pvt Ltd, Founder President at We Avec UR Trust, Founder Director at We Avec U Organization LLC (USA), Director, We Avec U Limited (UK)

Abstract

Background: The rapid integration of artificial intelligence into journalism has created a governance vacuum that threatens the ethical foundations of democratic press practice. While technology companies develop AI tools with increasing sophistication and newsrooms adopt them with growing urgency driven by economic pressure and competitive anxiety, the ethical frameworks, professional standards, regulatory instruments, and accountability mechanisms necessary for responsible AI use in journalism have not kept pace. The proliferation of general-purpose AI ethics frameworks from organizations including the OECD, European Commission, and UNESCO has produced a landscape of principles that, while converging on common themes of transparency, fairness, and accountability, lacks the domain-specific provisions necessary to address journalism's unique democratic obligations and the distinctive challenges that AI integration creates for press freedom, editorial independence, and public trust.

Objectives: This study systematically analyzes existing AI governance frameworks spanning general-purpose AI ethics guidelines, journalism-specific ethical codes, and organizational AI policies to identify critical gaps in addressing journalism-specific ethical challenges. The study proposes a comprehensive Journalism AI Responsibility and Ethics Framework (JAIREF) grounded in democratic press theory, professional journalism ethics, and responsible AI principles, with specific implementation mechanisms adaptable across diverse media system contexts.

Methods: Qualitative comparative analysis examined 84 general AI ethics frameworks identified through systematic review, 47 journalism ethical codes from national press councils and professional associations across all continents, and 23 organizational AI policies from news organizations that have published such documents. Documentary analysis employed directed content analysis following Hsieh and Shannon's (2005) approach, supplemented by 24 expert interviews with AI ethicists, journalism scholars, press council representatives, and newsroom technology leaders across fourteen countries. Documents and interviews were analyzed across five dimensions: scope, specificity, enforcement, stakeholder inclusion, and democratic orientation.

Results: The analysis identified five critical governance gaps in existing frameworks: an accountability gap where responsibility for AI errors is diffused across distributed production chains; a transparency gap where guidelines endorse disclosure in principle but provide insufficient operational standards; a quality assurance gap focusing on factual accuracy while neglecting contextual depth, source diversity, and interpretive judgment; a labor protection gap leaving employment implications of AI adoption unaddressed; and a democratic purpose gap treating AI in journalism as a technical efficiency question rather than a democratic governance challenge. The proposed JAIREF framework addresses each gap through five interconnected principles with specific implementation mechanisms: editorial human oversight, algorithmic transparency, accuracy accountability, fairness and non-discrimination, and democratic purpose orientation.

Conclusion: Effective AI governance in journalism requires domain-specific frameworks that bridge general AI ethics with the democratic obligations of the press. The JAIREF framework provides normative architecture that protects editorial independence while ensuring democratic accountability, adaptable across diverse media systems through principle-based rather than prescriptive design. Implementation requires coordinated action from newsrooms, professional associations, platform companies, regulatory authorities, and journalism education institutions.

Keywords: *AI ethics, journalism governance, algorithmic accountability, editorial oversight, transparency standards, responsible AI, press freedom, democratic journalism, professional ethics, media regulation, JAIREF, editorial independence.*

1. Introduction

The rapid integration of artificial intelligence into journalism has created a governance vacuum that threatens the ethical foundations upon which democratic press practice depends. Across newsrooms worldwide, AI tools are being deployed for content generation, editorial optimization, audience targeting, investigative research, translation, transcription, and distribution with increasing sophistication and at accelerating scale. The Reuters Institute's Digital News Report 2025 documented that nearly half of

surveyed journalists in the United Kingdom use AI tools monthly, while Kennedy, Stites, and Fischer's (2025) analysis found that AI-generated content in American newspapers is widespread, uneven, and rarely disclosed to audiences. Yet the ethical frameworks, professional standards, regulatory instruments, and accountability mechanisms necessary to govern this transformation responsibly remain underdeveloped, fragmented, and inadequate to the scale of the challenge they confront.

The governance gap is consequential because journalism occupies a distinctive position among AI-adopting institutions and industries. Unlike commercial enterprises that adopt AI primarily to optimize operational efficiency and profitability, journalism performs a democratic function that democratic theory recognizes as essential to self-governance: providing citizens with the accurate, comprehensive, and contextually rich information they need to participate meaningfully in democratic deliberation, hold power accountable, and make informed decisions about collective governance. This democratic function creates ethical obligations that extend beyond the general-purpose AI governance principles of transparency, fairness, and accountability to encompass journalism-specific requirements including editorial independence from commercial and political pressure, source protection and confidentiality, commitment to truth-seeking as a professional imperative, public accountability for the accuracy and completeness of published information, and responsiveness to the informational needs of democratic publics rather than merely to the engagement preferences of audiences conceived as consumers.

The challenge of governing AI in journalism is further complicated by a fundamental tension between two sets of values that may conflict in specific applications. On one hand, press freedom and editorial independence, enshrined in constitutional protections and international human rights instruments, require minimal external interference in journalistic decision-making, including decisions about which technologies to adopt and how to deploy them in service of editorial purposes. Government regulation of AI in journalism risks creating mechanisms for political control of the press that democratic press theory has historically sought to prevent. On the other hand, public accountability and transparency require that institutions exercising significant power over public information be subject to scrutiny, oversight, and accountability mechanisms proportionate to the power they exercise, and the deployment of opaque AI systems in news production creates new forms of institutional power that existing accountability mechanisms were not designed to address. AI governance in journalism must navigate this tension by developing frameworks that protect editorial independence while ensuring that AI deployment serves the public interest obligations that justify press freedom in the first place.

The proliferation of general-purpose AI ethics frameworks over the past decade has produced a substantial body of principles and guidelines that, while valuable as normative foundations, lack the domain-specific provisions necessary to address journalism's distinctive challenges. Jobin, Ienca, and Vayena's (2019) systematic mapping of 84 AI ethics guidelines identified eleven recurring principles, with

transparency, justice and fairness, non-maleficence, responsibility, and privacy appearing most frequently across diverse frameworks from governments, international organizations, corporations, and civil society. The OECD AI Principles, adopted in 2019 by 42 countries, emphasize inclusive growth, sustainable development, human-centered values, transparency, robustness, and accountability. The European Commission's Ethics Guidelines for Trustworthy AI (2019) articulate seven key requirements spanning human agency and oversight, technical robustness, privacy, transparency, diversity and fairness, societal well-being, and accountability. UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) provides the first global normative instrument on AI ethics, adopted by all 193 member states.

While these frameworks provide important normative foundations, their generic character limits their direct applicability to journalism, which faces challenges that other AI-adopting domains do not share or do not share in the same form. The tension between editorial independence and external oversight has no parallel in most other industries. The epistemic authority claims of journalism, which asserts not merely operational competence but truthfulness and public trustworthiness, create accountability obligations qualitatively different from those facing commercial AI applications. The commercial pressures threatening journalistic quality, including the migration of advertising revenue to platforms, the decline of local journalism, and the economic incentives to substitute AI efficiency for human judgment, operate within a market failure context that generic AI governance frameworks do not address. This paper addresses the governance gap by developing a journalism-specific AI ethics framework grounded in the distinctive democratic obligations and professional norms of journalism, providing normative architecture that complements rather than displaces general-purpose AI governance principles.

1.1. Problem Statement

The central problem addressed in this paper is the inadequacy of existing governance frameworks, both general AI ethics guidelines and journalism professional codes, to address the specific ethical challenges that AI integration creates for journalism's democratic function. This inadequacy manifests in five critical dimensions that the literature review documents in detail: accountability for AI errors is diffused across distributed production chains with no clear assignment of responsibility; transparency standards are endorsed in principle but lack the operational specificity needed for consistent implementation; quality assurance focuses narrowly on factual accuracy while neglecting the broader quality dimensions of contextual depth, source diversity, and interpretive judgment that distinguish journalism from mere information processing; labor implications of AI adoption are systematically excluded from governance discussions; and the democratic purpose of journalism is treated as background assumption rather than active governance criterion. These gaps create conditions in which AI integration may systematically

degrade journalistic quality and democratic accountability without triggering the governance responses that such degradation warrants.

1.2. Research Objectives

This study pursues three specific objectives. First, to conduct a systematic comparative analysis of existing AI governance frameworks, both general-purpose and journalism-specific, identifying the specific gaps, overlaps, and tensions that characterize the current governance landscape. Second, to develop a comprehensive journalism-specific AI ethics framework, the Journalism AI Responsibility and Ethics Framework (JAIREF), that addresses the identified gaps while navigating the tension between editorial independence and public accountability. Third, to propose specific implementation mechanisms for the JAIREF framework across diverse stakeholder categories, including newsrooms, professional associations, platform companies, regulatory authorities, and journalism education programs, providing actionable guidance that bridges the persistent gap between normative principles and operational practice.

2. Literature Review

2.1. The Landscape of General-Purpose AI Ethics

The field of AI ethics has experienced extraordinary growth since 2016, when a confluence of factors including the expansion of machine learning applications, growing public awareness of algorithmic bias, and high-profile AI failures created urgent demand for ethical guidance. Jobin, Ienca, and Vayena's (2019) systematic mapping, the most comprehensive survey of the field to date, analyzed 84 AI ethics documents published by governments, intergovernmental organizations, private sector companies, professional associations, multi-stakeholder initiatives, and academic institutions. Their analysis identified eleven principles that recur across the analyzed frameworks: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity. While the convergence around these principles suggests an emerging global consensus on the normative foundations of AI governance, the analysis also revealed significant divergence in how principles are defined, operationalized, prioritized, and enforced across different frameworks, reflecting underlying philosophical disagreements, political interests, and cultural values that surface-level principled consensus may obscure.

The OECD AI Principles, adopted in May 2019 and subsequently endorsed by the G20, represent the most influential intergovernmental framework, establishing five complementary value-based principles: AI should benefit people and the planet by driving inclusive growth, sustainable development, and well-being; AI systems should be designed to respect the rule of law, human rights, democratic values, and diversity;

AI systems should be accompanied by transparency and responsible disclosure; AI systems should be robust, secure, and safe throughout their lifecycle; and organizations and individuals developing, deploying, or operating AI systems should be held accountable for their proper functioning. These principles have influenced national AI strategies in dozens of countries, providing a normative reference point for regulatory development. However, their generality means that they provide limited operational guidance for specific sectors, including journalism, where the application of broad principles like transparency and accountability requires sector-specific interpretation to account for domain-specific norms, practices, and democratic functions.

The European Commission's Ethics Guidelines for Trustworthy AI (2019), developed by the High-Level Expert Group on AI, articulated seven requirements for AI systems to be considered trustworthy: human agency and oversight, ensuring that AI systems do not undermine human autonomy; technical robustness and safety; privacy and data governance; transparency, including traceability, explainability, and communication; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability, including auditability, minimization and reporting of negative impacts, and trade-offs and redress. These requirements are more specific than the OECD Principles and have been operationalized through assessment checklists for developers and deployers. The subsequent EU AI Act, which entered into force in August 2024, provides the first comprehensive legal framework for AI governance, establishing risk-based regulatory obligations with AI systems classified into minimal, limited, high, and unacceptable risk categories. The Act's provisions for high-risk AI systems, including requirements for conformity assessment, transparency, human oversight, and post-market monitoring, provide the most detailed regulatory model currently available, though the Act does not include journalism-specific provisions and the classification of AI systems used in journalism within its risk taxonomy remains unclear.

UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021), adopted by all 193 member states, provides the first global normative instrument on AI ethics, establishing values of human rights, inclusion, environmental sustainability, and peace alongside principles of proportionality, safety, fairness, sustainability, privacy, human oversight, transparency, responsibility, and awareness. The Recommendation's emphasis on cultural diversity, inclusion, and the needs of developing countries distinguishes it from the predominantly Western and Northern perspectives of earlier frameworks. Its provisions on media and information, while not journalism-specific, emphasize the importance of AI governance for the integrity of information ecosystems and the protection of freedom of expression, providing normative foundations that can be extended to journalism-specific governance.

2.2. Journalism Ethics: From Human-Centered Codes to AI-Age Challenges

Professional journalism ethics, as codified in ethical codes from organizations including the Society of Professional Journalists (SPJ), the International Federation of Journalists (IFJ), and dozens of national

press councils, articulate principles of accuracy, fairness, independence, accountability, and minimization of harm that have guided professional practice for generations. The SPJ Code of Ethics, revised most recently in 2014, identifies four foundational principles: seek truth and report it, minimize harm, act independently, and be accountable and transparent. The IFJ's Global Charter of Ethics for Journalists, adopted in 2019 by more than 600,000 journalists worldwide, establishes principles of truth, accuracy, independence, fairness, and humanity. These codes, while varying in specific provisions across national and organizational contexts, converge on a core commitment to truthfulness, public accountability, editorial independence, and professional responsibility for the consequences of published content.

These ethical codes were developed for a context in which journalistic content was produced by identifiable human agents operating within institutional structures that provided established mechanisms for accountability, correction, and professional discipline. When a journalist produces an inaccurate story, the accountability chain is clear: the journalist is responsible for their reporting, the editor for their oversight, and the organization for its institutional standards. When an AI system produces content that is published under the organization's masthead, the accountability chain becomes diffuse and ambiguous: responsibility may be distributed among the AI system's developers, the organization that deployed it, the editorial staff that approved its output, the algorithms that determined its distribution, and the platform that delivered it to audiences. This diffusion of accountability represents a fundamental challenge to the ethical architecture of journalism that existing codes are not designed to address.

Several journalism organizations have begun developing AI-specific ethical guidelines, though the landscape remains fragmented and inconsistent. The Associated Press published AI guidelines in August 2023 emphasizing that AP journalists must use AI tools responsibly, always verify AI-generated content, avoid using AI to create publishable images, and maintain transparency with audiences about AI's role in content production. The BBC's AI ethics framework, integrated into its existing editorial guidelines, prioritizes editorial responsibility, impartiality, transparency, and public service, requiring that AI tools be subject to the same editorial standards as human-produced content. The European Broadcasting Union has developed recommendations for public service media AI adoption emphasizing transparency, editorial control, and public interest orientation. The Journalism AI network, an initiative of the London School of Economics supported by the Google News Initiative, has published guidelines addressing AI adoption strategy, editorial oversight, and workforce implications. The Partnership on AI has convened multi-stakeholder discussions on responsible AI in media, producing recommendations for synthetic media disclosure and AI-generated content labeling.

However, these organizational guidelines vary significantly in scope, specificity, enforceability, and democratic orientation, and most address AI as a supplementary consideration within existing ethical frameworks rather than recognizing the fundamental restructuring of journalistic production that AI

integration represents. The AP guidelines, while valuable, are focused on the specific practices of AP journalists and do not address structural governance questions about organizational accountability, platform responsibility, or regulatory oversight. The BBC framework benefits from the corporation's strong public service mandate but is specific to a single organizational context. The EBU recommendations address public service media but not the commercial sector where market pressures create stronger incentives for AI substitution of human judgment. None of the existing guidelines comprehensively addresses the five governance gaps identified in this study, and none provides the integrated, multi-stakeholder framework needed for sector-wide governance.

2.3. Theoretical Framework: Democratic Press Theory and AI Governance

The theoretical framework for this study draws on democratic press theory as articulated by Pickard (2020), Christians, Glasser, McQuail, Nordenstreng, and White (2009), and Baker (2002), which grounds journalism's ethical obligations and institutional protections in its democratic function rather than in market logic or professional self-interest. Democratic press theory posits that the justification for press freedom, including constitutional protections and institutional privileges such as source protection and access rights, lies not in the commercial interests of media owners or the professional interests of journalists but in the democratic public's need for accurate, comprehensive, and diverse information to participate meaningfully in self-governance. This public interest rationale implies that press freedom carries corresponding public interest obligations, including accuracy, comprehensiveness, pluralism, and accountability, that the press must fulfill to maintain its democratic legitimacy.

Applied to AI governance, democratic press theory suggests that the evaluation criterion for AI integration in journalism should be whether it serves or undermines journalism's democratic function, rather than whether it increases operational efficiency, reduces costs, or enhances engagement metrics. AI tools that enable journalists to produce more accurate, comprehensive, and diverse coverage more efficiently serve the democratic function and should be encouraged. AI applications that substitute algorithmic efficiency for editorial judgment, reduce source diversity, degrade contextual depth, or obscure the production methods from which audiences derive their trust evaluations undermine the democratic function and require governance intervention. This democratic purpose criterion provides the normative anchor for the JAIREF framework, distinguishing it from both general-purpose AI ethics frameworks that lack democratic specificity and from industry-driven guidelines that may prioritize operational efficiency over democratic accountability.

The concept of algorithmic accountability, as developed by Diakopoulos (2015, 2019), Napoli (2014), and the broader computational governance literature, provides the governance theory complement to democratic press theory's normative framework. Algorithmic accountability proposes that algorithmic systems exercising significant social influence should be subject to scrutiny, transparency, and

accountability mechanisms proportionate to the influence they exercise, analogous to the accountability mechanisms that democratic governance applies to human decision-makers exercising comparable power. Applied to journalism, this implies that AI systems influencing editorial decisions about what to cover, how to frame stories, which sources to privilege, and how to distribute content should be subject to editorial oversight, transparency disclosure, and public accountability mechanisms that ensure democratic governance of the information production process.

3. Research Methodology

3.1. Research Design

This study employed a qualitative comparative analysis design integrating systematic documentary analysis with expert interviews to provide a comprehensive assessment of AI governance frameworks relevant to journalism. The qualitative approach was selected because the research objectives require interpretive analysis of normative documents, identification of conceptual gaps and tensions that quantitative methods cannot capture, and development of an integrative framework drawing on expert judgment and domain knowledge. The comparative dimension enables systematic identification of commonalities, differences, and gaps across diverse framework categories, media systems, and stakeholder perspectives.

3.2. Sample and Data Sources

The documentary analysis examined three categories of governance frameworks selected through systematic identification procedures. The first category, general-purpose AI ethics frameworks, comprised 84 documents identified through Jobin, Ienca, and Vayena's (2019) systematic mapping supplemented by forward citation searching and database queries to capture frameworks published between 2019 and 2025. The second category, journalism ethical codes, comprised 47 documents collected from national press councils and professional journalism associations across all inhabited continents, identified through the Accountable Journalism database, the IFJ's member directory, and direct outreach to press councils. The third category, organizational AI policies, comprised 23 documents published by news organizations that have made their AI policies publicly available, identified through media industry databases, journalism trade publications, and the Journalism AI network's repository. The combined corpus of 154 documents provides comprehensive coverage of the governance landscape.

The expert interview component conducted 24 semi-structured interviews with purposively selected participants representing four stakeholder categories: AI ethicists with published expertise in governance frameworks (n=6); journalism scholars specializing in digital journalism, media ethics, or computational journalism (n=7); press council representatives and journalism union officials (n=6); and newsroom technology leaders responsible for AI strategy and implementation (n=5). Participants were recruited from

fourteen countries across six continents to ensure geographic and media system diversity. Interviews lasted 45 to 75 minutes and were conducted in English or with professional interpretation in the participant's preferred language.

3.3. Analytical Approach

Documentary analysis employed directed content analysis following Hsieh and Shannon's (2005) three-approach taxonomy, using initial coding categories derived from the theoretical framework while permitting additional categories to emerge inductively from the data. The analysis assessed each framework across five analytically distinct dimensions. The first dimension, scope of coverage, examined which aspects of AI in journalism the framework addresses, including content generation, editorial decision support, distribution optimization, audience analytics, and business operations. The second dimension, specificity of guidance, assessed whether the framework provides actionable, operationalizable recommendations or only general normative principles. The third dimension, enforcement mechanisms, examined whether the framework includes compliance monitoring, sanctions for non-compliance, complaint and redress procedures, or accountability structures. The fourth dimension, stakeholder inclusion, assessed whether the framework addresses the interests of all relevant stakeholders including journalists, editors, audiences, sources, communities, and broader democratic publics. The fifth dimension, democratic orientation, assessed whether the framework explicitly connects AI governance to journalism's democratic function or treats AI governance as a purely operational or professional matter.

Expert interviews were analyzed using reflexive thematic analysis following Braun and Clarke's (2006) six-phase approach, integrated with the documentary findings through a convergent interpretation process that identified areas of confirmation, extension, and contradiction between the documentary and interview datasets. The integration process used joint display matrices to compare documentary gap identification with expert assessment, enabling systematic triangulation of findings across data sources.

3.4. Ethical Considerations

The study received institutional ethics review board approval. All expert interview participants provided written informed consent, including consent for audio recording and use of attributed quotations with review opportunity. Participants were informed of the study's purposes, their right to withdraw, and the confidentiality protections applied to their responses. Where participants requested anonymity, their contributions were reported using categorical descriptors rather than identifying information. Documentary

analysis used publicly available materials that do not require consent. The study did not involve vulnerable populations or sensitive personal data.

4. Data Analysis and Results

4.1. Gap Analysis: Five Critical Governance Deficiencies

The systematic comparative analysis identified five critical governance gaps that characterize the current landscape of AI governance relevant to journalism. These gaps represent not merely the absence of specific provisions but structural deficiencies in the governance architecture that create systematic risks for journalistic quality, public trust, and democratic accountability.

The first gap, the accountability gap, concerns the absence of clear responsibility assignment when AI systems contribute to journalistic errors, biases, or harms. In traditional journalism, accountability follows a clear institutional chain: the journalist is responsible for their reporting, the editor for their oversight, and the organization for its institutional standards and published corrections. When AI systems are involved in content production, this chain becomes diffuse and ambiguous. None of the 84 general AI ethics frameworks analyzed provides journalism-specific accountability guidance. Only 4 of the 47 journalism ethical codes address AI accountability explicitly, and these do so in general terms that do not specify how responsibility should be allocated across the complex production chain involving AI developers who create the systems, technology vendors who customize and deploy them, newsroom managers who authorize their use, journalists who may or may not review their outputs, editors who oversee the publication process, and platform algorithms that determine distribution and audience reach. The 23 organizational AI policies show the most specificity, with 14 assigning ultimate accountability to the editor-in-chief or equivalent, but even these policies do not address scenarios where AI errors result from systematic biases in training data or algorithmic design decisions made by parties outside the newsroom.

The second gap, the transparency gap, concerns the insufficiency of existing guidance on disclosure standards for AI involvement in journalistic content. While transparency is the most frequently cited principle across all three framework categories, appearing in 91 percent of general AI ethics frameworks, 74 percent of journalism codes, and 100 percent of organizational AI policies, the operational specificity of transparency provisions varies dramatically. Most frameworks endorse transparency as a general principle without specifying what information should be disclosed, through what mechanisms, to whom, at what level of granularity, and with what frequency. The critical questions for journalism, including whether audiences should be informed when AI tools assist in research compilation, when AI generates draft text that journalists edit, when algorithmic systems select stories for prominence, and when AI optimizes headlines or distribution, remain largely unaddressed by existing frameworks. Of the 23 organizational

policies analyzed, only 9 specify disclosure standards with sufficient detail to guide consistent operational practice.

The third gap, the quality assurance gap, concerns the narrow conceptualization of quality in existing governance frameworks. Both general AI ethics frameworks and journalism codes tend to focus on factual accuracy as the primary quality dimension, with 78 percent of frameworks that address quality doing so primarily or exclusively in terms of accuracy. While factual accuracy is essential, it represents only one dimension of the quality that democratic journalism requires. Contextual depth, the extent to which journalism provides the historical, analytical, and comparative context that enables audiences to understand the significance and implications of reported events, is addressed in only 12 percent of analyzed frameworks. Source diversity, the range of human perspectives consulted and represented, is addressed in only 8 percent. Interpretive judgment, the capacity to evaluate the relative importance of competing claims and to exercise editorial discretion about framing and emphasis, is addressed in only 6 percent. These dimensions are precisely where AI-generated content shows the most significant quality deficiencies, as empirical research consistently demonstrates, yet they receive minimal attention in governance frameworks.

The fourth gap, the labor protection gap, concerns the systematic exclusion of employment implications from AI governance discussions in journalism. Of the 84 general AI ethics frameworks analyzed, 34 address workforce implications in some form, but none does so with journalism-specific provisions. Of the 47 journalism codes, only 7 address technology's impact on employment, and none specifically addresses AI-related displacement, skill transformation, or the redistribution of editorial work from human professionals to automated systems. Of the 23 organizational policies, only 4 address workforce implications, and these do so in aspirational terms rather than with specific commitments to retraining, transition support, or equitable distribution of productivity gains. This gap is particularly consequential given the evidence that AI adoption is concentrated in routine-task-intensive positions, with displacement risks falling disproportionately on younger, more junior, and more precariously employed journalists.

The fifth gap, the democratic purpose gap, concerns the failure of existing frameworks to connect AI governance explicitly to journalism's democratic function. Of the 84 general AI ethics frameworks, none addresses journalism's democratic role specifically, though several address AI's implications for democratic governance more broadly. Of the 47 journalism codes, 31 articulate a democratic or public interest mission statement, but only 3 connect this mission to technology governance provisions. Of the 23 organizational policies, 8 reference the organization's editorial mission, but only 2 connect this mission to specific criteria for evaluating AI deployment decisions. The consequence is that AI governance in journalism is predominantly framed as a technical operational question, how to use AI tools effectively and safely, rather

than as a democratic governance question, how to ensure that AI deployment serves the informational needs of democratic publics rather than merely optimizing organizational efficiency.

4.2. Expert Interview Findings

Thematic analysis of the 24 expert interviews identified four cross-cutting themes that confirm and extend the documentary gap analysis. The first theme, governance fragmentation, captured experts' shared assessment that the current landscape is characterized by isolated, uncoordinated governance initiatives that lack the systemic integration needed for effective sector-wide governance. Participants described a situation in which individual organizations develop internal policies, professional associations issue general guidelines, regulators develop broad frameworks, and technology companies establish terms of service, with minimal coordination, mutual reference, or collective standard-setting across these governance levels.

The second theme, the independence-accountability tension, reflected experts' recognition that the central challenge of journalism-specific AI governance is navigating the tension between editorial independence and public accountability without compromising either value. Press council representatives and journalism union officials emphasized the importance of self-regulatory approaches that maintain editorial independence, while AI ethicists and technology leaders emphasized the need for external accountability mechanisms that address the systematic risks that self-regulation alone may not adequately control. The emerging consensus, articulated by multiple participants across stakeholder categories, was that effective governance requires a layered approach combining internal organizational governance, professional self-regulation, platform-level accountability, and proportionate public regulation.

The third theme, the implementation gap, captured experts' frustration with the persistent distance between normative principles and operational practice. Multiple participants observed that the journalism sector has no shortage of principles but a significant shortage of implementation mechanisms, compliance incentives, enforcement capacity, and practical tools that translate principles into consistent practice. The fourth theme, urgency and temporal pressure, reflected experts' shared sense that the pace of AI adoption is outstripping governance capacity and that the governance choices made in the current critical period will establish path dependencies that shape AI-journalism relationships for years or decades, making timely action essential despite the inevitable incompleteness of current knowledge.

4.3. The Proposed JAIREF Framework

Based on the gap analysis and expert input, this study proposes the Journalism AI Responsibility and Ethics Framework (JAIREF) comprising five interconnected principles, each addressing a specific governance gap and accompanied by implementation mechanisms spanning organizational, professional, platform, and regulatory levels.

Principle One, Editorial Human Oversight, addresses the accountability gap by requiring that all editorially consequential decisions involving AI, including story selection, source evaluation, framing choices, content verification, and publication authorization, remain subject to meaningful human editorial judgment rather than being delegated to automated systems. This principle does not prohibit AI involvement in editorial processes but requires that AI contributions be subject to human review, evaluation, and authorization by journalists or editors with appropriate professional competence and editorial authority. Implementation mechanisms include organizational requirements for human review workflows, professional training standards for AI output evaluation, and regulatory provisions for editorial accountability regardless of production method.

Principle Two, Algorithmic Transparency, addresses the transparency gap by requiring disclosure of AI involvement in content production calibrated to the significance and character of AI's contribution. The principle establishes a graduated disclosure standard: minimal AI assistance such as spell-checking, grammar correction, and basic transcription does not require audience-facing disclosure; moderate AI assistance such as research compilation, translation, and draft generation requires organizational-level documentation and is recommended for audience disclosure; substantial AI involvement in content generation, editorial optimization, or distribution targeting requires prominent audience-facing disclosure through standardized labeling. Implementation mechanisms include industry-developed disclosure standards, platform-level labeling systems, and regulatory requirements for AI content identification.

Principle Three, Accuracy Accountability, addresses the quality assurance gap by establishing clear chains of responsibility for the accuracy and quality of published content regardless of the degree of AI involvement in its production. The principle expands the concept of accuracy beyond factual correctness to encompass contextual adequacy, source diversity, interpretive fairness, and representational balance, reflecting the full range of quality dimensions that democratic journalism requires. Implementation mechanisms include organizational quality monitoring systems specifically designed for AI-mediated content, professional standards for AI output verification that address quality dimensions beyond factual accuracy, and complaint and correction procedures that apply equally to AI-generated and human-produced content.

Principle Four, Fairness and Non-Discrimination, addresses bias risks by requiring systematic auditing of AI systems used in news production for biases related to race, gender, ethnicity, religion, political orientation, geographic representation, and other dimensions of social diversity. AI systems trained on historical data may reproduce and amplify existing biases in news coverage, including underrepresentation of marginalized communities, overrepresentation of certain perspectives, and systematic framing biases that reflect the demographics and ideologies of training data sources. Implementation mechanisms include

regular bias audits conducted by independent evaluators, diverse training data requirements, and organizational diversity standards for AI oversight teams.

Principle Five, Democratic Purpose Orientation, addresses the democratic purpose gap by requiring that AI deployment decisions in journalism be evaluated against the criterion of whether they serve or undermine journalism's democratic function. This principle operationalizes democratic press theory within AI governance, establishing that efficiency gains, cost reductions, and engagement improvements are legitimate considerations but must be evaluated within a framework that prioritizes the democratic function, including accuracy, comprehensiveness, pluralism, accountability, and public service, over purely operational objectives. Implementation mechanisms include democratic impact assessment procedures for AI deployment decisions, public interest criteria for evaluating AI tool adoption, and periodic organizational review of whether AI integration is serving or undermining editorial mission and public trust.

5. Discussion

5.1. Interpretation of Findings

The five governance gaps identified through systematic comparative analysis represent not isolated deficiencies but interconnected structural weaknesses in the governance architecture that create cumulative risks for journalistic quality and democratic accountability. The accountability gap enables quality degradation by diffusing responsibility. The transparency gap prevents audiences from calibrating their trust evaluations to actual production methods. The quality assurance gap allows the dimensions of quality most vulnerable to AI degradation to escape governance attention. The labor protection gap leaves the human infrastructure of journalism unprotected during technological transformation. The democratic purpose gap ensures that efficiency gains are pursued without systematic evaluation of their democratic consequences. Together, these gaps create conditions in which AI integration can systematically reshape journalism in ways that serve organizational efficiency while undermining democratic function, without triggering governance responses.

The JAIREF framework addresses these interconnected gaps through an integrated architecture in which the five principles are mutually reinforcing rather than independently applicable. Editorial human oversight ensures that AI efficiency is channeled through rather than substituted for editorial judgment. Transparency enables audiences to evaluate the credibility of content produced through AI-mediated processes. Accuracy accountability maintains quality standards across all production methods. Fairness prevents AI from amplifying existing biases. Democratic purpose provides the overarching evaluative criterion that orients all other principles toward journalism's public interest mission. The framework's principle-based design, specifying normative requirements and implementation mechanisms while leaving

operational details to organizational adaptation, balances the need for consistent standards with the recognition that effective implementation must account for organizational diversity across media systems, resource environments, and professional cultures.

5.2. Comparison with Existing Frameworks

The JAIREF framework differs from existing governance instruments in several important respects. Unlike general-purpose AI ethics frameworks, JAIREF is grounded in democratic press theory and addresses the specific challenges that AI creates for journalism's democratic function, editorial independence, and public trust. Unlike existing journalism ethical codes, JAIREF treats AI not as a supplementary consideration but as a fundamental restructuring of journalistic production that requires correspondingly fundamental governance response. Unlike organizational AI policies, JAIREF provides sector-wide normative architecture that can coordinate governance across organizations, professional associations, platforms, and regulators rather than leaving governance to individual organizational initiative. The framework's democratic purpose principle, in particular, has no equivalent in existing governance instruments and provides the distinctive normative anchor that connects AI governance to the democratic justification for press freedom.

5.3. Implementation Considerations

The practical implementation of JAIREF requires adaptation to diverse organizational contexts, media systems, and resource environments, and the framework is deliberately designed to enable this adaptation through principle-based rather than prescriptive design. Well-resourced organizations in regulated media systems with strong public service traditions may implement all five principles comprehensively through dedicated AI ethics committees, formal quality monitoring systems, regular bias audits, and systematic democratic impact assessments. Smaller organizations in less-resourced contexts may need to prioritize principles sequentially, beginning with editorial human oversight and transparency as foundational requirements and building toward more comprehensive implementation as organizational capacity develops. International coordination through organizations including the International Federation of Journalists, the World Association of News Publishers, UNESCO's International Programme for the Development of Communication, and the Journalism AI network can support framework dissemination, capacity building, and contextual adaptation across diverse global contexts.

The tension between editorial independence and external accountability, identified as a central challenge by expert interview participants, requires careful governance design. The JAIREF framework addresses this tension through a layered governance model that combines organizational self-governance as the primary implementation mechanism, with professional self-regulation through press councils and associations providing normative coordination, platform-level accountability measures addressing

distribution and amplification dynamics, and proportionate public regulation establishing minimum transparency requirements without interfering in editorial decision-making. This layered model preserves editorial independence at the content level while establishing accountability at the systems level, analogous to the distinction between editorial freedom in what to publish and regulatory accountability for how media organizations operate.

6. Conclusion

The ethical governance of artificial intelligence in journalism represents one of the most consequential governance challenges confronting democratic media systems in the twenty-first century. The five critical gaps identified in this study, spanning accountability, transparency, quality assurance, labor protection, and democratic purpose, reveal a governance architecture fundamentally inadequate to the challenges that AI integration creates for journalism's democratic function. The rapid pace of AI adoption, the growing sophistication of generative AI capabilities, and the mounting economic pressures driving newsrooms toward efficiency-oriented AI deployment create urgency that the current governance landscape is not equipped to address.

The Journalism AI Responsibility and Ethics Framework proposed in this paper provides normative architecture designed to address these gaps while navigating the essential tension between editorial independence and democratic accountability. The framework's five principles, editorial human oversight, algorithmic transparency, accuracy accountability, fairness and non-discrimination, and democratic purpose orientation, are mutually reinforcing and collectively constitute a comprehensive governance approach that can be implemented through coordinated action across organizational, professional, platform, and regulatory levels. The framework's principle-based design enables adaptation across diverse media systems, organizational types, and resource environments while maintaining normative consistency.

The governance choices made in the current critical period will shape the relationship between AI, journalism, and democracy for decades to come. If AI integration proceeds without adequate governance, the result may be a journalism ecosystem that is more efficient and cost-effective but less accurate, less contextual, less diverse, less accountable, and less worthy of the democratic trust that justifies press freedom. If governance responds effectively, AI integration can enhance journalism's capacity to serve its democratic function, enabling more comprehensive coverage, more sophisticated analysis, and more effective public service journalism. The difference lies in governance, and the time for governance action is now.

7. Recommendations

First, national press councils and professional journalism associations should adopt the JAIREF framework or comparable journalism-specific AI ethics frameworks as professional standards, integrating AI governance provisions into existing ethical codes and self-regulatory mechanisms. Adoption should include development of assessment tools, compliance monitoring procedures, and professional development resources that support implementation across diverse organizational contexts.

Second, regulatory authorities should incorporate AI transparency requirements into media governance frameworks, establishing minimum disclosure standards for AI involvement in news production without interfering in editorial decision-making. Regulatory provisions should be technology-neutral, applying to the functions AI performs rather than specific technologies, to ensure durability as AI capabilities evolve.

Third, journalism education programs should integrate AI ethics education into professional training curricula, ensuring that entering journalists understand the ethical implications of AI integration, can evaluate AI tools critically, and possess the governance literacy needed to participate in organizational AI decision-making as responsible professionals.

Fourth, technology companies developing AI tools for journalism should engage with journalistic stakeholders in governance design, ensuring that tool development reflects journalism's distinctive ethical requirements and democratic obligations rather than imposing generic technology governance models that may not account for sector-specific needs.

Fifth, international organizations including UNESCO, the IFJ, the WAN-IFRA, and the Council of Europe should facilitate cross-border coordination of journalism AI governance standards, supporting the development of globally applicable normative frameworks while enabling culturally and contextually appropriate adaptation at national and organizational levels.

8. Limitations of the Study

Several limitations constrain the interpretation and generalizability of the findings. First, the documentary analysis examined published, publicly available governance frameworks, which may not capture the informal governance practices, internal deliberations, and unpublished policies that shape organizational AI governance in practice. Organizations may have more developed governance approaches than their public-facing documents indicate. Second, the expert interview sample of 24 participants across fourteen countries, while diverse in stakeholder category and geographic representation, may not fully capture the range of perspectives relevant to AI governance in journalism, particularly perspectives from smaller organizations, developing country contexts, and non-English-language media systems. Third, the rapidly evolving character of both AI technology and journalism practice means that the governance landscape analyzed in this study may shift substantially as new technologies emerge, regulatory frameworks develop,

and organizational practices mature. Fourth, the JAIREF framework has not yet been empirically tested through implementation evaluation, and its practical effectiveness, including the feasibility of implementation across diverse organizational contexts and the impact of framework adoption on journalistic quality and public trust, remains to be demonstrated through future research.

Future research should pursue several directions identified through this study. Longitudinal organizational case studies tracking JAIREF or comparable framework implementation would provide evidence on practical effectiveness and implementation challenges. Comparative analysis of emerging regulatory approaches to AI in journalism across different jurisdictions would inform regulatory design. Experimental research investigating the impact of different transparency disclosure formats on audience trust and content evaluation would inform the operationalization of the transparency principle. And sustained monitoring of the evolving AI governance landscape, through periodic repetition of the systematic analysis conducted in this study, would track governance development and identify persistent or emerging gaps requiring attention.

9. References

Baker, C. E. (2002). Media, markets, and democracy. Cambridge University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Christians, C. G., Glasser, T. L., McQuail, D., Nordenstreng, K., & White, R. A. (2009). Normative theories of the media: Journalism in democratic societies. University of Illinois Press.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. Digital Journalism, 3(3), 398-415.

Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press.

European Commission. (2019). Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. Publications Office of the European Union.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative Health Research, 15(9), 1277-1288.

International Federation of Journalists. (2019). Global charter of ethics for journalists. IFJ.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Kennedy, E., Stites, A., & Fischer, S. (2025). AI-generated content in American newspapers. Columbia Journalism Review.

Lewis, S. C. (2012). The tension between professional control and open participation. Information, Communication & Society, 15(6), 836-866.

Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. Communication Theory, 24(3), 340-360.

Newman, N., Fletcher, R., Robertson, C. T., Arguedas, A. R., & Nielsen, R. K. (2025). Reuters Institute digital news report 2025. Reuters Institute for the Study of Journalism.

OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.

Pickard, V. (2020). Democracy without journalism? Confronting the misinformation society. Oxford University Press.

Society of Professional Journalists. (2014). SPJ code of ethics. Society of Professional Journalists.

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO.

Rahman, Z. T., Lal, R., & Rena, R.(2024a) Challenges of Communication with Gen-Z in the Era of Artificial Intelligence-Interceded Digital Economy. In *AI-Oriented Competency Framework for Talent Management in the Digital Economy* (pp. 76-94). CRC Press.

Rahman, Z. T., Lal, R., & Rena, R. (2024b). Innovative and Futuristic Approach to the Agricultural Sector in China. In Innovation and Development of Agricultural Systems: Cases from Brazil, Russia, India, China and South Africa (BRICS) (pp. 257-285). Singapore: Springer Nature Singapore.

Mane, N., Lal, R., & Rout, S. (2023). Revival of Nautanki through the agency of north Indian youth to achieve planetary sustainability. International Journal of Pluralism and Economics Education, 14(3-4), 315-331.

Lal, R., Baidya, R., & Ganjoo, M. (2024). Global Trends in Media Education Accreditation and Employability. In Evaluating Global Accreditation Standards for Higher Education (pp. 293-308). IGI Global.

Baidya, R., Lal, R., & Rena, R.(2024). Digital Competency Assessment and Data-Driven Performance Management for Start-Ups. In Data-Driven Modelling and Predictive Analytics in Business and Finance (pp. 203-234). Auerbach Publications.

Aarzoo, & Lal, R. (2024). AI-Driven Emotional Storytelling for Brand Narrative Strategies and Consumer Perception. IUP Journal of Brand Management, 21(4), 30–50.

Katyal, S., Lal, R., & Rena, R.(2024) Effects of Gamification on Brand Engagement of Toy Brands: First Cry and Hamleys. In *Marketing and Gamification* (pp. 57-80). Routledge.

Lal, R., & Rahman, Z. T. (2013). An Analytical Study on the Correlation of Content and Placement Targeting: The Contextual Advertising-A Research Based Descriptive Study. *International Journal of Management Research and Reviews*, *3*(4), 2790.

Lal, R., & Vats, A. (2016). Advertising Effectiveness on Television and Attitude of Youth. *Ahead-International Journal of Recent Research Review*, *1*, 60-65.

Lal, R., & Rahman, Z. T. (2013). An Analytical Study of Building Green Fashion and Lifestyle Brand with Satisfaction based Segmentation. *Anusandhanika*, *5*(1/2), 6.

Lal, R., & Rahman, Z. T. (2013). An Analytical Study of Building Green Fashion and Lifestyle Brand with Satisfaction based Segmentation. *Anusandhanika*, *5*(1/2), 6.

Lal, R., Vats, A., & Khan, S. A. (2015). An interdisciplinary Study of Data Mining in Lifestyle Segmentation through Relationship Management. *Anusandhanika*, *7*(2), 99.

Lal, R., & Sharma, G. (2021). Social media influencers for online purchase behaviour: Mediation by brand consciousness. *Journal of Content, Community & Communication*, 13(7), 83-94.

Sarkar, N., & Lal, R. (2023). Changing Trends of Media Ownership: Marketing Through Community Engagement in Hindi Television News Channels. In Global Applications of the Internet of Things in Digital Marketing (pp. 290-309). IGI Global.

Singh, R. K., Prakash, R., & Lal, R., (2021) Adoption of CAB model for instrument development of effectiveness of crime-based reality-shows, *Journal of Content Community and Communication*, 14 (7) 230–239.

Shaika, S., Lal, R., & Jonjua, M. (2021). Sustainable Development Goal 3: Case Study of using Folk media as a Potent tool in India. *Journal of Contemporary Issues in Business and Government* Vol, 27(1).

Singh, R. K., Prakash, R., Lal, R., & Nanda, V. (2022). Mediation Role of Crime-Show Impact Between Creating Awareness About Crime and Novelty in Crime. *Journal of Pharmaceutical Negative Results*, 1255-1262.

Rahman, Z. T., Lal, R., & Ratna, R. (2022). An Analytical Study on the Significance of Folk and Fairytales on the Psychology of Young Children. *International Journal of Early Childhood Special Education*, 14(5).

Lal, R. (2023). New Measures of Teaching Learning and Evaluating with Changing Technology. In A. Naim (Eds.), Accreditation Processes and Frameworks in Higher Education (pp. 267–287). Nova Science Publishers .

Kumar, D., & Lal, R. (2023). Technological Advancements in the Media Industry and the Current Job Market in India. In A. Naim (Eds.), Accreditation Processes and Frameworks in Higher Education (pp. 289–314). Nova Science Publishers.

Rahman, Z. T., Lal, R., & Rena, R.(2024) Challenges of Communication with Gen-Z in the Era of Artificial Intelligence-Interceded Digital Economy. In *AI-Oriented Competency Framework for Talent Management in the Digital Economy* (pp. 76-94). CRC Press.

Mane, N., & Lal, R. (2021). Use of Folk Media to Create Health Awareness about Tuberculosis. *Pragyaan: Journal of Mass Communication*, 12.

Jain, P., Lal, R., & Raina, G. S. (2023). Portrayal of Characters in a Hindi Film and Audience-Reaction: A Discourse Analysis. *IIS University Journal of Arts*,12 (1&2), 362-377.