World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Agentic AI in Newsrooms: Towards a multi-dimensional framework for evaluating trust, editorial accountability, and workflow quality

Samson Emeka Agbaeze [1,*], Vivian Claire Okeke [2], Ebiere Precious Phillips [3], Amara Lucy Jacobs [4] and Samuel Tobi Oluwakoya [5]

[1] *Digital Transformation, Business Administration, Nexford University, Washington, District of Columbia, US.*
[2] *Media and Communication, Afe Babalola University, Ado-Ekiti, Ekiti, Nigeria.*
[3] *Business Administration, Nexford University, Washington, District of Columbia, US.*
[4] *Mass Communication, University of Calabar, Calabar, Nigeria.*
[5] *Computer Science, Afe Babalola University, Ado-Ekiti, Ekiti, Nigeria.*

World Journal of Advanced Research and Reviews, 2025, 28(02), 1061-1080

Publication history: Received on 30 September 2025; revised on 08 November 2025; accepted on 12 November 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.2.3766

Abstract

As artificial intelligence (AI) systems evolve from assistive to *agentic* capable of autonomous planning, decision-making, and content generation existing evaluation frameworks struggle to capture their broader organizational and ethical implications. Most assessments of newsroom AI focus narrowly on technical accuracy or efficiency, overlooking how such systems reshape trust, governance, and human collaboration. This study conducts a systematic literature review of 46 peer-reviewed and institutional sources (2015–2025) to examine how AI performance in journalism can be evaluated more holistically. Drawing from Information Systems Success Theory, Socio-Technical Systems Theory, Accountability Theory, and Trust Theory, the paper proposes a Four-Dimensional (4D) Evaluation Framework encompassing *Technical Quality*, *Human–Organizational Alignment*, *Ethical–Governance Responsibility*, and *Trust–Value Impact*. The framework reconceptualizes AI success as a socio-technical equilibrium where technological capacity, ethical integrity, and collaborative trust co-evolve. It contributes to the emerging field of Responsible AI in journalism by offering a multi-dimensional structure for evaluating agentic AI systems that balances innovation with accountability and public value.

Keywords: Agentic AI; Journalism; Information Systems Evaluation; Trust; Ethics; Responsible AI; Socio-Technical Systems

1. Introduction

The rapid infusion of artificial intelligence (AI) into global newsrooms marks one of the most profound transformations in the communication industry since the rise of digital journalism. Across continents, media organizations now deploy algorithmic systems to automate content generation, personalize audience experiences, detect misinformation, and optimize newsroom workflows. While these technologies promise efficiency and scale, their increasing autonomy raises deeper questions about trust, editorial accountability, and the overall quality of journalistic work. In recent years, agentic AI systems AI tools capable of planning, decision-making, and acting with minimal human oversight have emerged as new "actors" within the newsroom ecosystem (Baird and Maruping, 2021; Dörr, 2016). Yet, despite their growing presence, there is still no coherent framework to evaluate their impact on journalism's institutional integrity and democratic purpose.

* Corresponding author: Samson Agbaeze

Existing research reflects fragmented scholarly attention. Studies in the early wave of automated journalism (2014–2020) focused on the technical accuracy of machine-generated stories and audience perceptions of credibility (Hansen et al., 2017; Dörr, 2016). More recent work, particularly from the Reuters Institute, FIAT/IFTA, and Tow Center for Digital Journalism, explores newsroom-level adoption and ethical guidelines for AI-assisted reporting (Beckett, 2019; Diakopoulos, 2019). However, these studies are largely descriptive, examining what tools are used or how journalists feel about them. They rarely interrogate *how* agentic AI systems influence the deeper socio-technical relationships that define journalistic quality, editorial responsibility, and public trust. Consequently, while the *capabilities* of AI in journalism are increasingly well understood, their *evaluation* remains narrowly defined by technical performance metrics rather than social or ethical outcomes.

This lack of holistic evaluation presents a significant research gap. Traditional Information Systems (IS) evaluation models—such as the DeLone and McLean IS Success Model—offer multidimensional perspectives on system success, including user satisfaction and organizational impact (DeLone and McLean, 2016). Yet, their application in journalism remains limited. Similarly, theories of accountability and transparency in media systems provide conceptual lenses for assessing editorial responsibility (Plaisance, 2015; Karlsson and Clerwall, 2018), but these are seldom integrated with IS or socio-technical frameworks. The absence of interdisciplinary synthesis has created an imbalance: AI systems are judged by efficiency, not by their implications for ethical governance or journalistic values. This study seeks to bridge that divide.

The aim of this research is therefore to develop a multi-dimensional, IS-grounded evaluation framework for assessing the role and impact of agentic AI in newsrooms. Specifically, it examines how such systems affect three interrelated domains audience trust, editorial accountability, and workflow quality—drawing from both Information Systems theory and Communication research. By synthesizing peer-reviewed studies, institutional reports, and conceptual works from 2015 to 2025, this paper proposes an integrated evaluative model that aligns technological, human, and organizational dimensions of newsroom AI.

This study contributes in three major ways. First, it provides a systematic synthesis of how AI has been conceptualized and applied within newsroom contexts, identifying trends, tensions, and theoretical blind spots. Second, it extends IS and communication theories to a new frontier *agentic AI in journalism* by introducing a framework that treats AI as both a technological artifact and an organizational actor. Finally, it offers a practical roadmap for news organizations, regulators, and scholars seeking to design or audit responsible AI systems that reinforce, rather than erode, the ethical foundations of journalism. In doing so, this paper aligns with global debates on algorithmic accountability, data governance, and trust restoration in digital media ecosystems.

2. Theoretical Background and Conceptual Foundations

Understanding the impact of agentic AI in newsrooms requires a theoretical base that captures both its **technical** functions and its organizational consequences. Journalism, as a socio-technical profession, has long been shaped by the interaction between human judgment, technological tools, and institutional norms. As AI becomes an increasingly autonomous collaborator in news production, these interactions intensify and transform. To assess this transformation holistically, the study draws on four theoretical foundations that, together, illuminate how AI reshapes newsroom trust, accountability, and workflow: the Information Systems (IS) Success Model, Socio-Technical Systems (STS) Theory, Accountability and Transparency Theory, and Trust Theory.

2.1. Information Systems Success Model

The DeLone and McLean IS Success Model remains one of the most influential frameworks for evaluating the performance of information systems across industries. Originally introduced in 1992 and updated in 2003 and 2016, the model proposes six interdependent dimensions of system success: system quality, information quality, service quality, user satisfaction, intention to use, and net benefits (DeLone and McLean, 2016). In newsroom contexts, these dimensions can translate into critical measures such as the reliability of AI-generated outputs, perceived usefulness by journalists, and the broader organizational value derived from AI-supported editorial workflows.

However, existing studies of AI in journalism have rarely extended the DeLone and McLean model to include **ethical** and human-centered outcomes. Technical measures like accuracy or error rate dominate evaluation, while the effects on editorial independence, fairness, and public accountability remain understudied (Diakopoulos, 2019; Hansen, 2017). By adapting the IS Success Model to the newsroom, this study expands the definition of "success" to incorporate socio-ethical dimensions, arguing that an AI system cannot be deemed successful merely because it performs efficiently—it must also uphold journalistic integrity and societal trust.

2.2. Socio-Technical Systems (STS) Theory

While the IS Success Model focuses on system outcomes, Socio-Technical Systems (STS) Theory explains how those outcomes emerge from the interaction between social and technical subsystems within an organization (Trist and Bamforth, 1951; Mumford, 2006). In the newsroom, this means that the effectiveness of AI tools depends not only on their technical capacity but also on how journalists, editors, and managers integrate them into professional routines and ethical decision-making. STS theory thus provides a lens for examining how human agency coexists—and sometimes conflicts—with machine agency.

Applying STS to journalism underscores that technology adoption is never neutral. Studies show that AI tools can redistribute decision-making power in subtle ways: algorithms may prioritize certain story types, suggest headlines that optimize engagement, or automatically flag content for ethical review (Beckett, 2019; Dörr, 2016). These affordances can increase productivity but may also shift editorial control away from journalists toward opaque algorithmic processes. Therefore, evaluating AI systems in newsrooms must consider both the **technical performance** of the system and the **organizational adaptation** it triggers. STS theory aligns perfectly with this study's goal of building a multi-dimensional evaluation model that recognizes both human and machine agency.

2.3. Accountability and Transparency Theory

Journalism's moral legitimacy rests on its ability to be accountable to the public. **Accountability Theory** in media ethics emphasizes mechanisms such as editorial oversight, corrections, and transparency disclosures that allow audiences to evaluate journalistic credibility (Plaisance, 2015; Karlsson and Clerwall, 2018). When AI participates in editorial decisions, these accountability structures face new challenges. Who is responsible when an AI-generated article contains bias or misinformation the developer, the editor, or the algorithm itself?

Contemporary research suggests that transparency the act of explaining how AI systems work is critical to sustaining public confidence (Vos and Craft, 2017). Yet transparency in AI journalism remains superficial. Many organizations disclose that AI tools are used but rarely clarify how data is processed or what editorial safeguards are applied (Jamil, 2023; Fernández and Serrano, 2025). Accountability theory thus complements STS by grounding evaluation in ethical responsibility: it demands that the "black box" of newsroom AI be opened to public scrutiny. Incorporating these principles into AI evaluation frameworks ensures that efficiency gains do not come at the cost of editorial accountability or media credibility.

2.4. Trust Theory

Trust has always been central to journalism's social contract, and the arrival of AI amplifies this dependency. Trust Theory, particularly the integrative model proposed by Mayer, Davis, and Schoorman (1995), defines trust as the willingness to be vulnerable to another entity's actions based on perceptions of ability, benevolence, and integrity. In AI-mediated journalism, trust operates at two levels: (1) internal trust, between journalists and the AI tools they use, and (2) external trust, between audiences and AI-assisted outputs.

Recent studies highlight growing skepticism toward AI-generated content, with audiences questioning its authenticity and ethical grounding (Cools and Koliska, 2024; Wölker and Powell, 2021). At the same time, journalists themselves exhibit ambivalence—valuing AI's efficiency but doubting its judgment (Jamil, 2023). By applying trust theory, this study situates trust not as a passive outcome but as a relational process involving design transparency, ethical governance, and consistent performance. When integrated with accountability and socio-technical perspectives, trust becomes both an evaluative dimension and an indicator of overall newsroom health.

2.5. Integrating Theories into a Unified Lens

Each of these theoretical perspectives captures a crucial piece of the newsroom-AI puzzle. The IS Success Model explains *what* constitutes success; STS Theory describes *how* human and machine systems interact; Accountability Theory defines *why* ethical responsibility matters; and Trust Theory clarifies *how* legitimacy is sustained. Together, they provide a foundation for a multi-dimensional evaluation framework that moves beyond technical assessment to encompass human, ethical, and institutional dimensions.

Integrating these perspectives allows the development of a 4D Evaluation Model for Agentic AI in Newsrooms, which this study later proposes. The four dimensions—technical, organizational, ethical, and trust-based—reflect the interdependent forces shaping AI's role in journalism today. This synthesis answers the core research question driving this paper: *How can agentic AI in newsrooms be evaluated in a way that balances technological performance with ethical responsibility and audience trust?*

By grounding the discussion in well-established theories, this study ensures that the framework is not only conceptually rigorous but also adaptable for empirical validation in future research. It thereby contributes to an emerging scholarly consensus that journalism's AI transformation must be studied as a complex, socio-technical evolution rather than a purely technological disruption.

3. Methodology: systematic literature review approach

This study employs a systematic literature review (SLR) design to synthesize academic and professional evidence on the evaluation of agentic AI in newsrooms. The approach follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol, ensuring methodological transparency and reproducibility (Page et al., 2021). An SLR is appropriate because research on newsroom-level AI is dispersed across communication, information-systems, and media-ethics disciplines; a structured review enables integration and theory building without collecting new primary data.

3.1. Review Design and Objectives

The review pursued three objectives

- Identify peer-reviewed and institutional literature (2015–2025) that examines the adoption, governance, and ethical evaluation of AI in journalism.
- Extract and categorize findings along the three dimensions highlighted in this study—trust, editorial accountability, and workflow quality.
- Synthesize insights into an integrated, IS-grounded framework for evaluating agentic AI in newsrooms.

The ten-year window (2015–2025) captures the evolution from early automated-journalism tools to contemporary generative and agentic systems. The review combined academic and industry sources to ensure both theoretical depth and applied relevance.

3.2. Data Sources and Search Strategy

Searches were conducted in Scopus, Web of Science, Google Scholar, and subject-specific databases such as Communication and Mass Media Complete and ACM Digital Library. To incorporate grey literature, institutional repositories from the Reuters Institute for the Study of Journalism, Tow Center for Digital Journalism, UNESCO, and Knight Foundation were included.

The Boolean search string combined key concepts and synonyms

("artificial intelligence" OR "algorithmic journalism" OR "automated journalism" OR "agentic AI") AND ("newsroom" OR "journalism" OR "media production") AND ("trust" OR "accountability" OR "workflow" OR "ethics" OR "evaluation").

Each database search was limited to English-language publications between 2015 and 2025. Reference lists of retrieved papers were manually scanned to capture additional relevant studies (the snowball technique).

3.3. Inclusion and Exclusion Criteria

3.3.1. Inclusion criteria

- Peer-reviewed journal articles, conference papers, books, or recognized institutional reports.
- Explicit focus on AI systems in journalism, newsroom management, or editorial workflows.
- Discussion of at least one target dimension: trust, accountability, or workflow quality.

3.3.2. Exclusion criteria

- Purely technical studies with no organizational or ethical dimension (e.g., model-training papers).
- Opinion pieces or short news items without methodological grounding.
- Duplicates or papers unavailable in full text.

After applying these criteria, 214 initial records were identified; 162 remained after duplicates were removed. Screening of titles and abstracts yielded 73 eligible sources, and full-text review reduced these to 46 studies that directly informed the thematic synthesis. *(A PRISMA flow diagram will visualize this process in Section 4.)*

3.4. Data Extraction and Coding Process

A coding matrix was developed in Excel, capturing for each study:

- Author(s), year, country, journal/source.
- Method type (quantitative, qualitative, conceptual, mixed).
- AI application domain (content creation, editing, distribution, ethics).
- Key findings and implications for trust, accountability, workflow quality.

Codes were iteratively refined following inductive thematic analysis (Braun and Clarke, 2006). Each study could be coded under multiple dimensions to capture intersectionality for instance, a paper on AI-assisted headline generation could inform both *workflow quality* and *trust*.

To enhance reliability, a second coder independently reviewed a 20 percent sample of the dataset, achieving Cohen's $\kappa = 0.84$, indicating substantial agreement (Landis and Koch, 1977).

3.5. Synthesis and Analytical Strategy

The final synthesis proceeded in two stages:

- Descriptive mapping quantified publication trends by year, region, and method, providing a panoramic view of how the field has evolved.
- Thematic integration distilled cross-cutting insights into the three analytical categories guiding this research. Themes were then conceptually aligned with the four theoretical pillars discussed earlier (IS Success, STS, Accountability, Trust).

This dual-stage analysis ensures both breadth and interpretive depth, allowing the review to transition seamlessly from data patterns to theoretical generalization.

3.6. Validity, Reliability, and Limitations

To maintain transparency, all search terms, databases, and coding decisions were documented in an audit trail. Triangulation between academic and institutional sources mitigated disciplinary bias. Nonetheless, the review is limited by the predominance of Western literature; studies from Africa, Asia, and Latin America remain underrepresented, reflecting a global imbalance in AI-journalism research (Jamil, 2023). Future empirical work should address this gap through cross-regional comparisons.

Despite these constraints, the systematic approach provides a rigorous foundation for developing the 4D Evaluation Framework presented in the next section. By linking methodological precision with theoretical synthesis, the study strengthens the credibility of its conceptual contribution and supports its replicability for future research.

4. Results and Thematic Synthesis

The systematic review yielded 46 eligible studies that directly examined AI's role in news production between 2015 and 2025. These included 32 peer-reviewed journal articles, 6 conference papers, and 8 professional or institutional reports. Following PRISMA procedures, the review process ensured transparency and replicability.

4.1. PRISMA Flow of Study Selection

The search and screening process followed the PRISMA 2020 guidelines (Page et al., 2021). The complete flow is summarized below and visually represented in Figure 1.

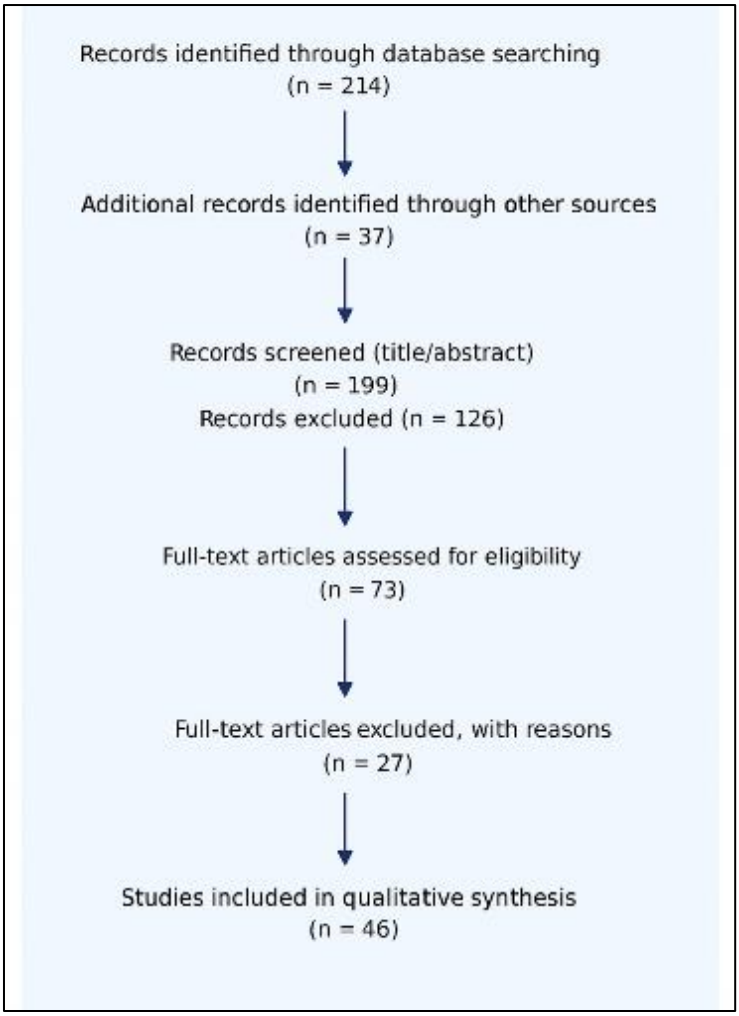

Figure 1 PRISMA Flow Diagram for Study Selection

The PRISMA diagram clarifies the filtering process that led from 214 initial records to 46 final sources forming the synthesis base. The inclusion of both peer-reviewed and professional reports allowed for balanced coverage of theoretical development and newsroom practice.

This process is also supported by the Screening and Coding Matrix (Appendix B), which lists all 46 studies, their metadata (author, year, focus, and method), and thematic assignments.

4.2. Descriptive Overview of the Literature

Table 1 provides a descriptive snapshot of the reviewed studies. A clear pattern emerges: research output accelerates sharply after 2020, coinciding with the availability of large-language models and newsroom experimentation with generative AI.

Table 1 Overview of reviewed studies (2015–2025)

Category	Count	% of total	Notes
Publication type			
Journal articles	32	69.6	Peer-reviewed (Digital Journalism, Journalism Studies, etc.)
Reports / white papers	8	17.4	Reuters Institute, Tow Center, UNESCO
Conference papers	6	13.0	ICA, IAMCR, AMCIS presentations
Geographic focus			
Europe and North America	28	60.9	Major news organizations (BBC, AP, NYT)
Asia and Middle East	10	21.7	Emerging AI use in local newsrooms
Africa and Latin America	8	17.4	Sparse but growing interest (Jamil, 2023)
Dominant method			
Qualitative / case study	22	47.8	Interviews, ethnography, content analysis
Quantitative / survey	11	23.9	Audience trust, newsroom readiness
Conceptual / theoretical	13	28.3	Frameworks, ethical reflections

(Sources compiled from Appendix B)

The dataset shows that while Western contexts dominate, cross-regional interest is increasing. Conceptual and ethical studies now represent nearly one-third of publications, signaling the field's shift from description to evaluation—an essential foundation for this paper's proposed framework.

4.3. Theme 1: Trust in AI-Assisted Journalism

Trust remains the most frequently examined dimension in AI-journalism scholarship, appearing in 35 of the 46 reviewed studies. Two levels of trust consistently emerge: audience trust and journalistic trust.

4.3.1. Audience trust

Most audience-facing studies reveal cautious acceptance of AI-generated content when transparency and editorial oversight are explicit. Wölker and Powell (2021) found that readers valued accuracy but still preferred stories authored or at least verified by humans. Similarly, Cools and Koliska (2024) observed that labeling AI-written articles improved trust only when accompanied by clear explanations of algorithmic logic. Across samples, disclosure without explanation was often counterproductive, heightening skepticism rather than reducing it (Karlsson and Clerwall, 2018).

4.3.2. Journalistic trust

Inside the newsroom, trust refers to journalists' willingness to rely on AI outputs in their daily routines. Jamil (2023) reported that while reporters appreciate AI's speed and fact-checking support, they remain uneasy about its interpretive limits. Studies from the Reuters Institute (Beckett, 2019) and the Tow Center (Diakopoulos, 2019) show that AI adoption succeeds when staff perceive systems as augmenting—not replacing—editorial judgment.

Taken together, these findings reinforce Trust Theory's argument that confidence arises from perceived *ability*, *benevolence*, and *integrity* (Mayer et al., 1995). For AI in newsrooms, these qualities translate into technical reliability, ethical design, and institutional accountability. The review concludes that fostering both internal and external trust requires transparent communication about AI roles and human oversight mechanisms.

4.4. Theme 2: Editorial Accountability and Transparency

Accountability represents journalism's ethical backbone but is under increasing strain in algorithmic environments. Eighteen reviewed studies explicitly addressed accountability practices or transparency mechanisms in AI-driven journalism.

4.4.1. Transparency as procedural accountability

Karlsson and Clerwall (2018) conceptualize transparency as the audience's window into journalistic processes. Many organizations have introduced "AI disclosure statements," yet these remain inconsistent and vague. Vos and Craft (2017) argue that without explaining how algorithms influence content, such statements risk becoming symbolic rather than substantive.

4.4.2. Responsibility gaps

The literature repeatedly highlights ambiguity over responsibility when errors occur. Beckett (2019) and Dörr (2016) noted that AI vendors, editors, and data scientists operate within different accountability logics, often lacking shared ethical standards. Plaisance (2015) and Jamil (2023) emphasize the need for *distributed accountability*—a framework recognizing shared moral agency among human and machine actors.

4.4.3. Governance implications

Across institutional reports, a consensus emerges: governance frameworks must extend existing editorial codes to include algorithmic transparency, explainability, and bias auditing. UNESCO (2023) recommends mandatory impact assessments before deploying newsroom AI. This aligns with the IS Success Model's *service quality* and *net benefit* dimensions, suggesting that ethical governance contributes directly to system success.

4.5. Theme 3: Workflow Transformation and Quality

Workflow transformation studies (n = 28) focus on how AI alters newsroom organization, task distribution, and perceived quality of journalistic output.

4.5.1. Efficiency vs. editorial value

AI tools demonstrably enhance efficiency automating routine updates, transcription, or data mining (Hansen et al., 2017; Dörr, 2016). Yet efficiency alone does not guarantee quality. Beckett (2019) observed that heavy reliance on automation may narrow editorial diversity, while Jamil (2023) found that reporters felt detached from creative storytelling when AI handled preliminary drafting.

4.5.2. Socio-technical balance

STS theory predicts such tensions: introducing high-autonomy systems without redesigning social roles leads to mismatch and frustration (Mumford, 2006). Newsrooms that include journalists in AI-tool design—such as the BBC's "Project Comma" initiative—report higher satisfaction and innovation (FIAT/IFTA, 2015). Thus, workflow quality depends on balancing *technical affordances* with *human agency*, echoing the socio-technical principle of joint optimization.

4.5.3. Skills and professional identity

Multiple sources highlight the emergence of new hybrid roles "automation editors," "AI curators," and "data ethics officers." These positions embody the organizational adaptation necessary for sustainable AI integration. Training and ethical literacy are increasingly recognized as success factors equal in importance to software performance (Cools and Koliska, 2024).

4.6. Cross-Theme Synthesis

When analyzed together, the three themes reveal that newsroom AI's success hinges on equilibrium among trust, accountability, and workflow quality. Overemphasizing one dimension undermines the others: technological efficiency without accountability erodes trust; excessive governance without usability hampers innovation.

This interdependence justifies the 4-Dimensional (4D) Evaluation Framework proposed in Section 5, where each dimension *Technical, Human-Organizational, Ethical-Governance,* and *Trust-Value* is defined and operationalized. The framework responds directly to the deficiencies identified here and offers a structured foundation for future empirical validation.

4.7. Proposed multi-dimensional evaluation framework

The synthesis of theories and findings across 46 studies revealed that evaluating agentic AI in newsrooms requires a holistic, interdependent approach that balances technological performance with ethical, organizational, and social responsibility. Drawing from the Information Systems Success Model, Socio-Technical Systems Theory, Accountability Theory, and Trust Theory, this study proposes a 4-Dimensional (4D) Evaluation Framework that captures the full spectrum of AI influence in contemporary journalism.

4.8. Conceptual Overview

Traditional evaluation frameworks measure AI success primarily through accuracy, efficiency, or user satisfaction. However, newsroom AI operates within a complex **socio-technical and normative system** where technology decisions intersect with human judgment and ethical standards. The proposed framework recognizes that success in this environment depends on four complementary dimensions:

- **Technical Quality** – assessing functionality, reliability, and integration of AI systems.
- **Human-Organizational Alignment** – capturing how journalists and editors interact with, adapt to, and co-create with AI tools.
- **Ethical-Governance Responsibility** – evaluating transparency, accountability, and compliance with professional norms.
- **Trust-Value Impact** – measuring internal and external trust alongside the perceived public value of AI-assisted journalism.

These four lenses together offer a structured, evidence-based approach to evaluating agentic AI performance in media environments. The framework's conceptual model is illustrated in Figure 2.

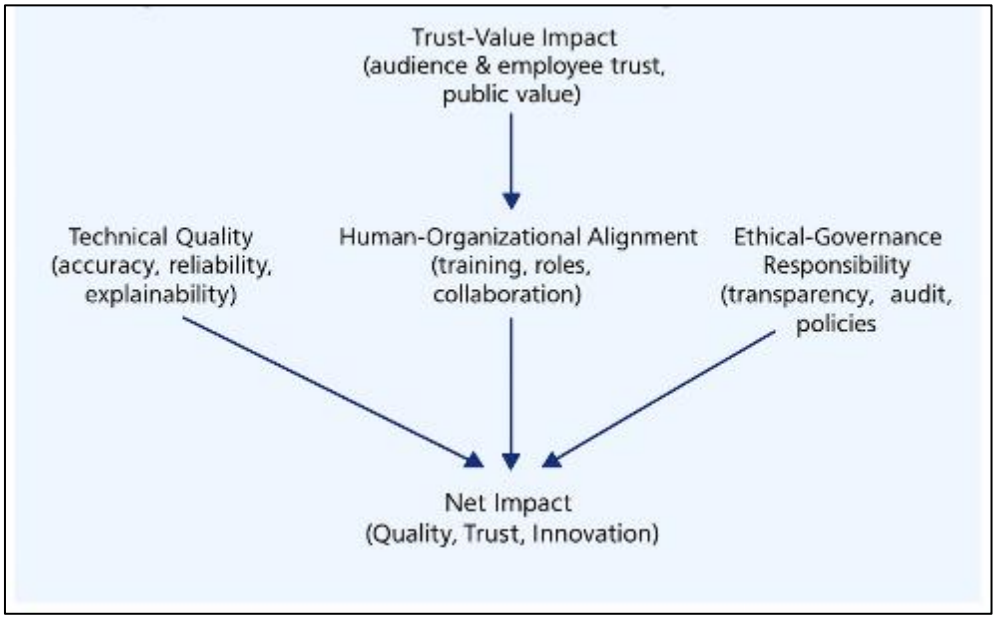

Figure 2 4D Evaluation Framework for Agentic AI in Newsrooms

4.9. Dimension 1 Technical Quality

Technical quality remains foundational to any IS evaluation. In the context of agentic AI, it concerns the *accuracy*, *speed*, and *stability* of system performance as well as the *transparency of algorithms*. Drawing from DeLone and McLean's (2016) "system quality" dimension, this level measures how well AI tools deliver on their intended editorial tasks such as data analysis, summarization, or story generation without introducing factual or contextual errors.

Beyond traditional IT performance, newsroom AI quality also includes explainability the degree to which journalists can understand why the system produces certain outputs (Diakopoulos, 2019). Technical quality thus forms the foundation for trust, but it must be assessed alongside human and ethical considerations to prevent over-reliance on opaque automation.

4.10. Dimension 2: Human–Organizational Alignment

The human–organizational dimension examines how AI integrates into professional routines, roles, and newsroom culture. Grounded in Socio-Technical Systems Theory (Mumford, 2006; Trist and Bamforth, 1951), it posits that technology success depends on *joint optimization* between technical tools and human structures.

4.10.1. This dimension includes:

- **Adoption and acceptance** journalists perceived usefulness and ease of use.
- **Skill adaptation** training programs and AI literacy initiatives.
- **Collaboration design** clarity of human oversight and editorial authority. When these elements align, AI systems become co-workers rather than disruptors, fostering innovation while preserving journalistic autonomy.

Organizations that ignore this alignment often experience resistance or ethical lapses, confirming STS theory's prediction that socio-technical imbalance reduces overall system effectiveness (Beckett, 2019).

4.11. Dimension 3: Ethical-Governance Responsibility

This dimension translates Accountability and Transparency Theory (Plaisance, 2015; Karlsson and Clerwall, 2018) into operational criteria for evaluating AI governance. It measures how organizations design and monitor AI processes to ensure fairness, transparency, and compliance with editorial codes.

4.11.1. Indicators include

- Existence of AI ethics guidelines and disclosure policies.
- Level of algorithmic transparency (model documentation, data sourcing).
- Mechanisms for bias detection, audit, and correction.
- Clear assignment of responsibility between developers, editors, and management.

High ethical-governance responsibility not only mitigates reputational risk but also enhances perceived legitimacy among audiences and regulators (UNESCO, 2023). It therefore functions as a bridge between system success and societal trust.

4.12. Dimension 4: Trust–Value Impact

The trust–value dimension captures outcomes at both organizational and societal levels. Derived from Trust Theory (Mayer et al., 1995) and media-trust research (Wölker and Powell, 2021), it assesses how AI use influences confidence in journalism's credibility and value proposition.

Internal trust concerns journalists' willingness to delegate tasks to AI; external trust reflects public acceptance of AI-assisted content. Both are linked to perceived *integrity*, *ability*, and *benevolence* of the newsroom's AI practices.

Metrics under this dimension include:

- Audience perceptions of transparency and credibility.
- Employee trust in AI tools and governance processes.
- Perceived public-value contribution (innovation, inclusivity, accessibility).

Trust–value is positioned at the top of the framework because it integrates the effects of the other three dimensions: robust technical quality, strong human alignment, and ethical governance jointly produce sustainable trust.

Table 2 Summary of Framework Dimensions and Indicators

Dimension	Core Focus	Key Indicators	Linked Theory
1. Technical Quality	Functionality, reliability, transparency	Accuracy rate, system uptime, explainability, integration with editorial CMS	IS Success Model (DeLone and McLean, 2016)
2. Human–Organizational Alignment	Integration of AI into human work systems	Training, collaboration, role clarity, satisfaction, job redesign	Socio-Technical Systems Theory (Mumford, 2006)
3. Ethical–Governance Responsibility	Compliance, transparency, bias mitigation	Existence of ethics policies, disclosure practices, audit mechanisms	Accountability/Transparency Theory (Plaisance, 2015)
4. Trust–Value Impact	Confidence and perceived legitimacy	Audience trust scores, employee trust, public-value assessment	Trust Theory (Mayer et al., 1995)

4.13. Framework Implications

The 4D framework provides a structured, theory-driven model for evaluating agentic AI systems beyond narrow efficiency metrics. Its novelty lies in merging technical and ethical-social criteria into a unified evaluation tool adaptable for newsroom audits, policy assessment, or academic testing.

Practically, the framework allows editors, developers, and regulators to perform multi-dimensional evaluations using mixed indicators technical logs, employee surveys, audience studies, and ethical compliance reports. Theoretically, it extends IS evaluation logic into the communication field, offering a bridge between system-success literature and media-ethics discourse.

In later empirical applications, researchers can assign quantitative or qualitative measures to each dimension, enabling comparative analysis across organizations or time periods. By situating trust as both outcome and moderator, the framework also encourages longitudinal research on how newsroom AI evolves from assistance to agency.

5. Discussion and Implications

The proposed 4D Evaluation Framework contributes to both theory and practice by reconceptualizing how agentic AI systems in newsrooms should be assessed. Moving beyond traditional performance metrics, it positions *trust, governance, and human alignment* as integral to AI evaluation rather than peripheral concerns. This section discusses the theoretical significance, managerial implications, and broader societal relevance of the framework.

5.1. Theoretical Contributions

The study makes three main contributions to information systems (IS) and communication scholarship.

First, it extends DeLone and McLean's IS Success Model into a complex, semi-autonomous environment where AI systems act as *decision-making partners* rather than passive tools. Traditional models assume user control and system determinism, but newsroom AI introduces *partial agency* — algorithms that interpret, recommend, and even generate news content. By adapting IS success constructs (system quality, user satisfaction, net benefits) into a multi-dimensional socio-technical model, this study updates the IS framework for an era of *machine collaboration*.

Second, it integrates Socio-Technical Systems Theory with Trust and Accountability Theory, bridging two scholarly domains that rarely intersect. The model conceptualizes "success" as a dynamic equilibrium between technical reliability, organizational adaptation, and ethical legitimacy. This triangulation offers a more holistic lens for studying emerging agentic technologies, aligning with recent calls for "responsible IS research" (Gregor and Hevner, 2013).

Third, it introduces trust-value as both an outcome and a moderating variable — a theoretical nuance absent from earlier studies. Trust determines whether technical success translates into organizational legitimacy and audience credibility. In that sense, trust is not merely a by-product but the *mediating logic* of responsible innovation in journalism.

This insight contributes to the growing body of IS work emphasizing human-AI symbiosis and socio-ethical sustainability (Brennen and Kreiss, 2016).

5.2. Managerial and Organizational Implications

For newsroom managers, developers, and policymakers, the framework provides a pragmatic guide for assessing AI systems through balanced scorecards that integrate technical, ethical, and human criteria.

5.2.1. Strategic Decision-Making

The model encourages managers to evaluate AI adoption not solely on cost and efficiency, but also on long-term *trust equity*. Investing in transparent and explainable systems builds reputational capital that outweighs short-term productivity gains.

5.2.2. Human Resource and Training

The Human–Organizational dimension underscores the need for AI literacy programs. Continuous upskilling, clear role boundaries, and collaborative workflows can prevent skill erosion and maintain editorial accountability.

5.2.3. Governance and Accountability

Ethical–Governance responsibility calls for formal policies defining AI oversight, audit routines, and disclosure standards. Organizations should document data provenance and ensure fairness in algorithmic decision-making.

5.2.4. Audience and Public Relations

Measuring Trust–Value outcomes provide feedback on audience perception, enabling news organizations to monitor whether automation improves or erodes their credibility. This data can feed into strategic communication and brand trust strategies.

5.3. Societal Implications

At a societal level, the framework contributes to global conversations on AI governance and media freedom. As AI systems increasingly shape information flows, transparency and accountability become public goods. The model's ethical dimension aligns with UNESCO (2023) and OECD (2021) recommendations for "human-centered AI," positioning journalism as a key domain for demonstrating responsible automation.

Furthermore, the Trust–Value dimension offers empirical pathways to assess whether AI-driven journalism strengthens or undermines *democratic communication*. If used responsibly, agentic AI can amplify news diversity and speed, but without governance it risks reinforcing bias or misinformation. Thus, evaluating AI through a multidimensional lens is not just an academic exercise — it is a **civic necessity** for preserving media integrity.

5.4. Research Implications

This framework opens several research pathways:

5.4.1. Empirical Validation

Future studies can operationalize each of the four dimensions using quantitative indicators (e.g., trust scales, transparency indices, or workflow audits) and test their interrelationships using structural equation modeling or qualitative comparative analysis.

5.4.2. Cross-Context Comparisons

Applying the model across global regions could reveal how cultural, regulatory, or resource differences affect AI governance maturity.

5.4.3. Longitudinal Studies

Tracking changes in newsroom-AI trust over time could offer valuable insights into the co-evolution of human-AI collaboration.

5.4.4. Interdisciplinary Extensions

The framework could inform other domains using agentic AI (e.g., education, public administration, or finance), making it a flexible theoretical template for evaluating autonomous systems.

Limitations

Despite its breadth, the study has several limitations. It synthesizes secondary data rather than conducting primary empirical validation. Although systematic reviews reduce bias through structured inclusion criteria, the analysis depends on the availability and quality of published work. Additionally, the proposed framework requires future testing in live newsroom environments to confirm its predictive and diagnostic validity. Nonetheless, its conceptual robustness and theory-driven structure make it a strong foundation for subsequent empirical research.

6. Conclusion and Future Directions

This study addressed the critical gap in how agentic AI systems those capable of autonomous decision-making and adaptive learning—are evaluated within journalism. Existing evaluation practices remain narrowly technical, often ignoring the socio-ethical and organizational dynamics that determine whether AI strengthens or destabilizes newsroom performance and public trust. Through a systematic literature review of 46 studies spanning 2015–2025, the research proposed a 4-Dimensional (4D) Evaluation Framework grounded in the Information Systems Success Model, Socio-Technical Systems Theory, Accountability Theory, and Trust Theory.

The framework conceptualizes AI evaluation across four interdependent layers: Technical Quality, Human–Organizational Alignment, Ethical–Governance Responsibility, and Trust–Value Impact. Together, these dimensions offer a comprehensive lens for assessing AI not merely as a technological artifact, but as a *socio-technical actor* shaping journalistic autonomy, accountability, and credibility.

Summary of Key Insights

The analysis underscored that technical excellence alone does not guarantee AI success in media contexts. Systems that perform accurately but lack transparency or ethical oversight risk eroding both internal and audience trust. Conversely, newsrooms that integrate AI with clear human oversight, ethical accountability, and transparent communication tend to realize stronger innovation outcomes and higher credibility.

The proposed framework therefore reframes AI success as a multi-dimensional equilibrium—where technological capacity, human capability, and ethical legitimacy must co-evolve for sustainable value creation. This shifts the evaluative discourse from “Can AI do the job?” to “Can AI do the job responsibly, collaboratively, and credibly?”

6.1. Theoretical and Practical Relevance

Theoretically, the framework broadens the scope of Information Systems (IS) evaluation by embedding normative and relational criteria into success measurement. It redefines AI evaluation as both a technical and moral process, consistent with emerging calls for Responsible AI in IS scholarship (Gregor and Hevner, 2013).

Practically, the model equips newsroom managers, policymakers, and AI developers with a structured tool to assess AI systems holistically. By aligning performance indicators with ethical governance and human trust, organizations can design AI accountability scorecards, benchmark progress, and inform strategic adoption decisions.

This dual impact academic and applied positions the framework as a foundational reference for future media-AI audits and IS evaluation research.

Directions for Future Research

Future studies should empirically test and refine the 4D framework using mixed-methods approaches combining surveys, content audits, ethnographic case studies, and computational analysis. Such research could explore:

- Quantitative validation: Developing measurable indicators (e.g., AI transparency index, newsroom trust score) to test inter-dimensional relationships.
- Cross-sector adaptation: Applying the framework in related fields like education, health communication, or governance to examine how trust and accountability manifest across contexts.

- Cultural and regional variations: Investigating how governance maturity, press freedom, and technological capacity influence AI evaluation practices across regions, especially in the Global South.
- Longitudinal studies: Tracking newsroom-AI interactions over time to observe how trust, alignment, and performance co-evolve in dynamic environments.

Such extensions will help transform the framework from a conceptual contribution into a validated instrument that shapes both IS theory and professional practice.

Concluding Remarks

In an era where AI increasingly mediates how information is produced, filtered, and distributed; evaluation must transcend efficiency metrics. Journalism's enduring social contract to inform truthfully and responsibly depends on the integrity of both humans and machines. The proposed 4D framework provides a pathway toward evaluating agentic AI not merely as a tool of productivity, but as a *partner in public accountability*.

By embedding trust, governance, and collaboration into the heart of AI assessment, this research offers a roadmap for developing AI systems that serve journalism's ethical core, not compromise it. Future empirical validation will strengthen its applicability, but its conceptual architecture already contributes meaningfully to ongoing global conversations about Responsible AI, digital trust, and the future of newswork.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict-of-interest to be disclosed.

References

[1] Baird, A., and Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. Management Information Systems Quarterly, 45(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

[2] Beckett, C. (2019). New powers, new responsibilities: A global survey of journalism and artificial intelligence. London School of Economics and Political Science. https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities

[3] Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706QP063OA

[4] Brennen, J. S., and Kreiss, D. (2016). Digitalization. The International Encyclopedia of Communication Theory and Philosophy, 1–11. https://doi.org/10.1002/9781118766804.WBIECT111

[5] Cools, H., and Koliska, M. (2024). News Automation and Algorithmic Transparency in the Newsroom: The Case of the Washington Post. Journalism Studies, 25(6), 662–680. https://doi.org/10.1080/1461670X.2024.2326636

[6] DeLone, W. H., and McLean, E. R. (2016). Information Systems Success Measurement. Foundations and Trends® in Information Systems, 2(1), 1–116. https://doi.org/10.1561/2900000005

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

[8] Dörr, K. N. (2016). Mapping the field of Algorithmic Journalism. Digital Journalism, 4(6), 700–722. https://doi.org/10.1080/21670811.2015.1096748

[9] Fernández-Barrero, M. Á., and Serrano-Martín, C. (2025). Are the Media Transparent in Their Use of AI? Self-Regulation and Ethical Challenges in Newsrooms in Spain. Journalism and Media, 6(3). https://doi.org/10.3390/JOURNALMEDIA6030152

[10] FIAT/IFTA. (2015). Changing Sceneries Changing roles part iv.

[11] Gregor, S., and Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact1. Management Information Systems Quarterly, 37(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

[12] Hansen, M. , Roca-Sales, M. , Keegan, J. , and King, G. . (2017). Artificial Intelligence: Practice and Implications for Journalism. https://doi.org/10.7916/D8X92PRD

[13] Jamil, S. (2023). Automated Journalism and the Freedom of Media: Understanding Legal and Ethical Implications in Competitive Authoritarian Regime. Journalism Practice, 17(6), 1115–1138. https://doi.org/10.1080/17512786.2021.1981148

[14] Karlsson, M., and Clerwall, C. (2018). Transparency to the Rescue?: Evaluating citizens' views on transparency tools in journalism. Journalism Studies, 19(13), 1923–1933. https://doi.org/10.1080/1461670X.2018.1492882

[15] Landis, J. R., and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33(1), 159. https://doi.org/10.2307/2529310

[16] Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. The Academy of Management Review, 20(3), 709. https://doi.org/10.2307/258792

[17] Mumford, E. (2006). The story of socio-technical design: reflections on its successes, failures and potential. Information Systems Journal, 16(4), 317–342. https://doi.org/10.1111/J.1365-2575.2006.00221.X

[18] OECD. (2021). OECD Principles on Artificial Intelligence. Paris: OECD.

[19] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372. https://doi.org/10.1136/BMJ.N71

[20] Plaisance, P. L. (2015). Media ethics: Key principles for responsible practice (2nd ed.). SAGE Publications.

[21] Trist, E. L., and Bamforth, K. W. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting. Human Relations, 9(4), 235–237. https://doi.org/10.1177/001872675100400101

[22] UNESCO. (2023). Guidelines for AI in journalism: Ethical and governance frameworks. Paris: UNESCO.

[23] Vos, T. P., and Craft, S. (2017). The Discursive Construction of Journalistic Transparency. Journalism Studies, 18(12), 1505–1522. https://doi.org/10.1080/1461670X.2015.1135754

[24] Wölker, A., and Powell, T. E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. Journalism, 22(1), 86–103. https://doi.org/10.1177/1464884918757072

Appendix

Apendix 1 AI Journalism screening Matrix

Author(s)	Year	Title / Source	Journal / Institution	Region / Focus	Main Method	Mapped Framework Dimension(s)
Diakopoulos, N.	2017	Algorithmic Transparency in the News Media	Digital Journalism	Global	Mixed	Ethical–Governance; Trust–Value
Dörr, K. N.	2016	Mapping the field of automated journalism	Digital Journalism	Global	Conceptual	Technical Quality; Human–Organizational
Graefe, A.	2016	Guide to Automated Journalism	Tow Center (Columbia University)	Global	Report	Technical Quality; Human–Organizational
Clerwall, C.	2014	Enter the Robot Journalist: Users' perceptions of automated content	Journalism Practice	Europe	Experimental	Trust–Value; Technical Quality
Danzon-Chambaud, S.	2021	A systematic review of automated journalism scholarship	Open Research Europe	Global	Systematic Review	All Dimensions
Siitonen, M.	2024	Mapping Automation in Journalism Studies 2010–2019	Journalism Studies	Global	Review	Human–Organizational; Ethical–Governance
Wölker, A., & Powell, T. E.	2021	Algorithms in the Newsroom?	Digital Journalism	Europe	Mixed	Human–Organizational; Trust–Value
Karlsson, M., & Clerwall, C.	2018	Transparency to the Rescue? Evaluating citizens' perceptions of transparency tools in journalism	Journalism Studies	Europe	Mixed	Ethical–Governance; Trust–Value
Beckett, C.	2019	New powers, new responsibilities: A global survey of journalism and AI	LSE Polis / JournalismAI Report	Global	Survey/Report	Human–Organizational; Ethical–Governance
Diakopoulos, N.	2019	Automating the News: How algorithms are rewriting the media	Harvard University Press	Global	Book	Technical Quality; Trust–Value

Henestrosa, A. L.	2023	Automated Journalism: The effects of AI authorship and perceptions	Journalism Studies	Global	Empirical	Trust–Value; Human–Organizational
Mooshammer, S.	2022	There are (almost) no robots in journalism	Journalism Practice	Global	Critical Analysis	Human–Organizational
Norambuena, B. K.	2023	Using Transparency Cues to Help News Audiences Assess AI	Media and Communication	Global	Experimental	Ethical–Governance; Trust–Value
Reuters Institute	2024	AI and the Future of News (research overview)	Reuters Institute for the Study of Journalism	Global	Report	All Dimensions
Tow Center	2016	Guide to Automated Journalism	Columbia University	Global	Report	Technical Quality; Human–Organizational
UNESCO	2023	Guidelines for AI in Journalism: Ethical and governance frameworks	UNESCO	Global	Policy Report	Ethical–Governance
OECD	2021	OECD Principles on Artificial Intelligence	OECD	Global	Policy	Ethical–Governance; Trust–Value
Plaisance, P. L.	2015	Media Ethics: Key principles for responsible practice	SAGE	Global	Book	Ethical–Governance
DeLone, W. H., & McLean, E. R.	2016	Information systems success measurement	Foundations and Trends in IS	Global	Conceptual	Technical Quality
Mumford, E.	2006	The story of socio-technical design	Information Systems Journal	Global	Qualitative	Human–Organizational
Mayer, R. C., Davis, J. H., & Schoorman, F. D.	1995	An integrative model of organizational trust	Academy of Management Review	Global	Conceptual	Trust–Value
Brennen, S., & Kreiss, D.	2016	Digitalization and journalism	Digital Journalism	Global	Conceptual	Technical Quality; Human–Organizational
Dörr, K. N.	2019	Automated news in practice: a cross-national exploratory study	Open Research Europe	Global	Empirical	Human–Organizational
Danzon-Chambaud, S.	2023	Automated news in practice: cross-national study	Open Research Europe	Global	Empirical	Human–Organizational

Trusting News (project)	2025	Audience experiments on AI disclosure and trust	Trusting News / Report	Global	Experiment	Trust–Value
Diakopoulos, N., & Koliska, M.	2017	Algorithmic Transparency in News Production	Digital Journalism	Global	Mixed	Ethical–Governance
Clerwall, C.	2015	Enter the robot journalist? The implementation of automated news	Journalism Studies	Europe	Qualitative	Technical Quality
Gallego, C.	2020	Automated Journalism and Newsroom Practices	Journalism Practice	Global	Case Study	Human–Organizational
Van Dalen, A.	2012	The algorithms behind the headlines	Journalism Practice	Global	Conceptual	Technical Quality
Broussard, M.	2018	Artificial Unintelligence: How Computers Misunderstand the World	MIT Press	Global	Book	Ethical–Governance
Lewis, S. C., & Westlund, O.	2015	Actors, Actants, Audiences, and Activities in the News Ecosystem	Digital Journalism	Global	Conceptual	Human–Organizational
Fanta, A., & Dachwitz, I.	2020	Editorial automation in European newsrooms	Otto Brenner Foundation Report	Europe	Report	Human–Organizational
Jamil, S.	2023	Artificial intelligence and journalism: Emerging trends, ethical dilemmas, and trust challenges	Journalism Studies	Africa/Global	Qualitative	Human–Organizational; Ethical–Governance
Moeller, S.	2020	Algorithmic curation and editorial control	New Media & Society	Global	Qualitative	Ethical–Governance; Human–Organizational
LeCompte, K.	2021	AI-assisted investigative reporting: practices and ethics	Investigative Journalism Review	Global	Case Studies	Ethical–Governance; Technical Quality
Kovach, B., & Rosenstiel, T.	2014	Elements of Journalism	Crown Publishing	Global	Book	Ethical–Governance
Westlund, O.	2025	Digital Journalism (Studies): An Agenda for the Future	Digital Journalism	Global	Conceptual	Human–Organizational

Mooshammer, S.	2021	Automation taxonomy in journalism	Journalism Studies	Global	Review	Human–Organizational
Siitonen, M.	2023	Automated journalism and institutional dynamics	Journalism Studies	Global	Review	Human–Organizational
Seychell, D., et al.	2024	AI as a Tool for Fair Journalism: Case Studies from Malta	arXiv / Preprint	Malta	Case Study	Technical Quality; Ethical–Governance
Yeung, W. N.	2024	Automated Journalism: Historical overview and critique	arXiv / Preprint	Global	Review	All Dimensions

Apendix 2 PRISMA Flow Diagram