Scientific Research Publishing

Open Journal of Social Sciences, 2026, 14(11), 359-381
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Verifiable Journalism in the Generative AI Era

—A Provenance-by-Design Architecture for News Authenticity, AI Disclosure, and Editorial Accountability

ASM Johirul Islam

MSIT, Emporia State University, Emporia, KS, USA
Email: aislam@g.emporia.edu

How to cite this paper: Islam, ASM J. (2026). Verifiable Journalism in the Generative AI Era. *Open Journal of Social Sciences, 14,* 359-381.
https://doi.org/10.4236/jss.2026.1411020

Received: October 3, 2026
Accepted: November 20, 2026
Published: November 23, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Generative artificial intelligence has changed the economics of content production faster than journalism has changed the infrastructure used to prove where news came from, who altered it, and how human editorial responsibility was exercised. The central problem is therefore not only the detection of false content. It is the absence of a durable and interoperable chain of provenance across the news-production lifecycle. Existing responses remain fragmented: fact-checking is largely post-publication; machine-learning detectors are probabilistic and vulnerable to distribution shift; platform moderation is opaque; and content-credential standards provide important authenticity metadata but do not, by themselves, establish newsroom governance, protected editorial auditability, or factual truth. This study develops Verifiable Journalism Blockchain Architecture 2.0 (VJBA 2.0), a conceptual and design-science-oriented sociotechnical architecture that integrates C2PA-compatible public content credentials, decentralized identity, cryptographic versioning, AI-use disclosures, distributed timestamp commitments, and privacy-preserving newsroom audit controls. The paper makes four contributions. First, it reframes the research problem from fake-news detection to provenance assurance and trust calibration. Second, it distinguishes cryptographic authenticity from factual accuracy, preventing provenance claims from being overstated as truth claims. Third, it introduces a dual-boundary design in which public verification evidence is separable from confidential newsroom records and protected sources. Fourth, it develops theoretically grounded propositions and an evaluation protocol for testing technical integrity, organizational accountability, audience interpretation, and governance legitimacy. The framework is positioned as complementary to emerging standards such as C2PA rather than as a replacement for them. Its novelty lies in connecting public content provenance to the internal editorial chain of custody and in specifying how these records can remain independently auditable without placing sensitive journalistic material on a public ledger.

Keywords

Verifiable Journalism, Content Provenance, Generative AI, C2PA, Blockchain, Decentralized Identity, Editorial Accountability, Misinformation, Trust Calibration, Newsroom Governance

1. Introduction

Digital journalism now operates in an information environment where convincing text, images, audio, and video can be generated, transformed, and redistributed at extremely low cost. The concern is no longer limited to obviously fabricated stories. A more difficult problem arises when authentic media are placed in false contexts, when legitimate articles are silently modified, when AI-assisted rewriting obscures responsibility, or when synthetic material is mixed with human reporting in ways that audiences cannot inspect. These conditions weaken a basic assumption on which journalism depends: that readers can connect a published claim to an accountable source, publication process, and editorial record.

The scale of the credibility problem is visible in current audience research. The Reuters Institute Digital News Report 2025, based on almost 100,000 online news consumers across 48 markets, reported that overall trust in news remained at 40%. The same report found broad public scepticism about generative AI in journalism: respondents expected AI to make news cheaper and more current, but less transparent, less accurate, and less trustworthy (Newman et al., 2025). These attitudes matter because AI is already being used for transcription, summaries, headline generation, translation, rewriting, recommendation, and in some settings automated article production. The practical challenge for newsrooms is therefore not whether AI will enter the production process, but how its use can be disclosed, governed, and verified without destroying legitimate editorial flexibility.

Research on misinformation has produced sophisticated taxonomies and detection methods, but much of that work begins after questionable content is already circulating. Wardle and Derakhshan (2017) distinguish misinformation, disinformation, and malinformation, while Zhou and Zafarani (2020) classify fake-news detection into knowledge-, style-, propagation-, and source-based approaches. Such methods are valuable, yet they typically estimate whether content is deceptive; they do not create a persistent record of who authored an item, which transformations occurred, when editorial decisions were made, or whether provenance metadata survived cross-platform distribution. Vosoughi, Roy, and Aral (2018) further showed that false news can diffuse farther, faster, deeper, and more broadly than truth on social media, underscoring why purely post-hoc correction is structurally disadvantaged.

At the same time, the technical landscape has advanced. The Coalition for Content Provenance and Authenticity (C2PA) now specifies a standardized mechanism for cryptographically signed content credentials that can carry assertions

about origin and transformations. IPTC has developed guidance for expressing trust and credibility information in news-industry metadata. W3C Decentralized Identifiers provide a standards-based foundation for verifiable identity. These developments substantially change the research baseline. A credible contribution can no longer be framed as simply putting a news hash on a blockchain. The more difficult research question is how interoperable provenance credentials can be integrated with newsroom identity, editorial version control, AI-use disclosure, confidential audit records, governance, and audience-facing verification in a way that is technically defensible and institutionally realistic.

1.1. Problem Statement

The problem addressed in this study is the lack of an end-to-end, journalism-specific provenance architecture that links public-facing authenticity evidence to the internal editorial chain of custody while preserving confidentiality and avoiding the false inference that provenance equals truth. Current approaches solve only parts of the problem. Fact-checking evaluates claims but usually does so after publication. AI detectors estimate whether content appears synthetic but can produce uncertain or unstable classifications. Platform moderation can reduce distribution but is controlled by private intermediaries. C2PA credentials can describe origin and editing history, yet the standard is intentionally general-purpose and opt-in; it does not define a complete newsroom accountability model. Blockchain prototypes can provide immutability, but immutable storage does not establish that the original input was accurate, nor does it solve identity governance, source confidentiality, correction policy, or audience comprehension.

This fragmentation creates four concrete failure points. First, provenance breaks across organizational and platform boundaries because credentials, CMS histories, corrections, and publication metadata are often stored in separate systems. Second, AI assistance is frequently disclosed as a simple label, even though meaningful accountability requires information about the type of AI use, materiality, human review, and responsible editor. Third, centralized editorial logs can be changed or selectively disclosed, while fully public ledgers would create unacceptable risks for source protection, privacy, and newsroom security. Fourth, audiences may interpret a cryptographic authenticity marker as proof that a claim is factually true, even though cryptography can verify custody and signatures, not the truth of the underlying reporting. A useful architecture must therefore establish verifiable process evidence while explicitly limiting the scope of what that evidence proves.

1.2. Rationale of the Study

The study is timely for three reasons. The first is technological. Generative AI has reduced the cost of creating persuasive synthetic content and increased the number of transformations that may occur before a story reaches the public. The

second is institutional. News organizations face persistent trust deficits while also adopting AI under commercial pressure to produce more content with fewer resources. The third is regulatory and standards-related. Article 50 of the European Union AI Act, applicable from August 2026, introduces transparency obligations concerning machine-readable marking of AI-generated or manipulated content and disclosure in specified cases involving deepfakes and public-interest text. In parallel, C2PA and IPTC are creating a practical interoperability layer for provenance and publisher metadata. These developments make provenance assurance a concrete design and governance problem rather than a speculative one.

There is also a strong theoretical rationale. Trust in news is not generated by technology alone. It depends on signals, institutions, accountability relationships, and the cost of verification. Signaling Theory explains why credible, difficult-to-fake provenance cues may reduce information asymmetry (Spence, 1973). Transaction Cost Economics explains why lowering the effort required to inspect authorship, revision history, and publication timing can change verification behavior (Williamson, 1985). Agency Theory focuses attention on monitoring and responsibility when reporters, editors, publishers, platforms, and audiences have different incentives (Jensen & Meckling, 1976). Institutional Theory helps explain whether provenance practices become normalized through professional standards, regulation, and competitive imitation (DiMaggio & Powell, 1983). Bringing these perspectives into a design framework allows the study to ask not only whether provenance is technically possible, but under what conditions it can become credible and usable.

1.3. Research Questions and Objectives

The paper is guided by four research questions: (RQ1) What trust and accountability failures remain after current fact-checking, AI-detection, and content-credential approaches are considered? (RQ2) How can public provenance evidence be linked to the newsroom editorial chain of custody without exposing confidential sources or sensitive internal material? (RQ3) Which architectural mechanisms can represent AI involvement, editorial changes, and publisher responsibility in interoperable and auditable form? (RQ4) How should a verifiable-journalism architecture be evaluated so that technical integrity is not confused with factual truth or audience trust?

Accordingly, the objectives are to synthesize the current provenance and journalism literature, identify the unresolved design gap, develop VJBA 2.0 as a standards-compatible conceptual architecture, derive theoretically grounded propositions, and specify an empirical evaluation protocol for future prototype and field testing.

1.4. Novel Contributions

The contribution of this study is deliberately narrower and stronger than a claim

that blockchain can make journalism trustworthy. Blockchain-based news provenance has already been explored in prior research and industry prototypes, including the New York Times News Provenance Project and later blockchain-based verification mechanisms. C2PA also provides a mature technical foundation for signed provenance. The novelty of VJBA 2.0 is therefore the integration logic around those building blocks.

Contribution	What is new in VJBA 2.0?	Why it matters?
C1. Provenance-by-design framing	Moves the unit of analysis from detecting questionable content to maintaining a verifiable chain of identity, transformation, approval, publication, and correction.	Prevention and accountability are addressed before and after publication, rather than only after misinformation spreads.
C2. Dual-boundary architecture	Separates public provenance commitments from confidential newsroom evidence; sensitive material remains off-chain while hashes, approvals, and time commitments can be independently verified.	Addresses the tension between transparency and source protection.
C3. Standards bridge	Uses C2PA-compatible content credentials and W3C-style decentralized identity as interoperability layers, while distributed ledgers are limited to commitments and audit functions where immutability adds value.	Avoids building a closed blockchain ecosystem that competes with emerging standards.
C4. AI-use accountability schema	Represents AI use as a structured editorial event including task type, materiality, model/tool identifier where appropriate, human reviewer, and approval state rather than a binary "AI-generated" label.	Supports meaningful disclosure and organizational responsibility.
C5. Truth/provenance separation	Explicitly defines cryptographic provenance as evidence of origin and custody, not proof of factual accuracy.	Prevents a central category error in authenticity systems and supports calibrated audience interpretation.
C6. Multi-level evaluation	Combines cryptographic, workflow, governance, privacy, and audience measures rather than using invented performance targets.	Creates a falsifiable research program for prototype and field validation.

2. Literature Review

2.1. Information Disorder and the Limits of Content-Level Detection

The misinformation literature demonstrates that false or misleading content is not a single technical category. Wardle and Derakhshan (2017) distinguish misinformation from intentional disinformation and from malinformation, where genuine information is used harmfully. This distinction has architectural implications: a system can verify that a file was signed by a recognized publisher without establishing that the publisher's claim is correct, complete, or ethically contextualized. Similarly, a genuine photograph can be used in a misleading story while retaining perfect provenance. Provenance must therefore be treated as one evidentiary layer in a broader journalistic verification process.

Computational fake-news research has produced useful methods based on linguistic features, external knowledge, social propagation, and source characteris-

tics (Shu et al., 2017; Zhou & Zafarani, 2020). However, machine-learning classification remains probabilistic and context-sensitive. Detection systems can also become strategically vulnerable as generative models improve and as adversaries optimize against classifiers. Recent scholarship on AI-mediated authenticity verification warns that provenance neglect can cause systems to treat low-quality or synthetic evidence as if it were independent corroboration (Yanesa et al., 2026). This reinforces the need to preserve information about origin and custody alongside content-level analysis.

Empirical and behavioral scholarship also helps explain why provenance evidence is necessary but insufficient: social-media misinformation has institutional and economic drivers (Allcott & Gentzkow, 2017), and individuals' judgments about false news are shaped by cognitive and attention-related factors (Pennycook & Rand, 2021). Accordingly, VJBA does not substitute a verified publisher identity for scrutiny of the underlying claim.

2.2. Generative AI, Journalism, and Audience Trust

Generative AI changes both production and verification. Newsrooms can use AI to support research, transcription, summarization, translation, data analysis, and editing, but the same tools can fabricate quotations, citations, images, or video. Reuters Institute research in 2025 found that audiences are more comfortable with supportive uses of AI when humans remain involved, while they are more sceptical of heavily automated news production (Newman et al., 2025). This suggests that the design problem is not solved by an AI label alone; the label must communicate the human accountability structure around the AI-assisted process.

Disclosure may also have unintended effects. Schilke and Reimann (2025) report experimental evidence that disclosure of AI use can reduce trust in the person or organization using AI under some conditions. This finding complicates simple transparency narratives. A provenance system should not assume that more disclosure automatically produces more trust. The more defensible goal is trust calibration: giving audiences enough reliable context to distinguish known from unknown provenance and to judge the degree of human responsibility for a published item.

2.3. Content Credentials, Provenance Standards, and Identity

C2PA represents the most important contemporary standardization effort in digital content provenance. Its specification provides a model for signed manifests and assertions describing how an asset was created or transformed, with cryptographic bindings that make subsequent tampering detectable. C2PA is explicitly intended to enable interoperable provenance across applications, and current versions include mechanisms relevant to AI-generated or transformed content (C2PA, 2026). The standard is a major advance because it shifts provenance from proprietary metadata to a verifiable ecosystem model.

For journalism, C2PA is complemented by IPTC standards that structure news metadata and by IPTC guidance on expressing trust and credibility information. IPTC has also worked to make C2PA easier to use in news-industry workflows (IPTC, 2025). W3C Decentralized Identifiers (DIDs) provide another relevant foundation by allowing controllers to prove control over identifiers using cryptographic methods without dependence on a single centralized identity provider (W3C, 2022). Together, these standards suggest that a journalism architecture should reuse interoperable identity and content-credential mechanisms rather than invent competing formats.

At the same time, provenance standards have limits. C2PA is not a factual truth engine, and its trustworthiness depends on the security of claim generators, credential issuance, key management, trust lists, and user interpretation. A 2026 preprint by Golaszewski and colleagues presents an independent security critique of C2PA and argues that high-stakes deployments require stronger assurance around the protocol's security goals. Whether every claim in that preprint survives peer review or subsequent specification changes, the broader design lesson is important: provenance systems require threat modelling, independent validation, and careful communication of what a credential does and does not prove.

2.4. Blockchain and Distributed-Ledger Approaches to Journalism

Blockchain has been proposed for timestamping, tamper evidence, licensing, decentralized publication, and fake-news traceability. Hasan and Salah (2019) demonstrated how blockchain and smart contracts could support deepfake-related provenance. The New York Times News Provenance Project, developed with IBM, explored a Hyperledger-based approach for tracking contextual metadata for photojournalism and found that transparency cues had to be designed carefully for audiences; simply emphasizing the word blockchain could itself reduce comprehension or trust (Ivancsics & Ghosh, 2020). Evans et al. (2020) similarly treated provenance as a way to reveal reuse and origins of news text so that users could make more informed judgments about publishers and articles.

Later blockchain research has continued to explore storage, verification, and traceability of news data. These studies establish that distributed ledgers can make certain records difficult to rewrite and can support independently inspectable timestamps. They also reveal a recurring weakness: many proposals treat immutability as if it were synonymous with credibility. The "garbage in, garbage forever" problem remains. A ledger can preserve a false statement perfectly. VJBA 2.0 therefore uses distributed ledgers only for verifiable commitments, ordering, and audit evidence; it does not use consensus as a substitute for editorial verification.

2.5. Regulatory and Risk-Management Context

The policy context is moving toward explicit transparency for synthetic content. The European Commission's 2026 guidance on Article 50 of the AI Act clarifies obligations concerning machine-readable marking of AI-generated or manipulated output and disclosure of certain deepfakes and public-interest text. For public-interest text, the regulatory framework also recognizes the significance of human review and editorial responsibility. This is directly relevant to journalism because it makes the quality of the human accountability record as important as the technical marker itself (European Commission, 2026).

Risk management guidance points in the same direction. NIST's Generative AI Profile extends the AI Risk Management Framework with considerations for generative systems across their lifecycle (Autio et al., 2024). A provenance architecture for journalism should therefore be evaluated not only for cryptographic integrity but for governance, privacy, misuse, explainability, and operational resilience.

2.6. Research Gap

The literature now contains strong components but not a complete journalism-specific integration. Misinformation research explains deceptive content and human susceptibility. AI research explains new production and verification risks. C2PA provides interoperable asset provenance. IPTC provides newsroom metadata structures. DIDs provide identity primitives. Blockchain provides tamper-evident commitments. What remains insufficiently specified is how these components should be assembled across the full editorial lifecycle, including pre-publication versions,

Research stream	Established contribution	Remaining limitation addressed here
Misinformation/ fake-news detection	Content classification, propagation analysis, cognitive and source-based indicators.	Usually reactive; does not create a durable editorial chain of custody.
C2PA/content credentials	Interoperable signed manifests and tamper-evident provenance assertions.	General-purpose; does not by itself define newsroom accountability, confidential audit governance, or factual truth.
IPTC news metadata	Structured metadata for news production and syndication; trust/credibility guidance.	Metadata can be informative without being independently anchored to an immutable audit history.
Blockchain journalism prototypes	Timestamping, immutability, distributed verification, experimental provenance.	Often overclaims "trust" and under-specifies source confidentiality, standards interoperability, and organizational governance.
AI disclosure research	Shows disclosure can influence perceptions and may sometimes reduce trust.	Needs richer, process-aware disclosure and trust-calibration design.
VJBA 2.0	Integrates public standards, newsroom chain of custody, AI-use accountability, dual privacy boundaries, and distributed audit commitments.	Conceptual artifact still requires prototype, security testing, and field validation.

human approvals, AI-use events, corrections, confidential audit evidence, and audience-facing trust calibration. This is the gap addressed by VJBA 2.0.

3. Research Approach and Theoretical Foundation

3.1. Design-Science Orientation

This paper is a conceptual design study. It follows the logic of design-science research in information systems, where a problem is analyzed, an artifact is designed to address that problem, and evaluation criteria are specified so the artifact can later be tested (Hevner et al., 2004; Gregor & Hevner, 2013). VJBA 2.0 is therefore not presented as an empirically validated production system. It is a design artifact and associated research model intended to guide implementation, comparative testing, and organizational field studies.

The requirements were derived through a structured, purposive synthesis rather than a systematic review or an exhaustive standards audit. The starting corpus comprised the misinformation, audience-trust, design-science, and news-provenance studies discussed in Section 2, complemented by directly applicable specifications and guidance (C2PA, 2026; IPTC, 2025; W3C, 2022; European Commission, 2026; Autio et al., 2024). Sources were selected for their relevance to one or more of six design questions: content-origin continuity, identity and authority, tamper-evident workflow history, accountable AI involvement, source confidentiality, and audience interpretation. Preference was given to normative primary standards and original empirical or conceptual studies; conflicting requirements were treated as design trade-offs rather than evidence of consensus. Each candidate requirement was mapped from a documented limitation or standard capability to an architectural control and a prospective test. This transparent, problem-driven derivation is not claimed to establish the completeness or prevalence of requirements across all newsrooms.

A traceability example is requirement R1, interoperable verification of publication origin and subsequent transformations. The C2PA (2026) manifest, signature, and asset-binding model motivates portable public credentials, while the limitations of newsroom-level accountability identified in Section 2 motivate a protected internal event chain. The design decision is therefore Layer 3's signed C2PA-compatible publication envelope linked by a controlled identifier to Layer 5's protected editorial record, rather than publication of raw newsroom documents. The proposed test is to reproduce a story across supported channels and verify both manifest continuity and correct, access-controlled retrieval of the internal audit trail. Likewise, R2 (authorization, not merely identifier control) maps W3C (2022) DID control to the organizational vetting and delegated signing workflow in Section 4.2; R3 (confidentiality despite public commitments) maps source-protection concerns to the keyed, batched, access-controlled approach in Section 5; and R4 (interpretable disclosure) links the audience and AI-disclosure literature to Sections 4.5 and 6.4. These mappings are design rationales, not empirical proof of control effectiveness.

3.2. Transaction Cost Economics

Verification is costly. Readers rarely have the time or expertise to contact sources, inspect a publisher's correction history, or compare cryptographic hashes. Editors also incur costs when reconstructing who changed a story and why. Transaction Cost Economics suggests that institutions and technologies can reduce these frictions (Williamson, 1985). VJBA 2.0 therefore treats provenance metadata as a mechanism for reducing the cost of answering narrow but important questions: Who signed this item? Is this the same version that was published? What editorial transformations were recorded? Was AI use declared? Was a correction issued?

3.3. Signaling Theory and Trust Calibration

In conditions of information asymmetry, observable and costly-to-fake signals can communicate otherwise hidden qualities (Spence, 1973). Cryptographic signatures, verified organizational identities, and signed correction histories can function as such signals. However, signaling theory does not imply that every signal will increase trust. If audiences do not understand a credential, or if they perceive AI disclosure negatively, the signal may reduce confidence. The appropriate dependent concept is therefore calibrated trust rather than maximum trust.

3.4. Agency Theory and Editorial Accountability

Perceived usefulness and ease of use may influence whether newsroom staff adopt provenance tools (Davis, 1989). The future newsroom pilot will therefore examine perceived utility and interaction effort alongside institutional pressures, rather than assuming that a standards-compliant system will be adopted simply because it is technically available.

Journalism contains nested principal-agent relationships: publishers delegate to editors, editors to reporters, audiences rely on publishers, and platforms mediate distribution. Agency Theory predicts accountability problems when actions are difficult to observe or incentives diverge (Jensen & Meckling, 1976). A signed chain of editorial events can strengthen monitoring without requiring that every internal document be public. VJBA 2.0 is designed around this distinction between verifiable event commitments and protected evidence.

3.5. Institutional Theory and Adoption

Technical feasibility is insufficient if newsrooms do not adopt the system. Institutional Theory explains how professional norms, regulation, and competitive imitation can normalize practices across an organizational field (DiMaggio & Powell, 1983). C2PA adoption, emerging AI transparency rules, platform requirements, and newsroom codes of ethics can create normative, coercive, and mimetic pressures toward provenance practices. These pressures also create risks of symbolic compliance, making independent auditability important.

3.6. Research Propositions

Proposition	Statement	Primary theoretical basis
P1	Content accompanied by independently verifiable provenance will reduce perceived uncertainty about authorship and modification history relative to identical content without provenance evidence.	Transaction Cost Economics; Signaling Theory
P2	Structured AI-use disclosures that identify human review and editorial responsibility will produce better trust calibration than binary "AI-generated" labels.	Signaling Theory; Human-AI trust
P3	Tamper-evident version commitments and signed correction events will increase perceived editorial accountability, provided that the verification interface clearly distinguishes provenance from factual accuracy.	Agency Theory; Accountability
P4	A dual-boundary architecture that keeps sensitive newsroom evidence off-chain while publishing verifiable commitments will achieve higher organizational acceptability than designs that place detailed newsroom records on a public ledger.	Agency Theory; Privacy-by-design
P5	Interoperability with C2PA/IPTC identity and metadata standards will increase adoption intention relative to a proprietary blockchain-only provenance scheme.	Institutional Theory; Technology adoption
P6	The effect of provenance signals on audience credibility judgments will be moderated by news-brand familiarity, AI literacy, and understanding of what the credential verifies.	Signaling Theory; Trust calibration

4. VJBA 2.0 System Architecture

4.1. Design Principles

VJBA 2.0 is built on five principles. First, provenance is evidence about origin and process, not a truth certificate. Second, standards interoperability is preferred to proprietary metadata. Third, sensitive source material should not be placed on a public or broadly shared ledger. Fourth, every material AI-assisted or editorial transformation should be representable as an accountable event. Fifth, verification should remain possible even when content is copied across platforms or a newsroom later changes its website (Figure 1).

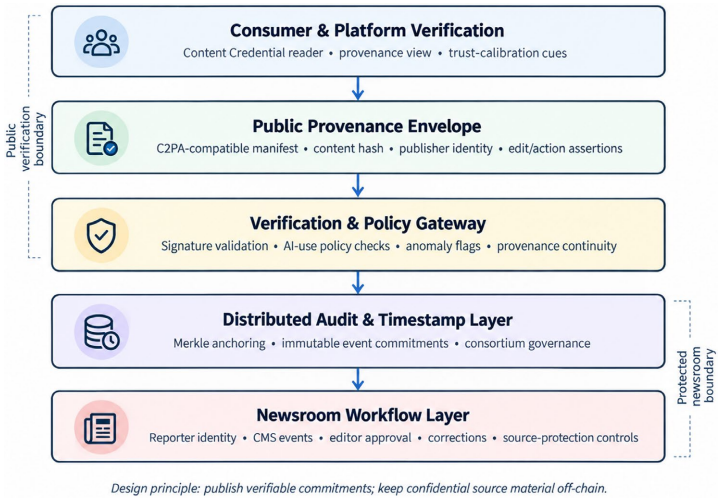


Figure 1. VJBA 2.0 provenance-by-design architecture and public/protected trust boundaries.

4.2. Layer 1: Verified Newsroom and Contributor Identity

Identity-validation workflow. First, a designated newsroom registration authority verifies the publisher's legal or operational existence and control of its claimed publication domain, using documentary checks, a domain challenge, and an authorized officer's approval; an independent consortium registrar or accredited verifier may cross-check these assertions where cross-publisher recognition is needed. Second, the newsroom identity administrator verifies employment or contractual authorization and the relevant role of each contributor, assigns least-privilege permissions, and issues a time-limited, revocable role credential (optionally bound to a DID). Third, a controlled credential registry binds the verified publisher and authorized role to an approved C2PA signing service and its signing-certificate identifier or key reference. A release gate checks current role authorization, credential status, separation-of-duties rules, and certificate validity before the service signs a C2PA manifest. Fourth, the verifier validates both the content signature and the publisher authorization attestation; failure or expiry is reported separately. A DID proof only establishes control of an identifier, not the truth of a claimed organization or a person's editorial authority. Pseudonymous contributors can be verified internally without disclosing their identities publicly, subject to independent oversight and revocation.

The identity layer binds authorized organizational and contributor identities to cryptographic keys. News organizations can maintain institutional credentials, while individual journalists or editors can use organization-issued credentials or decentralized identifiers. The objective is not to expose personal identity in every publication. Pseudonymous or role-based credentials may be appropriate for high-risk reporting. What matters is that the organization can prove that a publishing event was authorized by a recognized credential and that credential status can be revoked when necessary. W3C DID concepts are relevant because they separate identifier control from a single identity provider and support cryptographically verifiable control (W3C, 2022).

Ledger anchoring is conditional, not a prerequisite for provenance security. A newsroom-controlled signed append-only log with hash chaining and independently issued trusted timestamps is the baseline: it already supplies tamper evidence and credible time bounds when keys, timestamps, and auditors are independent. A permissioned multi-organization ledger may add value under a narrower threat model in which a publisher and its log operator might collude to suppress or fork historical entries, a single timestamp authority is not an acceptable institutional dependency, or multiple publishers need a common non-equivocating ordering of syndication and correction events. Independent validators, separately administered by organizations without common control, would co-sign periodic Merkle-root commitments and publish consistency evidence. A ledger does not prevent an initially dishonest entry, protect compromised endpoints, guarantee validator independence, or replace archival and disclosure policy. If external timestamping plus independent log mirrors sufficiently address these

threats, the ledger should be omitted; the comparative evaluation in Sections 6.2 and 8.2 will test whether any incremental benefit warrants its cost and governance burden.

4.3. Layer 2: Editorial Version and Timestamp Commitments

Every material version of a story is assigned a content hash and a timestamped event record. Rather than writing the full article to a blockchain, the system stores a cryptographic commitment and a minimal event type. Multiple events can be batched in a Merkle tree and periodically anchored to a distributed ledger. This design reduces cost and avoids unnecessary disclosure while preserving tamper evidence. If a newsroom later modifies a published story, the new version produces a new hash and references the preceding event, creating a verifiable sequence rather than silently overwriting history.

4.4. Layer 3: Public Content Credential Envelope

Editorial rule for material AI involvement. For purposes of VJBA 2.0, AI use is material and requires public-facing disclosure when generated or substantially transformed output becomes audience-visible journalistic content; when it creates or materially changes a reported factual assertion, quotation, translation affecting meaning, image, audio, or video; or when it influences a consequential editorial selection or summary that a reasonable reader would consider relevant to assessing human responsibility. Routine spelling correction, formatting, and purely internal brainstorming that neither changes published meaning nor supplies reported facts remain recorded according to internal policy but ordinarily do not trigger a public AI-use label. Borderline cases are escalated to the responsible editor, with a written rationale and public-disclosure decision. A structured record states the task, materiality determination, human verification and approval, and disclosure status; the public assertion reveals only appropriate fields. Mandatory legal disclosures, including applicable Article 50 obligations, take precedence over this newsroom rule (European Commission, 2026). Signatures authenticate only what the newsroom declares and records: they cannot prove that every AI interaction occurred in a logged system or that all AI involvement was reported. Audit sampling, tool-access monitoring where proportionate, training, and whistleblower protections are complementary controls.

At publication, the system creates or extends a C2PA-compatible manifest containing the publisher identity, content bindings, relevant actions, and selected provenance assertions. This public credential travels with the asset where platform support exists. For text and web articles, VJBA 2.0 treats the public manifest as a portable summary of provenance, while the ledger anchor functions as a persistence mechanism for the newsroom's signed event commitments. This relationship is intentionally complementary: C2PA carries interoperable assertions; the distributed audit layer provides an external commitment to the newsroom event history.

4.5. Layer 4: AI-Use Accountability Record

A binary AI label is too crude for newsroom accountability. VJBA 2.0 models AI use as a structured event with fields such as task category (e.g., transcription, translation, summarization, image generation, code assistance), materiality, model or tool identifier where policy permits, input-output retention policy, human reviewer, approval state, and whether the AI-generated element is visible to the audience. The public credential may expose only the subset necessary for meaningful disclosure, while the protected audit record retains additional details for internal review. This approach aligns with the direction of C2PA's AI-related provenance mechanisms and with regulatory emphasis on machine-readable marking and disclosure, while preserving editorial responsibility as a human-governance concept rather than a machine property.

4.6. Layer 5: Protected Newsroom Audit Trail

The protected audit layer records workflow events that are relevant to accountability but unsuitable for public disclosure: approvals, correction decisions, access-control changes, policy exceptions, and cryptographic references to supporting evidence. Confidential sources, interview notes, raw documents, and sensitive communications remain in newsroom-controlled repositories. The ledger stores only the minimum commitment needed to prove that a record existed in a given state at a given time. Access to detailed evidence is governed by newsroom policy, legal privilege, source-protection rules, and, where appropriate, independent audit procedures.

4.7. Layer 6: Verification and Trust-Calibration Interface

The consumer-facing interface answers a limited set of questions in plain language: Is the publisher identity verified? Has this file or article changed since the signed version? What kinds of editing actions were declared? Was generative AI materially involved? Has the publisher issued a correction? Can the provenance chain be validated? The interface must also display a clear limitation notice: verified provenance does not mean that every factual claim is true. This distinction is central to VJBA 2.0 because users may otherwise overgeneralize cryptographic integrity into epistemic certainty.

4.8. Event Schema and Data Minimization

Privacy attack-and-control example. Suppose a confidential source submits one of several predictable documents and the newsroom anchors an unsalted SHA-256 hash of the document with an exact-minute timestamp. An adversary who possesses likely candidates can hash each one and match a public commitment, revealing which document was used; correlating the timestamp with observed meetings or upload activity may additionally identify the source. Merely keeping the file off-chain does not prevent this dictionary attack or metadata correlation. VJBA instead commits to canonical event records through a secret-keyed com-

mitment or a cryptographically strong, non-reused random nonce retained only in the protected repository; commitments are aggregated into periodically published Merkle roots, with publication schedules decoupled from individual editorial events. Public records omit source identifiers, document sizes, fine-grained event times, and stable cross-story pseudonyms unless clearly required. Keys and nonces are access controlled, rotated and backed up; designated auditors can receive selective opening proofs under documented authorization. Padding, delayed batches, minimum batch sizes and retention limits reduce metadata leakage, but a small newsroom or exceptional public event may still be linkable. Threat tests should include candidate-document guessing, temporal correlation, insider misuse and unauthorized proof requests, and report remaining privacy risk rather than asserting perfect anonymity.

Event field	Public credential	Protected audit record	Rationale
Content hash/asset binding	Yes	Yes	Required to verify integrity and continuity.
Publisher identity	Yes	Yes	Supports organizational accountability.
Individual contributor identity	Optional/role-based	Yes, access controlled	Protects journalists where public identification creates risk.
Timestamp/ordering proof	Yes	Yes	Supports chronology and correction history.
AI-use category	Yes when material	Yes	Enables meaningful disclosure.
Model/tool details	Policy dependent	Yes where retained	Avoids exposing unnecessary operational or security details.
Raw source material	No	Off-chain repository only	Essential for source protection and data minimization.
Editorial rationale/legal review	Summary or no	Access controlled	Balances accountability with privilege and newsroom confidentiality.
Correction event	Yes	Yes	Allows public verification of version history and remediation.

5. Threat Model and Security Assumptions

Assess a provenance architecture against realistic adversaries rather than idealized trust. VJBA 2.0 considers at least six threat classes: key theft or unauthorized signing; malicious insiders who create false but properly signed records; replay or stripping of credentials during platform distribution; compromised claim-generation software; collusion among consortium participants; and interface-level deception that causes users to misunderstand a valid credential. None of these threats is solved by adding a blockchain alone.

The design therefore depends on operational controls such as hardware-backed key storage where feasible, credential revocation, role separation, independent timestamp anchors, signed software releases, audit sampling, and incident-response procedures. C2PA's trust model and security guidance should be incorporated directly into implementation. The 2026 security critique of C2PA also motivates formal threat analysis before using VJBA in high-stakes settings (Golaszewski

et al., 2026).

A second security principle is graceful degradation. If a platform removes provenance metadata, the news item should not automatically be treated as false; rather, its provenance becomes unavailable or unverifiable. Conversely, a valid credential should not immunize content from fact-checking. The system should represent states such as verified provenance, broken provenance, unknown provenance, revoked credential, and signature failure without converting them into simplistic true/false labels.

6. Evaluation Framework

6.1. Why the Evaluation Model Is Changed

The earlier version of this manuscript used numerical baseline and target values for authenticity rates, latency, incident frequency, disclosure compliance, and audience trust without an empirical dataset supporting those numbers. Those values are removed in this revision. A conceptual architecture should not present invented performance targets as if they were observed industry statistics. VJBA 2.0 instead defines measurable evaluation dimensions and a staged validation protocol. Establish quantitative targets only after testing a prototype and baseline system under the same conditions (Figure 2).

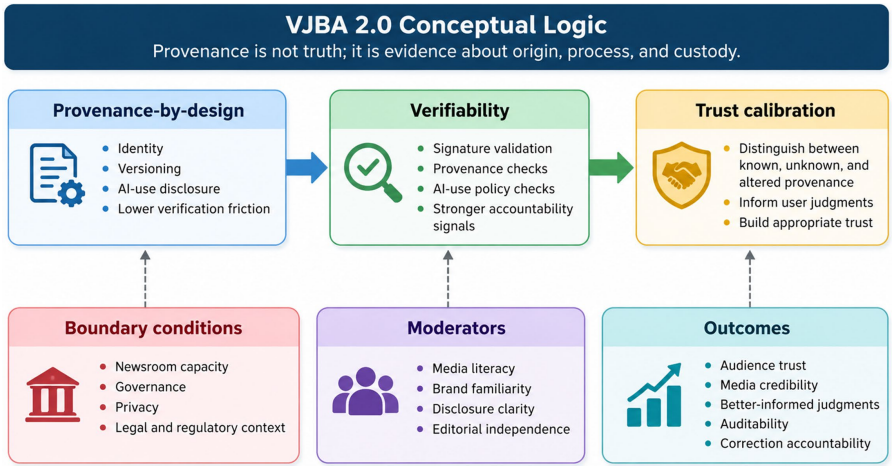


Figure 2. Conceptual research logic linking provenance-by-design to verifiability, trust calibration, and accountability outcomes.

6.2. Technical Evaluation

Primary trust-calibration metrics will be Brier scores for binary origin, custody and fact-accuracy judgments, calculated separately on adjudicated items; absolute calibration error and reliability plots across pre-specified confidence bins will complement them. For unresolved items, analysis will examine appropriate uncertainty and abstention. The key overtrust test is whether participants assign elevated probability of factual accuracy to a cryptographically valid but factually false item. Planned comparisons across interface arms will estimate effects on these dis-

tinct outcomes, with random article effects, prior brand familiarity and AI literacy as prespecified covariates; sample size and effect thresholds will be set by a priori power analysis. The study will report discriminative accuracy, calibration and response times, avoiding the assumption that higher mean trust is an improvement.

The experiment will use a preregistered factorial set of stimuli in which actual provenance status (valid, broken, absent, or validly signed but organizationally unauthorized) varies independently from fact-checker-established factual accuracy (accurate, inaccurate, or unresolved). For each article, participants will separately report 0 - 100% confidence that (a) the claimed publisher originated or authorized the item, (b) its chain of custody and revisions are intact, and (c) its central factual claim is accurate; an overall credibility rating is elicited only as a distinct secondary outcome. Ground-truth labels for origin and custody will be generated from controlled signed-event histories, whereas factual labels will be established in advance using independently documented source evidence and blinded expert adjudication. Unknown or unresolved claims will not be forced into binary accuracy scoring.

Dimension	Measure	Example test
Integrity	Percentage of published versions whose hashes and signatures validate.	Replay a corpus of original and modified articles and test detection of unauthorized changes.
Provenance continuity	Percentage of legitimate transformations that retain an unbroken credential chain.	Export, syndicate, republish, and update content across supported systems.
Verification latency	Time to validate identity, manifest, signatures, and ledger anchor.	Benchmark browser/mobile verification under realistic network conditions.
Resilience	Recovery from key revocation, missing manifests, node outage, or stripped metadata.	Inject failure cases and record false acceptance/false rejection behavior.
Interoperability	Successful validation by independent C2PA-compatible tools and news systems.	Cross-validate manifests produced by different implementations.

6.3. Organizational and Governance Evaluation

Field evaluation should measure workflow burden, policy compliance, correction practices, audit completeness, source-protection concerns, and perceived fairness of consortium governance. Interviews with reporters, editors, security teams, legal counsel, standards specialists, and newsroom management can identify whether the architecture supports or obstructs professional norms. Adoption intention should be compared between a standards-compatible implementation and a proprietary blockchain-only implementation to test P5.

6.4. Audience Evaluation

Audience experiments should compare several interface conditions: no provenance information; a simple "verified" badge; a C2PA-style credential view; and a VJBA 2.0 view that explicitly separates provenance from factual truth and provides structured AI-use information. Outcomes should include perceived credi-

bility, confidence calibration, comprehension of what the credential verifies, willingness to inspect provenance, and susceptibility to the false belief that a valid signature proves factual accuracy. This design can also test whether AI disclosure produces a trust penalty and whether information about human editorial review moderates that effect (Schilke & Reimann, 2025).

6.5. Proposed Empirical Program

Stage	Method	Primary question
Stage 1: Prototype	Implement CMS plug-in, signing service, C2PA manifest generation, event ledger, and verification widget.	Can the architecture be implemented without placing sensitive source material on-chain?
Stage 2: Security test	Threat modelling, penetration testing, key-compromise simulation, credential stripping, replay tests.	What attacks break provenance continuity or create false confidence?
Stage 3: Controlled user experiment	Randomized comparison of provenance/disclosure interfaces.	How do interface designs affect comprehension, trust calibration, and credibility judgments?
Stage 4: Newsroom pilot	Deploy in one or more volunteer newsrooms for a defined set of stories.	What operational burdens, governance conflicts, and workflow adaptations appear?
Stage 5: Longitudinal study	Observe corrections, republishing, credential continuity, and audit events over time.	Does provenance improve accountability behavior and post-publication traceability?

7. Comparative Positioning

VJBA 2.0 should be evaluated against complementary approaches rather than presented as a universal replacement. Fact-checkers evaluate factual claims and context. Platform moderation controls distribution. AI detectors offer probabilistic signals. C2PA provides interoperable asset provenance. VJBA 2.0 adds a newsroom-specific accountability and audit layer that connects these public credentials to a protected editorial chain of custody.

Capability	Fact-checking	AI detector	C2PA/content credentials	VJBA 2.0
Factual claim assessment	Strong	Limited /probabilistic	Not its purpose	Not its purpose; integrates references to editorial verification evidence
Cryptographic asset provenance	Usually no	No	Core capability	Core capability via C2PA-compatible envelope
Editorial version chain	Partial/manual	No	Can represent actions, implementation dependent	Explicit, timestamped newsroom event chain
Structured AI-use accountability	Variable	Detects/estimates AI use	Supports AI provenance assertions	Adds newsroom policy, reviewer, materiality, approval state

Continued

Protected internal audit trail	No	No	Outside general scope	Explicit protected boundary with off-chain confidential evidence
Distributed timestamp commitment	No	No	Optional/ecosystem dependent	Used selectively for independent auditability
Source confidentiality model	Professional practice	No	Not a newsroom governance model	Data-minimizing design; sensitive sources remain off-chain
Audience truth/provenance distinction	Fact-check verdicts	Often opaque probability	Depends on UI	Required trust-calibration principle

Before deploying a consortium, a newsroom should compare three configurations using the same event corpus and attacker assumptions: (A) signed hash-chained append-only logs with internal timestamps; (B) the same logs with independent timestamping and external mirrors or transparency witnesses; and (C) independently governed multi-party ledger anchoring of batched roots. Evaluation should measure detection of deletion, reordering, split-view and backdating under collusion; verification latency; cost; availability; metadata exposure; and governance capture. Ledger deployment is justified only where scenario C demonstrably reduces a material risk left unresolved by B, especially cross-institutional equivocation or coordinated suppression, and where validator diversity and dispute resolution can be sustained. Otherwise configuration B is the simpler defensible default.

8. Discussion

8.1. Theoretical Contribution

The first theoretical contribution is a shift from trust maximization to trust calibration. Signaling mechanisms are valuable only when users understand what the signal means. A provenance credential should increase confidence about origin or custody when those properties are verified, but it should not necessarily increase confidence in factual accuracy. This distinction refines the application of Signaling Theory to digital news by separating the object of the signal from broader judgments about truth.

Second, VJBA 2.0 extends the transaction-cost perspective by treating verification friction as a design variable. The architecture lowers the cost of checking identity, version continuity, correction history, and AI involvement, but it does not eliminate the higher-order cost of assessing evidence and factual claims. This boundary clarifies what a provenance infrastructure can realistically change in the information market.

Third, the dual-boundary model contributes to accountability design. Agency Theory often emphasizes monitoring, but journalism cannot maximize transpar-

ency indiscriminately because source confidentiality and editorial independence are core institutional values. VJBA 2.0 introduces selective verifiability: organizations can commit cryptographically to the existence and sequence of protected records without publishing their contents. This provides a testable mechanism for balancing accountability and confidentiality.

8.2. Practical Implications for Newsrooms

For news organizations, the architecture suggests an incremental adoption path. A newsroom does not need to begin with a complex blockchain consortium. It can start by issuing signed content credentials, strengthening identity and key management, and generating verifiable correction histories. Distributed timestamp anchoring and protected consortium audit functions can be added where there is a concrete governance benefit. This avoids technology-first adoption and reduces the risk that blockchain becomes an expensive layer with no clear operational purpose.

The architecture also implies that AI policies should become machine-actionable. Instead of relying only on narrative newsroom policies, CMS workflows can require reporters or editors to declare material AI assistance, identify the responsible reviewer, and record approval. Those records can later support internal quality assurance, public disclosure, and regulatory compliance without forcing the newsroom to reveal prompts, confidential source material, or commercially sensitive workflows.

8.3. Policy and Standards Implications

Policy should avoid equating machine-readable provenance with truth. Regulatory and platform rules can require disclosure and tamper-evident marking, but the resulting credentials should be presented as evidence about process. The European AI Act's distinction between machine-readable marking and human-facing disclosure is useful in this respect because technical detectability and audience comprehension are separate requirements (European Commission, 2026).

Standards bodies and news organizations should also develop consistent vocabularies for editorial AI use. "AI-generated" is often too broad. A translation checked by an editor, a synthetic image, an automatically generated earnings summary, and a chatbot-produced draft raise different accountability questions. A cross-industry schema for task type, materiality, human review, and editorial responsibility would make provenance more meaningful than a single badge.

9. Limitations and Future Research

The principal limitation is that VJBA 2.0 remains a conceptual design. The propositions have not yet been tested, and the architecture has not been benchmarked against production workloads. Claims about improved accountability or audience understanding must therefore be treated as hypotheses, not results. A prototype is necessary before performance, cost, or trust effects can be quantified.

A second limitation is governance. A permissioned consortium can reduce de-

pendence on one publisher, but it can also reproduce industry concentration. Governance rules for membership, validator selection, dispute resolution, credential revocation, and audit access require independent study. Polycentric governance principles may be useful for preventing a small group of large organizations from controlling the trust infrastructure (Ostrom, 1990).

Third, cryptographic identity can create safety risks if implemented without role-based or pseudonymous options. Investigative journalists, reporters in authoritarian environments, and confidential contributors may require stronger anonymity and compartmentalization. Future designs should explore verifiable organizational authorization without universal public identification of individual journalists.

Fourth, content credentials can be stripped, lost, or become inaccessible across platforms. Future research should compare embedded credentials, external manifest repositories, watermarking, soft binding, and ledger-anchored recovery mechanisms. The appropriate architecture may differ for text, photographs, audio, recorded video, and live streams.

Finally, the audience layer needs careful experimentation. Transparency is not automatically beneficial, and too much technical detail can confuse users. Future studies should investigate which provenance cues improve understanding without creating automation bias, authority bias, or the mistaken belief that authenticated content is necessarily accurate.

10. Conclusion

Journalism's emerging authenticity problem cannot be reduced to detecting fake content. In a media environment shaped by generative AI, the deeper challenge is to preserve a trustworthy chain of evidence about origin, transformation, editorial responsibility, publication, and correction. Current tools address parts of that chain, but the chain itself remains fragmented. VJBA 2.0 responds by treating verifiable journalism as an infrastructure and governance problem. It combines public content credentials, standards-based identity, cryptographic version commitments, structured AI-use disclosure, protected newsroom audit evidence, and a consumer-facing verification layer. Its central design rule is deliberately modest: provenance can prove aspects of custody and process; it cannot prove truth. That limitation is not a weakness. It is what makes architecture conceptually defensible. The paper's novel contribution is therefore not "blockchain for news" in isolation. It is a standards-compatible, dual-boundary model that links portable public provenance to accountable newsroom processes while preserving source confidentiality and separating authenticity from factual accuracy. The next step is empirical-implementation, adversarial testing, newsroom pilots, and controlled audience studies are needed to determine whether the architecture can improve verifiability and accountability without creating new forms of false confidence.

Declaration on the Use of Generative AI

The author used generative artificial intelligence tools to assist with the organiza-

tion, structuring, language refinement, and grammatical correction of the manuscript. AI tools were also used to improve clarity, coherence, and presentation of the written content. The author remained responsible for reviewing, verifying, revising, and approving the final manuscript, including its arguments, interpretations, citations, references, and overall scholarly content.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives, 31,* 211-236. https://doi.org/10.1257/jep.31.2.211

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1).* National Institute of Standards and Technology.

Coalition for Content Provenance and Authenticity (C2PA) (2026). *Content Credentials: C2PA Technical Specification, Version 2.4.* https://spec.c2pa.org/

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13,* 319-340. https://doi.org/10.2307/249008

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review, 48,* 147-160. https://doi.org/10.2307/2095101

European Commission (2026). *Guidelines on Transparency Obligations for Providers and Deployers of AI Systems under Article 50 of the AI Act.* Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/

Evans, N., Edge, D., Larson, J., & White, C. (2020). News Provenance: Revealing News Text Reuse at Web-Scale in an Augmented News Search Experience. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-8). ACM. https://doi.org/10.1145/3334480.3375225

Golaszewski, E., Krawetz, N., Sherman, A. T., Zieglar, E., Matukumalli, S. K., Yus, R. et al. (2026). *Verifying Provenance of Digital Media: Why the C2PA Specifications Fall Short.* https://arxiv.org/abs/2604.24890

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly, 37,* 337-355. https://doi.org/10.25300/misq/2013/37.2.01

Hasan, H. R., & Salah, K. (2019). Combating Deepfake Videos Using Blockchain and Smart Contracts. *IEEE Access, 7,* 41596-41606. https://doi.org/10.1109/access.2019.2905689

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly, 28,* 75-106. https://doi.org/10.2307/25148625

International Press Telecommunications Council (IPTC) (2025). *Expressing Trust and Credibility Information in IPTC Standards (Version 1.1).* https://www.iptc.org/std/guidelines/trust-and-credibility/

Ivancsics, B., & Ghosh, B. (2020). *A Technology of Trust in a Trustless Information Economy: Insights into The News Provenance Project.* Brown Institute for Media Innovation. https://brown.stanford.edu/news-provenance-project/

Jensen, M. C., & Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency

Costs and Ownership Structure. *Journal of Financial Economics, 3,* 305-360. https://doi.org/10.1016/0304-405x(76)90026-x

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, University of Oxford.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press. https://doi.org/10.1017/cbo9780511807763

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences, 25,* 388-402. https://doi.org/10.1016/j.tics.2021.02.007

Schilke, O., & Reimann, M. (2025). The Transparency Dilemma: How AI Disclosure Erodes Trust. *Organizational Behavior and Human Decision Processes, 188,* Article ID: 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter, 19,* 22-36. https://doi.org/10.1145/3137597.3137600

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics, 87,* 355-374. https://doi.org/10.2307/1882010

Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of True and False News Online. *Science, 359,* 1146-1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. Free Press.

World Wide Web Consortium (W3C) (2022). *Decentralized Identifiers (DIDs) v1.0. W3C Recommendation*. https://www.w3.org/TR/did-core/

Yanes, N. A., Jitkajornwanich, K., Kee, K. F., & Perlmutter, D. D. (2026). The Original Sins of Algorithmic Authenticity Verification: Provenance Neglect and Model Collapse in AI-Mediated News and Information. *Journal of Broadcasting & Electronic Media, 70,* 496-516. https://doi.org/10.1080/08838151.2026.2691175

Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys, 53,* Article No. 109. https://doi.org/10.1145/3395046