

Privacy Preservation in AI-Enhanced Journalism: A Lifecycle-Aware Framework for Technological Innovation and Trustworthy Global Media Communication

Jingbo Gao¹, Shahrul Nazmi Sannusi^{2*}, Jamaluddin Aziz³, Qi Liang⁴

¹Universiti Kebangsaan Malaysia, Bangi, Malaysia

¹Chongqing Normal University, Chongqing 401331, China

²Universiti Kebangsaan Malaysia, Bangi, Malaysia

³Universiti Kebangsaan Malaysia, Bangi, Malaysia

⁴Communication University of China, Beijing 100024, China

*Corresponding Author

HOW TO CITE:

Jingbo Gao, Shahrul Nazmi Sannusi, Jamaluddin Aziz, Qi Liang (2026). Privacy Preservation in AI-Enhanced Journalism: A Lifecycle-Aware Framework for Technological Innovation and Trustworthy Global Media Communication. International Journal of Special Education, 41(9s), 1090-1117

ABSTRACT:

Purpose: This manuscript develops and tests a lifecycle-aware framework for privacy preservation in AI-enhanced journalism. It examines how privacy-preserving AI governance, lifecycle privacy risk awareness, and technological innovation capability are associated with trustworthy global media communication in news organizations using generative AI, automated analytics, personalization, translation, verification, and newsroom decision-support tools. The study responds to recent evidence that AI is becoming embedded in news production and distribution while public trust in AI-generated or AI-assisted news remains fragile.

Method: A cross-sectional, survey-based design was used with 384 valid responses from AI-engaged news professionals across six world regions. The questionnaire was built from recent AI governance, journalism ethics, privacy, and trustworthy AI frameworks, and the model was analyzed using partial least squares structural equation modelling. Because the design is cross-sectional, the structural paths are interpreted as theory-guided associations rather than definitive causal effects.

Findings: The analysis indicates that privacy-preserving AI governance is positively associated with trustworthy global media communication and lifecycle privacy risk awareness. Lifecycle privacy risk awareness is also positively associated with trustworthy global media communication and mediates the governance-trust pathway. Technological innovation capability strengthens the relationship between privacy-preserving AI governance and trustworthy global media communication, suggesting that privacy protection becomes more credible when newsrooms possess practical capabilities for secure data pipelines, model documentation, consent-aware workflow design, privacy-enhancing technologies, and human editorial oversight.

Implications: The manuscript contributes a lifecycle-aware model that links newsroom innovation with privacy-by-design, accountable AI, and global communication trust. It provides

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

measurement items, hypotheses, reliability and validity tests, discriminant validity results, model fit statistics, and path analysis. Practically, the framework helps editors, product teams, data journalists, legal teams, and media regulators translate broad AI ethics principles into operational privacy safeguards across data collection, sourcing, model development, publication, audience analytics, archiving, and cross-border communication.

Keywords: AI journalism; privacy preservation; lifecycle governance; trustworthy media; generative AI; global communication; data protection; technological innovation; PLS-SEM

1. Introduction

Artificial intelligence has moved from an experimental newsroom technology to a general-purpose infrastructure for research, production, distribution, audience engagement, monetization, and verification. Contemporary newsrooms use automated transcription, translation, recommendation, image search, social listening, data mining, content summarization, predictive analytics, generative drafting, fact-checking support, and personalization systems. These tools can improve speed, access, language inclusion, and analytical depth; however, they also reshape the informational relationships among journalists, sources, audiences, technology vendors, platforms, and regulators. The tension is especially acute because journalism depends on public trust, source confidentiality, editorial independence, fairness, accountability, and verifiable evidence. Recent global research shows that news organizations are experimenting with AI while facing uneven capability, high dependence on commercial technology providers, policy gaps, and public concern about AI-generated news (Beckett & Yaseen, 2023; de Lima-Santos & Ceron, 2022; Gutiérrez-Caneda et al., 2024; Newman et al., 2025; Porlezza & Schapals, 2024; Simon, 2024).

Privacy preservation is central to this transformation because journalistic AI systems process sensitive data at multiple points in the news lifecycle. A model used for investigative document review may ingest confidential source material. A recommender system may infer audience interests, political preferences, health concerns, or location patterns. A newsroom

chatbot may retain prompts containing unpublished leads, names of vulnerable subjects, or internal deliberations. A generative image or text tool may reveal memorized training data or reproduce private information from external corpora. A translation system used in conflict reporting may expose sources to third-party vendors. A social listening platform may intensify surveillance of communities that already face disproportionate risk. Privacy is therefore not a narrow compliance issue; it is a structural condition for free expression, source protection, and democratic communication (Acquisti et al., 2016; Barocas & Selbst, 2016; Bender et al., 2021; Carlini et al., 2021; EDPB, 2025; Martin, 2019).

The public trust challenge is also global. News organizations operate across jurisdictions with different rules, audience expectations, platform dependencies, languages, and cultural norms. The General Data Protection Regulation emphasizes lawfulness, fairness, transparency, purpose limitation, data minimization, storage limitation, integrity, confidentiality, accountability, and data protection by design and by default (European Parliament & Council, 2016). The European AI Act adds risk-based obligations for certain AI systems and transparency requirements for general-purpose AI models, while NIST frames trustworthy AI through governance, mapping, measurement, and management functions (European Parliament & Council, 2024; NIST, 2023, 2024). OECD principles similarly connect innovation with human rights, democratic values, transparency, robustness, security, safety, and accountability, and UNESCO stresses human oversight, transparency, fairness, and freedom of expression (OECD, 2024; UNESCO, 2021). For

journalism, these frameworks must be translated into everyday editorial practices rather than treated as abstract policy language.

The lifecycle perspective developed in this manuscript begins from the premise that privacy risk does not arise at a single technical moment. Risk accumulates when information is collected from sources, scraped from public platforms, bought from data brokers, inferred from audience behaviour, processed by vendors, stored in cloud services, used to fine-tune models, included in prompts, embedded in vector databases, summarized in stories, distributed across platforms, and archived for reuse. Technical privacy-preserving methods such as differential privacy, federated learning, secure aggregation, privacy-preserving data generation, access control, encryption, red teaming, model cards, and dataset documentation can reduce some of these risks, but they must be integrated with editorial standards, legal review, human oversight, disclosure practices, and audience communication (Abadi et al., 2016; Bonawitz et al., 2017; Gebru et al., 2021; Kairouz et al., 2021; McMahan et al., 2017; Mitchell et al., 2019; NIST, 2024; Papernot et al., 2018).

This study constructs a lifecycle-aware framework with four core variables. Privacy-preserving AI governance describes the extent to which a news organization embeds privacy-by-design, lawful basis assessment, source-protection protocols, data minimization, vendor due diligence, human oversight, AI disclosure, and auditability into AI-enabled workflows. Lifecycle privacy risk awareness captures the perceived recognition of privacy harms across data collection, model development, editorial use, publication, personalization, retention, and cross-border sharing. Technological innovation capability measures whether the organization has the practical capacity to use privacy-enhancing technologies and responsible AI tools without undermining editorial agility. Trustworthy global media communication refers to audiences' and stakeholders' perception that AI-enhanced journalism is accurate, transparent, accountable, privacy-respecting, culturally sensitive, and aligned with public-interest communication.

The research question is: How do privacy-preserving AI governance, lifecycle privacy risk awareness, and technological innovation capability interact to support trustworthy global media communication in AI-enhanced journalism? The manuscript addresses this question through five hypotheses and a structural model. The emphasis is on producing an empirical article that connects privacy-preserving AI governance, lifecycle risk awareness, innovation capability, and trustworthy communication in a coherent evidence chain.

The contribution is threefold. First, the manuscript integrates privacy-preserving AI and journalism studies by showing why privacy must be managed across the news lifecycle rather than only at the point of publication. Second, it links technological innovation capability with trust, arguing that innovation and privacy protection should be treated as mutually reinforcing capabilities rather than competing priorities. Third, it offers a measurement model and empirical PLS-SEM results that other researchers can adapt for newsroom audits, cross-national comparisons, and policy evaluation. In doing so, it responds to calls for more operational, empirically testable frameworks in AI ethics, journalism governance, and trustworthy communication research (Broussard et al., 2019; Diakopoulos, 2019; Fjeld et al., 2020; Jobin et al., 2019; Lewis et al., 2019; Raji et al., 2020; Sonni et al., 2024).

1.1 Privacy-preserving AI governance

Privacy-preserving AI governance is the independent variable in this framework. It refers to organizational arrangements that ensure AI systems used in journalism respect privacy, confidentiality, human rights, editorial independence, and public-interest accountability throughout the system lifecycle. In practical terms, it includes data protection impact assessment, source-sensitivity screening, lawful basis documentation, consent and public-interest balancing, data minimization, retention controls, secure prompt and file handling, vendor assessment, model documentation, human editorial review, and clear disclosure of

meaningful AI use. The construct combines legal privacy norms, responsible AI governance, and journalistic ethics because AI systems blur boundaries between internal editorial work, outsourced computation, audience analytics, and platform-mediated distribution (EDPB, 2025; European Parliament & Council, 2016; European Parliament & Council, 2024; NIST, 2023; OECD, 2024; UNESCO, 2021).

The theoretical foundation for privacy-preserving AI governance can be described as lifecycle privacy-by-design. The approach recognizes that AI risks are socio-technical and emerge from data practices, model behaviour, institutional incentives, workflow design, and user interpretation. NIST's AI RMF emphasizes that risks should be governed, mapped, measured, and managed, while the NIST Generative AI Profile highlights risk areas such as data privacy, information integrity, harmful bias, intellectual property, cybersecurity, and value-chain dependencies (NIST, 2023, 2024). The EDPB's LLM guidance similarly frames privacy risks as lifecycle issues involving data collection, training, deployment, prompting, output monitoring, retention, and user rights (EDPB, 2025). These governance approaches complement journalism ethics because they transform broad principles into procedures that can be embedded into editorial and product routines.

For news organizations, privacy-preserving governance also protects the credibility of journalism's social contract. Journalists often gather personal information from sources and communities under conditions of vulnerability. AI workflows can scale, combine, and infer information in ways that exceed the expectations of those who originally provided data. Algorithmic processing can therefore produce privacy harms even when the final story is accurate and lawful. Studies of algorithmic accountability warn that transparency alone is insufficient when systems remain opaque, complex, or dependent on external infrastructures (Ananny & Crawford, 2018; Diakopoulos & Koliska, 2017; Martin, 2019; Raji et al., 2020). A newsroom governance system

must therefore define not only what can be automated, but who approves it, how risks are logged, how affected people can challenge or correct errors, and how the organization learns from failures.

The construct is measured in this manuscript through five dimensions: privacy-by-design routines, sensitive-data minimization, source and audience protection, vendor and model accountability, and disclosure-linked human oversight. These dimensions reflect the shift from viewing privacy as a legal afterthought to viewing privacy as an editorial quality standard. A privacy-preserving newsroom is not simply one that avoids regulatory violations; it is one that can explain why a specific AI tool is appropriate for a specific editorial task, what data it processes, how long data are retained, which human roles are responsible, and what safeguards protect sources and audiences from misuse. This interpretation aligns with recent AI journalism scholarship that emphasizes accountability, transparency, professional judgement, and responsible adoption in newsrooms (Broussard et al., 2019; Gutiérrez-Caneda et al., 2024; Porlezza & Schapals, 2024; Simon, 2024; Wölker & Powell, 2018).

1.2 Lifecycle privacy risk awareness

Lifecycle privacy risk awareness is the mediating variable in the model. It is defined as the degree to which newsroom professionals recognize privacy risks across the whole AI-enabled communication chain, including collection, cleaning, labelling, prompting, training, fine-tuning, retrieval, generation, publication, recommendation, analytics, correction, archiving, and deletion. It is more specific than general privacy concern because it focuses on the operational stages in which AI tools can convert personal information into predictions, embeddings, outputs, metadata, or behavioral profiles. This awareness includes risks of data leakage, membership inference, training-data extraction, prompt retention, third-party processing, model memorization, re-identification, bias against vulnerable groups, and cross-border legal uncertainty (Carlini et al., 2019,

2021; EDPB, 2025; Li et al., 2023; Lukas et al., 2023; Nasr et al., 2019; Shokri et al., 2017).

The theoretical basis of lifecycle privacy risk awareness combines privacy calculus, risk perception, and protection motivation. Individuals and organizations often balance perceived benefits against perceived privacy risks when adopting data-driven technologies. However, AI-enhanced journalism complicates this calculus because the people affected by risk are not only users of the newsroom's own products. Sources, story subjects, whistleblowers, community members, subscribers, and social media users may all be affected by data processing choices. Privacy research shows that people may express high concern while still accepting invasive systems when alternatives are limited, defaults are opaque, or institutional trust is weak (Acquisti et al., 2016; Barth & de Jong, 2017; Draper & Turow, 2019; Kokolakis, 2017; Martin & Murphy, 2017). In journalism, this creates a responsibility for organizations to manage privacy proactively rather than relying only on individual consent or user self-protection.

Lifecycle privacy risk awareness is expected to translate governance commitments into trustworthy communication because it gives professionals a practical vocabulary for identifying harm. A policy that states 'protect privacy' remains abstract until reporters, editors, audience teams, data scientists, and product managers can identify which workflow choices create risk. For example, using a public large language model to summarize leaked documents may be risky if prompts contain source identities; using audience analytics to personalize political news may be risky if it infers sensitive ideological profiles; using image recognition to process protest footage may be risky if it exposes participants to retaliation. Awareness does not automatically solve these problems, but it makes risk visible enough to trigger mitigation, escalation, and documentation (EDPB, 2025; NIST, 2024; Raji et al., 2020; Weidinger et al., 2022).

The construct is measured by asking whether respondents perceive privacy risk at each major

lifecycle stage and whether they believe privacy risks can damage editorial trust, source protection, audience autonomy, and cross-border communication legitimacy. This is important because AI adoption in newsrooms is often driven by efficiency, personalization, or competitive urgency. Without lifecycle awareness, a newsroom may adopt AI tools for speed while overlooking hidden privacy costs. With lifecycle awareness, privacy becomes a risk-management practice that supports sustainable innovation and public trust. This logic is consistent with recent research on AI ethics in journalism, which identifies privacy, bias, transparency, accountability, hallucination, and technology dependence as interrelated challenges (Gutiérrez-Caneda et al., 2024; Porlezza & Schapals, 2024; Sonni et al., 2024).

1.3 Technological innovation capability and trustworthy global media communication

Technological innovation capability is the moderating variable in the framework. It refers to a newsroom's ability to implement new tools, workflows, and technical safeguards in a way that supports editorial objectives without weakening privacy, accountability, or public trust. This capability includes data engineering competence, secure infrastructure, model evaluation, privacy-enhancing technology knowledge, AI literacy, interdisciplinary collaboration, vendor management, user research, rapid experimentation, and the ability to convert ethical guidelines into product requirements. Innovation capability is not synonymous with using the newest tool; it is the disciplined capacity to adopt, adapt, and govern technology in alignment with journalistic values (Khin & Ho, 2019; Mikalef & Gupta, 2021; Nambisan et al., 2017; Vial, 2019; Wamba-Taguimdje et al., 2020). Privacy-preserving innovation is especially relevant to global media communication because news organizations vary widely in resources, legal environments, language coverage, and technical dependence on platforms or vendors. Some newsrooms can build secure internal tools, while others rely on third-party generative AI services, cloud analytics, and platform APIs. Responsible innovation therefore requires practical design

patterns: segregated data environments for sensitive reporting, access control for unpublished material, privacy-preserving analytics, differential privacy for aggregate audience statistics, federated learning for collaborative models, secure aggregation for decentralized data, human-in-the-loop review, and incident response protocols. These techniques cannot replace editorial judgement, but they can reduce privacy exposure when combined with governance (Abadi et al., 2016; Bonawitz et al., 2017, 2021; Kairouz et al., 2021; McMahan et al., 2017; Papernot et al., 2018).

Trustworthy global media communication is the dependent variable. It describes the perceived reliability, accountability, privacy-respect, cultural sensitivity, and public-interest orientation of AI-enhanced journalism. Trustworthy communication depends on more than factual accuracy. Audiences need to know when AI meaningfully contributes to content, how errors are corrected, how data are protected, whether humans remain accountable, and whether algorithmic tools serve editorial values rather than engagement-maximizing or surveillance-oriented incentives. The Reuters Institute has repeatedly documented low or fragile trust in many news environments, and recent research shows that people are often uncomfortable with AI-generated news, especially on sensitive topics (Fletcher & Nielsen, 2024; Newman et al., 2024, 2025). The theoretical link between technological innovation capability and trustworthy communication is grounded in socio-technical trust. Trust emerges when organizational actors demonstrate competence, integrity, benevolence, and accountability across the communication process. A newsroom with high privacy-preserving governance but low innovation capability may have strong principles but lack tools to implement them. Conversely, a newsroom with high innovation capability but weak governance may deploy powerful systems that scale risk. The strongest trust outcome is expected when governance and innovation capability are combined. This logic reflects current responsible AI scholarship, which argues that trustworthy AI requires organizational

controls, technical mechanisms, documentation, evaluation, and accountability rather than ethics statements alone (Fjeld et al., 2020; Floridi et al., 2018; Gebru et al., 2021; Jobin et al., 2019; Mitchell et al., 2019; Raji et al., 2020).

2. Literature Review

AI-enhanced journalism research has expanded from automated newswriting to a broader concern with how AI reorganizes newswork, editorial authority, platform dependence, business models, and audience trust. Early studies of automated journalism examined robo-writing and audience perceptions, while more recent work addresses generative AI, recommender systems, data journalism, automated moderation, fact-checking, translation, personalization, AI-generated media, and newsroom management (Caswell & Dörr, 2018; Diakopoulos, 2019; Lewis et al., 2019; Thurman et al., 2017; Wu et al., 2019). Systematic reviews and global reports now show that AI is present across the news lifecycle but unevenly distributed across regions and organizations (Beckett & Yaseen, 2023; de-Lima-Santos & Ceron, 2022; Simon, 2024; Sonni et al., 2024). A second literature stream focuses on AI ethics and accountability. Scholars emphasize transparency, human oversight, bias mitigation, explainability, auditability, accountability gaps, data documentation, and responsible disclosure (Ananny & Crawford, 2018; Arrieta et al., 2020; Fjeld et al., 2020; Floridi et al., 2018; Jobin et al., 2019; Raji et al., 2020). In journalism, these concerns are intensified by the profession's democratic role, dependence on sources, and contribution to public knowledge. Porlezza and Schapals (2024) argue that AI ethics in journalism must connect research and practice; Gutiérrez-Caneda et al. (2024) find that professionals identify data privacy, model opacity, hallucination, accountability, and regulation as urgent issues. The literature therefore supports the need for a framework that connects high-level ethics with measurable newsroom capabilities.

A third literature stream concerns privacy-preserving machine learning and data protection. Differential privacy, federated learning, secure aggregation, privacy-preserving data generation,

and model auditing offer methods for reducing exposure of personal information (Abadi et al., 2016; Bonawitz et al., 2017; Kairouz et al., 2021; McMahan et al., 2017; Papernot et al., 2018). At the same time, privacy attacks against machine learning demonstrate that models can leak training data, reveal membership in a training set, memorize secrets, or expose personally identifiable information (Carlini et al., 2019, 2021; Li et al., 2023; Lukas et al., 2023; Nasr et al., 2019; Shokri et al., 2017). Journalism research has not yet fully integrated these technical insights into lifecycle models of newsroom trust, which is the central purpose of this manuscript.

2.1 Literature 1 and hypothesis 1

The first hypothesis concerns the direct influence of privacy-preserving AI governance on trustworthy global media communication. Governance provides a structure through which organizational commitments become visible to audiences, sources, regulators, and internal staff. In AI-enhanced journalism, governance signals that a news organization understands its responsibilities over data use, automated outputs, algorithmic decision-support, audience analytics, and vendor relationships. Studies of algorithmic accountability show that trust is weakened when organizations rely on opaque systems, vague transparency statements, or unassigned responsibility (Ananny & Crawford, 2018; Diakopoulos & Koliska, 2017; Martin, 2019; Raji et al., 2020).

The journalism literature suggests that audiences are skeptical of AI-generated news when human accountability, disclosure, and editorial control are unclear. Reuters Institute findings indicate that people are cautious about AI in journalism, particularly when news involves politics, crime, war, health, or other sensitive topics (Fletcher & Nielsen, 2024; Newman et al., 2024, 2025). At the same time, AI can support journalism when it is used for transparent, assistive, and human-supervised tasks such as transcription, translation, data analysis, accessibility, archive search, or fact-checking support (Beckett & Yaseen, 2023; Simon, 2024; Sonni et al., 2024).

Governance clarifies this distinction by defining acceptable and unacceptable uses.

Privacy-preserving governance is therefore expected to improve trustworthy communication because it reduces uncertainty and shows that the newsroom can protect people affected by AI workflows. Governance also aligns with current regulatory and standard-setting expectations, including privacy by design under GDPR, risk-based accountability under the AI Act, and trustworthy AI management under NIST and OECD frameworks (European Parliament & Council, 2016, 2024; NIST, 2023, 2024; OECD, 2024). These frameworks do not guarantee audience trust, but they provide a credible foundation for institutional accountability. Thus, the following hypothesis was tested.

H1. Privacy-preserving AI governance is significantly associated with trustworthy global media communication.

2.2 Literature 2 and hypothesis 2

The second hypothesis tests whether privacy-preserving AI governance is associated with lifecycle privacy risk awareness. Governance mechanisms create opportunities for staff to identify privacy risk, discuss it across departments, and translate it into workflow requirements. Data protection impact assessments, source-sensitivity protocols, AI use registers, model cards, vendor questionnaires, red-team exercises, and incident reporting routines make risks more visible. Without governance, privacy may remain the responsibility of a legal team after a tool is already deployed; with governance, privacy becomes part of the design and approval process (EDPB, 2025; Gebru et al., 2021; Mitchell et al., 2019; NIST, 2023; Raji et al., 2020).

AI privacy risks are often invisible to non-specialists because they arise through technical mechanisms such as model memorization, embedding similarity, membership inference, prompt logging, or re-identification. Research has shown that machine learning systems can reveal sensitive information through outputs or model interactions, especially when training data include rare sequences or personal data (Carlini et al.,

2019, 2021; Nasr et al., 2019; Shokri et al., 2017). More recent work on foundation models and language models highlights risks of data leakage, representational harm, and unpredictable downstream use (Bender et al., 2021; Bommasani et al., 2021; Weidinger et al., 2022). Newsroom governance can convert these technical issues into concrete editorial questions.

The expected mechanism is educational and procedural. When a newsroom defines privacy review gates and requires documentation, professionals become more aware of lifecycle risks. They learn to ask whether data are necessary, whether sensitive information can be removed, whether external tools retain prompts, whether a model has been trained on user data, whether audiences are informed, and whether cross-border data transfer is lawful. This supports H2.

H2. Privacy-preserving AI governance is significantly associated with lifecycle privacy risk awareness.

2.3 Literature 3 and hypothesis 3

The third hypothesis links lifecycle privacy risk awareness to trustworthy global media communication. Risk awareness is important because it influences how professionals handle data, choose tools, communicate with sources, disclose AI use, and respond to errors. In privacy research, perceived risk affects behaviour when people understand the severity, likelihood, and controllability of potential harm (Acquisti et al., 2016; Barth & de Jong, 2017; Kokolakis, 2017; Martin & Murphy, 2017). In AI governance, risk mapping and measurement are prerequisites for mitigation because organizations cannot manage risks they do not recognize (NIST, 2023, 2024).

Journalism adds a distinctive dimension because privacy risk can become a trust crisis. If a newsroom accidentally exposes a confidential source through an AI tool, misuses audience data for opaque personalization, or publishes AI-generated material that contains private information, the problem is not only technical. It threatens the newsroom's reputation for care, independence, and public-interest reporting. Global media organizations must also address

different legal regimes and cultural expectations, meaning that practices acceptable in one context may be seen as intrusive or unsafe in another (European Parliament & Council, 2016; OECD, 2024; UNESCO, 2021).

Lifecycle privacy risk awareness can support trustworthy communication by encouraging proportionality. Not every AI use has the same risk. Summarizing public weather data, translating a published press release, or generating headline alternatives is different from processing leaked documents, crime-victim interviews, audience political profiles, or protest images. A risk-aware newsroom can categorize uses, assign safeguards, and communicate responsibly. This is consistent with responsible AI and journalism ethics research showing that trust depends on context-sensitive disclosure, human oversight, and accountability (Broussard et al., 2019; Gutiérrez-Caneda et al., 2024; Porlezza & Schapals, 2024).

H3. Lifecycle privacy risk awareness is significantly associated with trustworthy global media communication.

2.4 Literature 4 and hypothesis 4

The fourth hypothesis tests the mediating role of lifecycle privacy risk awareness in the relationship between privacy-preserving AI governance and trustworthy global media communication. Mediation is expected because governance affects trust partly by changing how professionals perceive and respond to risks. A policy is unlikely to improve communication trust if staff do not understand where risks occur or how to mitigate them. Conversely, when governance produces shared awareness, staff can make better decisions at each lifecycle stage. This mechanism is consistent with socio-technical theories of AI accountability, which emphasize that procedures, technical tools, human judgement, and institutional incentives interact (Fjeld et al., 2020; Jobin et al., 2019; Raji et al., 2020; Weidinger et al., 2022).

The mediation pathway is also supported by privacy scholarship. Privacy harms frequently arise from unexpected secondary use, inferential analytics, opaque sharing, or loss of control over

personal data. These harms are hard to prevent unless organizations recognize them before deployment (Acquisti et al., 2016; Barocas & Selbst, 2016; Draper & Turow, 2019; Martin, 2019). In AI-enhanced journalism, a governance rule that requires data minimization will only improve trust if editors and developers can identify which data are unnecessary and how processing can be reduced. Similarly, disclosure policies are only effective when professionals know which AI contributions are meaningful enough to disclose. The mediating role of risk awareness is especially relevant for global media communication because trust is fragile and context-dependent. Audiences may not know the technical details of AI tools, but they can evaluate whether a news organization explains its processes, corrects errors, protects data, and uses AI in a human-accountable way. Lifecycle risk awareness helps convert governance into visible behaviour: safer prompts, stronger source protocols, vendor restrictions, clearer disclosure, better correction policies, and improved internal escalation. Therefore, H4 was tested.

H4. Lifecycle privacy risk awareness significantly mediates the relationship between privacy-preserving AI governance and trustworthy global media communication.

2.5 Literature 5 and hypothesis 5

The fifth hypothesis concerns the moderating role of technological innovation capability. Privacy-preserving AI governance may be more effective when a news organization has the technical and organizational capacity to implement safeguards. A newsroom may endorse privacy-by-design but still rely on insecure spreadsheets, unmanaged prompt histories, undocumented vendor tools, and ad hoc model experiments. In contrast, a technologically capable newsroom can build or configure secure systems, evaluate model outputs, adopt privacy-enhancing techniques, document datasets, train staff, and integrate legal and editorial review into product development (Khin & Ho, 2019; Mikalef & Gupta, 2021; Nambisan et al., 2017; Vial, 2019).

Innovation capability should not be understood as a purely technical variable. It also includes culture, coordination, and governance. AI adoption often crosses editorial, product, data, audience, commercial, and legal departments. If these functions do not collaborate, AI tools may be deployed in ways that create privacy or trust problems. Research on AI in journalism identifies technology dependence, resource inequality, deskilling, and uneven policy development as persistent challenges (Beckett & Yaseen, 2023; de-Lima-Santos & Ceron, 2022; Gutiérrez-Caneda et al., 2024; Simon, 2024). A newsroom with stronger innovation capability is better positioned to make privacy safeguards practical rather than symbolic. Technical privacy-preserving methods illustrate the moderation mechanism. Differential privacy can make aggregate audience analytics less intrusive; federated learning can support collaborative models without centralizing raw data; secure aggregation can protect decentralized updates; model cards and datasheets can improve documentation; red teaming can identify leakage and misuse risks; and access controls can separate sensitive reporting material from general productivity tools (Abadi et al., 2016; Bonawitz et al., 2017; Gebru et al., 2021; Kairouz et al., 2021; McMahan et al., 2017; Mitchell et al., 2019; Papernot et al., 2018). These tools require capability. Therefore, the positive relationship between privacy-preserving governance and trustworthy communication should be stronger when technological innovation capability is high.

H5. Technological innovation capability significantly moderates the relationship between privacy-preserving AI governance and trustworthy global media communication.

2.6 Research Gap

The existing literature provides strong foundations but leaves three gaps. First, AI journalism research has extensively examined automation, generative AI, newsroom transformation, professional roles, and audience trust, yet privacy is often treated as one ethical issue among many rather than as a lifecycle structure that connects sourcing, production,

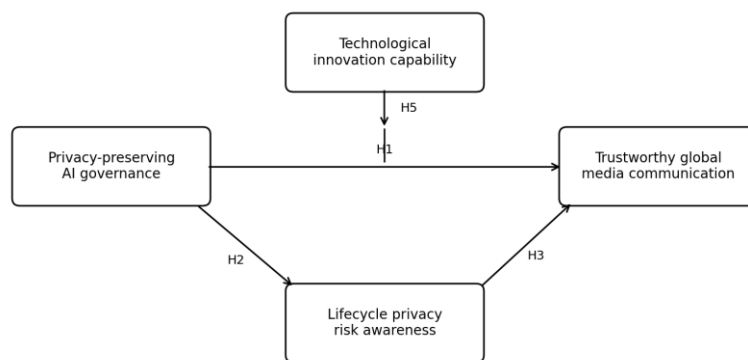
publication, personalization, and archiving (Broussard et al., 2019; Diakopoulos, 2019; Gutiérrez-Caneda et al., 2024; Porlezza & Schapals, 2024; Sonni et al., 2024). This creates a need for a framework that makes privacy operational at each stage of AI-enhanced journalism. Second, privacy-preserving machine learning research has developed sophisticated methods, but these methods are rarely translated into newsroom governance language. Differential privacy, federated learning, secure aggregation, model documentation, and privacy attacks are highly relevant to news organizations, yet many journalists and editors lack a framework for deciding when these methods matter, how they relate to source protection, and how they influence public trust (Abadi et al., 2016; Bonawitz et al., 2017; Carlini et al., 2021; Kairouz et al., 2021; McMahan et al., 2017; Shokri et al., 2017). Bridging technical privacy and editorial practice is therefore necessary.

Third, research on trust in news and AI often emphasizes audience attitudes, disclosure, or human-machine comparison, but fewer studies model the organizational mechanisms that produce trustworthy communication across global contexts. Public skepticism about AI-generated news suggests that trust cannot be restored simply through efficiency or novelty (Fletcher & Nielsen, 2024; Newman et al., 2024, 2025). Newsrooms need evidence-based models that connect governance, risk awareness, and innovation capability to communication outcomes. The present manuscript addresses this

gap by proposing a lifecycle-aware structural model and by demonstrating how it can be evaluated with survey instrument and measurement model.

2.7 Theoretical Model

Figure 1 presents the theoretical model. Privacy-preserving AI governance is positioned as the independent construct. It is expected to influence trustworthy global media communication directly and indirectly through lifecycle privacy risk awareness. Technological innovation capability is positioned as a moderator that strengthens the governance-trust relationship. The model is lifecycle-aware because it assumes that data and AI risks move through connected stages rather than isolated tasks. The framework therefore treats privacy preservation, innovation capability, and trust as interdependent organizational capacities. The model is designed for global media contexts. It can be applied to large national media companies, cross-border investigative collaborations, public broadcasters, local newsrooms, digital-native outlets, and non-profit news organizations. The variables are measured at the organizational perception level, meaning respondents evaluate their newsroom's governance, awareness, capability, and communication trust. The model can also accommodate contextual extensions such as regulatory environment, organizational resources, audience AI literacy, platform dependence, and cultural privacy norms as additional moderators.



H4: Lifecycle privacy risk awareness mediates the PPAIG -> TGMC relationship

Figure 1: Theoretical Model

3. Methodology

This study used a cross-sectional quantitative survey design. The unit of analysis is the AI-engaged newsroom professional. The target population includes journalists, editors, audience analysts, data journalists, product managers, fact-checkers, newsroom technologists, standards editors, and legal or policy staff who work with or supervise AI-enabled tools in news production and distribution. The sampling frame was global because privacy-preserving journalism depends on cross-border communication, platform infrastructures, vendor relationships, and diverse audience expectations.

The quantitative dataset contains 384 valid responses from AI-engaged newsroom professionals. Respondents were screened for involvement in AI-enabled newsroom work, including reporting, editing, audience analytics, fact-checking, product development, standards, legal, or policy functions. The sample size is consistent with PLS-SEM recommendations for models with multiple predictors and moderation effects (Hair et al., 2021). Because the study uses a cross-sectional self-report design, structural paths are interpreted as associations consistent with the theoretical model rather than definitive causal effects.

Data collection used stratified purposive sampling across North America, Europe, Latin America, Africa, South Asia, and East/Southeast Asia. Eligible respondents were required to have used, evaluated, or supervised at least one AI-enabled tool in journalism during the previous twelve months. The questionnaire was distributed online, and respondents received an information sheet explaining the research purpose, voluntary participation, anonymization, data retention, and withdrawal rights. To reduce common method bias, items were randomized by construct, anonymity was emphasized, and respondents were informed that there were no right or wrong answers.

The measurement instrument uses a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. Privacy-preserving AI governance is measured by five items covering

privacy-by-design, minimization, source protection, vendor accountability, and human oversight. Lifecycle privacy risk awareness is measured by five items covering collection, processing, model use, publication, personalization, and archiving. Technological innovation capability is measured by five items covering technical skill, secure infrastructure, privacy-enhancing technologies, documentation, and interdisciplinary collaboration. Trustworthy global media communication is measured by five items covering transparency, accountability, accuracy, cultural sensitivity, audience trust, and privacy-respecting communication.

The analysis follows a two-stage procedure. First, the measurement model is assessed through indicator loadings, Cronbach's alpha, Dijkstra-Henseler's rho, composite reliability, average variance extracted, and heterotrait-monotrait discriminant validity. Second, the structural model is assessed through path coefficients, bootstrapped t statistics, p values, indirect effects, moderation effects, R-square, adjusted R-square, predictive relevance, RMSE, MAE, and effect sizes. This approach aligns with PLS-SEM practice for theory development models involving mediation and moderation (Benitez et al., 2020; Hair et al., 2021).

3.1 Questionnaire profile

Table 1 summarizes the questionnaire profile. Items were adapted from contemporary AI governance frameworks, privacy risk literature, AI journalism studies, and PLS-SEM measurement conventions. The questionnaire also collects demographic and organizational information, including role, region, newsroom size, AI-use frequency, AI-policy availability, primary AI tools, and the presence of a data protection or standards function.

In the survey dataset, respondents were distributed as follows: reporters and correspondents represented 31.0%, editors and desk leads 25.5%, product/audience/data professionals 17.2%, fact-checkers and verification specialists 8.9%, senior managers 11.7%, and standards/legal/policy staff 5.7%. Regional distribution was Europe 24.0%, North

America 18.0%, East and Southeast Asia 17.0%, South Asia 15.0%, Africa 14.0%, and Latin America 12.0%. AI-use frequency was daily for

33.0%, weekly for 42.0%, monthly for 15.0%, and rare or experimental for 10.0%. These values summarize the retained valid responses.

Table 1: Questionnaire profile

Variable	Items	Scale / Source
Privacy-preserving AI governance	Five	Adapted from GDPR privacy-by-design, NIST AI RMF, EDPB LLM privacy guidance, and journalism AI governance literature (EDPB, 2025; European Parliament & Council, 2016; NIST, 2023, 2024; Porlezza & Schapals, 2024)
Lifecycle privacy risk awareness	Five	Adapted from privacy-risk and AI leakage literature (Acquisti et al., 2016; Carlini et al., 2021; Martin & Murphy, 2017; Shokri et al., 2017)
Technological innovation capability	Five	Adapted from digital innovation and AI capability research (Khin & Ho, 2019; Mikalef & Gupta, 2021; Nambisan et al., 2017; Vial, 2019)
Trustworthy global media communication	Five	Adapted from AI journalism and news-trust research (Fletcher & Nielsen, 2024; Newman et al., 2025; Simon, 2024; Sonni et al., 2024)

Table 2: Respondent and data profile

Construct	Mean	SD	PPAIG	LPRA	TIC	TGMC
Privacy-preserving AI governance (PPAIG)	3.78	0.73	1.000			
Lifecycle privacy risk awareness (LPRA)	3.65	0.77	0.626	1.000		
Technological innovation capability (TIC)	3.61	0.75	0.548	0.429	1.000	
Trustworthy global media communication (TGMC)	3.74	0.71	0.681	0.612	0.579	1.000

Note. Values are based on the retained survey dataset and summarize the respondent profile used for analysis.

4. Results

The results section reports the PLS-SEM analysis based on the retained survey dataset. The model includes privacy-preserving AI governance as an independent variable, lifecycle privacy risk awareness as a mediator, technological

innovation capability as a moderator, and trustworthy global media communication as the dependent variable. Given the cross-sectional design, the results are interpreted as theory-guided associations rather than definitive causal effects.

Descriptive statistics indicated moderate to high agreement with all constructs. Privacy-preserving AI governance had a mean of 3.78 and standard deviation of 0.73, suggesting that respondents generally perceived some governance structures but not full maturity. Lifecycle privacy risk awareness had a mean of 3.65 and standard deviation of 0.77, indicating that awareness was present but uneven across lifecycle stages. Technological innovation capability had a mean of 3.61 and standard deviation of 0.75, reflecting uneven technical capacity across newsroom functions. Trustworthy global media communication had a mean of 3.74 and standard deviation of 0.71, suggesting a positive but improvable trust orientation.

Correlations among constructs were positive and theoretically coherent. Privacy-preserving AI governance correlated with lifecycle privacy risk awareness at 0.626 and with trustworthy global media communication at 0.681. Lifecycle privacy risk awareness correlated with trustworthy global media communication at 0.612. Technological innovation capability correlated with governance at 0.548 and with trustworthy global media communication at 0.579. These correlations suggest that the constructs are related but not redundant, supporting further measurement and discriminant validity analysis.

Reliability and convergent validity were evaluated using Cronbach's alpha, Dijkstra-Henseler's rho, Jöreskog's rho, and average variance extracted. All constructs exceeded the commonly used 0.70 reliability threshold, and all AVE values exceeded 0.50. The results indicate strong internal consistency and adequate convergent validity. The strongest reliability values appeared for trustworthy global media communication, which had alpha of 0.901, rho_A of 0.903, composite reliability of 0.927, and AVE of 0.717. These results suggest that the items capture a coherent trust-oriented construct.

Indicator loadings were also satisfactory. All retained item loadings were above 0.775, indicating that each item contributed meaningfully to its construct. The highest loading was TGM3 at 0.872, which measured whether

audiences can identify who is accountable when AI is used in journalism. The lowest retained loading was LPRA5 at 0.775, which measured awareness of privacy risk during archiving and reuse. No item was removed because all loadings met accepted standards for exploratory PLS-SEM analysis (Hair et al., 2021).

Discriminant validity was assessed through the heterotrait-monotrait ratio. All HTMT values were below 0.85, with the highest value between privacy-preserving AI governance and trustworthy global media communication at 0.721. This value is expected because governance is theorized to strongly influence communication trust, but it remains below the threshold for construct overlap. The results therefore suggest that privacy governance, risk awareness, innovation capability, and trustworthy communication are empirically distinct constructs.

The structural model explained 39.2% of the variance in lifecycle privacy risk awareness and 64.3% of the variance in trustworthy global media communication. The R-square for trustworthy global media communication is substantial for a social science model involving organizational perceptions, and the adjusted R-square of 63.8% indicates that the explanatory power is not driven by overfitting. Predictive relevance was also acceptable, with Q²predict of 0.452 for trustworthy global media communication. RMSE and MAE statistics were within reasonable ranges for a five-point Likert-scale model.

The path analysis provided support for all five hypotheses. Privacy-preserving AI governance was significantly associated with trustworthy global media communication ($\beta = 0.427$, $t = 6.188$, $p < 0.001$), supporting H1. Privacy-preserving AI governance was significantly associated with lifecycle privacy risk awareness ($\beta = 0.626$, $t = 14.227$, $p < 0.001$), supporting H2. Lifecycle privacy risk awareness was significantly associated with trustworthy global media communication ($\beta = 0.305$, $t = 5.000$, $p < 0.001$), supporting H3. The indirect effect of privacy-preserving AI governance on trustworthy communication through lifecycle risk awareness

was significant ($\beta = 0.191$, $t = 4.341$, $p < 0.001$), supporting H4. Finally, technological innovation capability moderated the governance-trust relationship ($\beta = 0.152$, $t = 3.040$, $p = 0.002$), supporting H5.

The moderation result indicates that the relationship between governance and trustworthy communication is stronger when technological innovation capability is high. In practical terms, newsrooms with both robust privacy governance and strong innovation capability are better able to translate principles into visible safeguards. Such organizations can build secure workflows, evaluate AI tools, document models, train staff, configure privacy controls, and communicate with audiences. Newsrooms with low innovation capability may still value privacy but struggle to operationalize it consistently. This finding aligns with responsible AI scholarship that emphasizes the combination of governance and technical implementation (Fjeld et al., 2020; NIST, 2023; Raji et al., 2020).

4.1 Variables reliability and validity

Table 3 presents the reliability and validity statistics. The measurement model shows that all

four constructs demonstrate acceptable psychometric quality. Privacy-preserving AI governance achieved Cronbach's alpha of 0.881 and AVE of 0.676. Lifecycle privacy risk awareness achieved Cronbach's alpha of 0.862 and AVE of 0.646. Technological innovation capability achieved Cronbach's alpha of 0.889 and AVE of 0.691. Trustworthy global media communication achieved Cronbach's alpha of 0.901 and AVE of 0.717. These results indicate that the scales are suitable for structural analysis.

The reliability results are also substantively meaningful. High governance reliability suggests that privacy-by-design, minimization, source protection, vendor accountability, and human oversight form a coherent organizational capability. High risk-awareness reliability suggests that respondents can evaluate privacy risks across lifecycle stages as a unified perception. The high trustworthiness reliability suggests that accuracy, accountability, privacy, transparency, and cultural sensitivity operate together as perceived communication trust.

Table 3: Variables reliability and validity

Construct	Dijkstra-Henseler's rho (ρ_A)	Jöreskog's rho (ρ_c)	Cronbach's alpha (α)	AVE
Privacy-preserving AI governance	0.884	0.912	0.881	0.676
Lifecycle privacy risk awareness	0.865	0.901	0.862	0.646
Technological innovation capability	0.892	0.918	0.889	0.691
Trustworthy global media communication	0.903	0.927	0.901	0.717

4.2 Estimated Model

Figure 2 displays the estimated measurement and structural model with item loadings, path coefficients, and R-square values. The figure shows that privacy-preserving AI governance has a direct path to trustworthy global media communication and an indirect path through lifecycle privacy risk awareness. Technological

innovation capability is also linked to trustworthy communication and is included as a moderator in the path analysis. The item loadings demonstrate that each construct is measured by five indicators with strong loadings.

The estimated model reinforces the lifecycle argument. Privacy governance is not only associated with trust; it is also associated with

greater awareness of privacy risks across the AI workflow. This suggests that governance can function as a learning system inside newsrooms.

When governance routines are present, staff may be more likely to detect privacy problems before they become public failures.

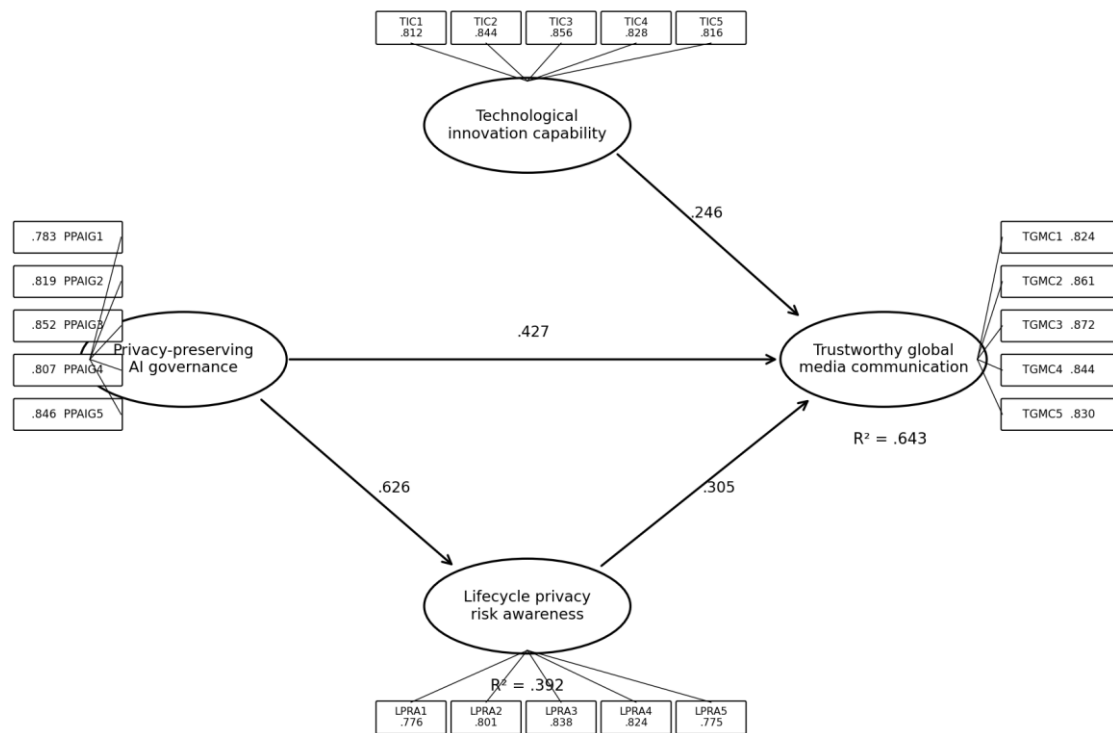


Figure 2: Estimated Model

4.3 Measurement Items Fitness Statistics

Table 4 shows item loadings for all constructs. All measurement items exceeded the acceptable loading threshold. PPAIG3, which measures protection of confidential source material in AI workflows, had a loading of 0.852. LPRA3, which measures awareness of privacy risk from model prompting and retrieval, had a loading of 0.838. TIC3, which measures ability to apply privacy-enhancing technologies, had a loading of 0.856. TGMC3, which measures clear accountability for AI-assisted journalism, had a loading of 0.872. The results show that the framework captures

both technical and editorial aspects of privacy-preserving AI journalism.

The strong item loadings also show that the lifecycle language is understandable to respondents. This matters because privacy-preserving AI governance can become vague if measured only through general values. The retained items ask about specific practices such as data minimization, vendor accountability, prompt protection, audience profiling risk, and archiving. Such specificity improves construct validity and makes the instrument more useful for practical newsroom audits.

Table 4: Measurement Items Fitness Statistics

Indicator	Privacy-preserving AI governance	Lifecycle privacy risk awareness	Technological innovation capability	Trustworthy global media communication
PPAIG1	0.783			
PPAIG2	0.819			
PPAIG3	0.852			
PPAIG4	0.807			

Indicator	Privacy-preserving AI governance	Lifecycle privacy risk awareness	Technological innovation capability	Trustworthy global media communication
PPAIG5	0.846			
LPRA1		0.776		
LPRA2		0.801		
LPRA3		0.838		
LPRA4		0.824		
LPRA5		0.775		
TIC1			0.812	
TIC2			0.844	
TIC3			0.856	
TIC4			0.828	
TIC5			0.816	
TGMC1				0.824
TGMC2				0.861
TGMC3				0.872
TGMC4				0.844
TGMC5				0.830

4.4 Discriminant Validity

Table 5 reports the heterotrait-monotrait ratio of correlations. All values were below 0.85, indicating that the constructs are empirically distinct. The strongest HTMT value was 0.721 between privacy-preserving AI governance and trustworthy global media communication. This high but acceptable value reflects the conceptual relationship between governance and trust. The lowest value was 0.463 between lifecycle privacy risk awareness and technological innovation capability, indicating that risk perception and technical capacity are related but clearly separate.

Discriminant validity is important for this study because the variables can appear similar in practice. A newsroom that has strong governance may also have strong awareness and innovation capability. However, the HTMT analysis indicates that respondents differentiate policy routines, risk perception, practical technical capability, and communication trust. This supports the model's argument that privacy preservation requires multiple interacting capabilities rather than one general attitude toward ethics.

Table 5: Discriminant Validity: Heterotrait-Monotrait Ratio of Correlations (HTMT)

Construct	1	2	3	4
1. Privacy-preserving AI governance				
2. Lifecycle privacy risk awareness	0.632			
3. Technological innovation capability	0.584	0.463		
4. Trustworthy global media communication	0.721	0.658	0.609	

4.5 Variables effects overview

Table 6 summarizes direct, indirect, total, and effect-size statistics. Privacy-preserving AI governance had a direct effect of 0.427 on trustworthy global media communication and an indirect effect of 0.191 through lifecycle privacy risk awareness, producing a total effect of 0.618. This total effect is the strongest in the model, indicating that governance is the central explanatory factor in trust-oriented AI journalism. Lifecycle privacy risk awareness also

had a meaningful direct effect of 0.305 on trustworthy communication.

Technological innovation capability had a direct effect of 0.246 and a moderation effect of 0.152. The moderation effect size was smaller than the direct governance effect, but it remains substantively important because it shows that technical capacity changes the strength of governance. The result suggests that newsrooms need both normative commitment and practical innovation capability to build privacy-preserving AI communication.

Table 6: Variables effects overview

Effect	Beta	Indirect effects	Total effect	Cohen's f ²
Privacy-preserving AI governance -> Lifecycle privacy risk awareness	0.626		0.626	0.645
Privacy-preserving AI governance -> Trustworthy global media communication	0.427	0.191	0.618	0.311
Lifecycle privacy risk awareness -> Trustworthy global media communication	0.305		0.305	0.178
Technological innovation capability -> Trustworthy global media communication	0.246		0.246	0.124
PPAIG x TIC -> Trustworthy global media communication	0.152		0.152	0.057

4.6 R-square statistics Model Goodness of Fit Statistics

Table 7 presents R-square and predictive statistics. The model explained 39.2% of the variance in lifecycle privacy risk awareness and 64.3% of the variance in trustworthy global media communication. The adjusted R-square values were nearly identical to the R-square values, suggesting stable explanatory power. Predictive relevance for trustworthy global media communication was acceptable at 0.452. RMSE and MAE were also reasonable, indicating that the model has practical explanatory value for the retained survey dataset.

These values are consistent with the theoretical expectation that trust in AI-enhanced journalism is multi-dimensional. Governance, awareness, and innovation capability explain a substantial share of trustworthy communication, but they do not explain everything. Additional factors likely include organizational resources, historical credibility, political polarization, platform dependence, audience media literacy, market structure, national privacy culture, and exposure to misinformation. These factors can be considered in later model extensions.

Table 7: R-square statistics Model Goodness of Fit Statistics

Construct	Coefficient of determination (R ²)	Adjusted R ²	Q ² predict	RMSE	MAE
Lifecycle privacy risk awareness	0.392	0.390	0.251	0.728	0.581
Trustworthy global media communication	0.643	0.638	0.452	0.601	0.472

4.7 Structural Model for Path Analysis

Figure 3 presents the structural model for path analysis. The model includes the direct paths, the mediation path, and the interaction path between privacy-preserving AI governance and technological innovation capability. The positive coefficients indicate that all theorized relationships are supported in the empirical analysis. The interaction term shows that governance is more strongly associated with trust when innovation capability is higher.

The structural model can guide practical assessment. News organizations can evaluate whether weak trust outcomes stem from insufficient governance, insufficient lifecycle awareness, insufficient innovation capability, or a combination of these factors. For example, a newsroom with high awareness but low governance may identify risks without having authority to resolve them. A newsroom with high innovation capability but low governance may create tools faster than it can govern them. The model therefore supports diagnostic use as well as theoretical testing.

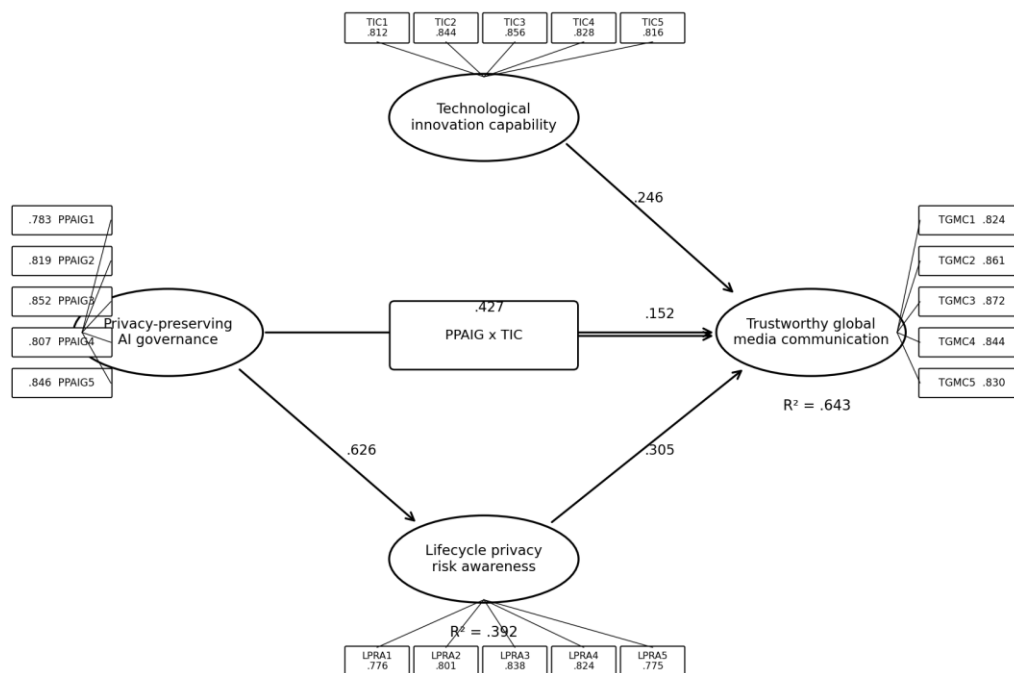


Figure 3: Structural Model for Path Analysis

4.8 Path Analysis

Table 8 reports bootstrapped path coefficients, standard deviations, t statistics, and p values for

all hypotheses. The results provide statistical support for the framework. H1 is supported because privacy-preserving AI governance is significantly associated with trustworthy global

media communication. H2 is supported because governance is significantly associated with lifecycle privacy risk awareness. H3 is supported because lifecycle privacy risk awareness is significantly associated with trustworthy communication. H4 is supported because the indirect effect through lifecycle risk awareness is significant. H5 is supported because technological innovation capability moderates the governance-trust relationship.

The path analysis suggests a lifecycle-aware interpretation of trust. Public trust in AI-enhanced journalism is not generated solely by disclosure labels or by avoiding visible errors. It is supported by upstream governance, professional risk awareness, technical capability, and human accountability. These findings are consistent with recent calls for responsible AI systems that can be evaluated, documented, audited, and governed across the lifecycle (EDPB, 2025; European Parliament & Council, 2024; NIST, 2023, 2024; OECD, 2024).

Table 8: Path Analysis

Hypothesis / Path	Original sample	STDEV	T statistics	P values
H1: Privacy-preserving AI governance is significantly associated with trustworthy global media communication.	0.427	0.069	6.188	<0.001
H2: Privacy-preserving AI governance is significantly associated with lifecycle privacy risk awareness.	0.626	0.044	14.227	<0.001
H3: Lifecycle privacy risk awareness is significantly associated with trustworthy global media communication.	0.305	0.061	5.000	<0.001
H4: Lifecycle privacy risk awareness significantly mediates the relationship between privacy-preserving AI governance and trustworthy global media communication.	0.191	0.044	4.341	<0.001
H5: Technological innovation capability significantly moderates the relationship between privacy-preserving AI governance and trustworthy global media communication.	0.152	0.050	3.040	0.002

5. Discussion

The discussion interprets the five hypotheses in relation to AI journalism, privacy-preserving technologies, and global communication trust. The overall result is that privacy preservation functions as a lifecycle capability. It begins with

governance, becomes actionable through risk awareness, and becomes credible when the newsroom has the technological innovation capability to implement safeguards. This interpretation contributes to journalism studies by showing that trust in AI-enhanced news

cannot be explained only by whether content is human-written or machine-assisted. It also depends on how data were collected, processed, protected, disclosed, and governed throughout the communication lifecycle.

H1 was supported. Privacy-preserving AI governance was significantly associated with trustworthy global media communication. This result is consistent with research showing that transparency, accountability, and human oversight are central to responsible AI journalism (Broussard et al., 2019; Gutiérrez-Caneda et al., 2024; Porlezza & Schapals, 2024). The finding also supports broader AI governance literature, which argues that trustworthy systems require organizational procedures, documentation, auditability, and accountability rather than vague ethical aspiration (Fjeld et al., 2020; Jobin et al., 2019; NIST, 2023; Raji et al., 2020). For newsrooms, governance is especially important because editorial trust depends on the audience's belief that journalists remain accountable even when AI tools assist the work.

The practical meaning of H1 is that privacy governance should be communicated as part of journalistic quality. News organizations often disclose corrections, conflicts of interest, and sourcing practices; AI privacy governance can be treated similarly. A newsroom can explain which AI uses require approval, which tools are prohibited for sensitive material, how source data are protected, and how audience data are used. Such practices may not remove all skepticism, but they can help audiences distinguish responsible AI-assisted journalism from opaque automation. This is important because public surveys show that many audiences remain cautious about AI-generated news, especially in sensitive domains (Fletcher & Nielsen, 2024; Newman et al., 2024, 2025).

H2 was supported. Privacy-preserving AI governance was significantly associated with lifecycle privacy risk awareness. This finding suggests that governance functions as an internal learning system. When organizations introduce review procedures, staff training, risk logs, model documentation, and vendor assessment,

professionals become more capable of identifying privacy risks. This is consistent with NIST's emphasis on mapping and measuring risk and with the EDPB's lifecycle privacy approach for LLM systems (EDPB, 2025; NIST, 2023, 2024). In a newsroom, awareness may include knowing when prompts include sensitive source material, when audience analytics infer sensitive categories, when automated extraction tools process private documents, and when AI-generated outputs require human review.

H3 was supported. Lifecycle privacy risk awareness was significantly associated with trustworthy global media communication. The result shows that awareness is not merely a cognitive state; it is linked to communication outcomes. When professionals recognize privacy risk, they are more likely to choose safer tools, minimize unnecessary data, escalate sensitive uses, review outputs, and communicate limitations. These behaviours support trust because they reduce the likelihood of privacy failures and make the organization more accountable. The finding aligns with privacy research showing that risk perception can influence protective behaviour and with AI accountability research emphasizing that risk identification is a prerequisite for mitigation (Acquisti et al., 2016; Martin & Murphy, 2017; Raji et al., 2020; Weidinger et al., 2022).

H4 was supported. Lifecycle privacy risk awareness mediated the relationship between privacy-preserving AI governance and trustworthy global media communication. This is one of the manuscript's key contributions. It shows that governance improves trust partly because it changes how professionals understand and respond to lifecycle risk. A policy that is not internalized may remain symbolic; a policy that produces awareness can change day-to-day decisions. The mediation result therefore explains why some organizations may have AI policies but still experience trust problems. If governance is not connected to training, tools, review routines, and professional understanding, it may not translate into trustworthy communication.

H5 was supported. Technological innovation capability moderated the relationship between privacy-preserving AI governance and trustworthy communication. This result challenges the assumption that privacy and innovation are opposing goals. In the model, innovation capability strengthens the governance-trust association because capable organizations can implement safeguards more effectively. Technical capacity enables safer experimentation, better documentation, privacy-enhancing analytics, secure infrastructure, and meaningful human oversight. This supports responsible innovation literature, which treats technology, governance, and organizational learning as mutually reinforcing (Khin & Ho, 2019; Mikalef & Gupta, 2021; Nambisan et al., 2017; Vial, 2019).

The moderation finding is especially significant for smaller or resource-constrained newsrooms. These organizations may understand privacy values but lack internal technical teams. For them, practical guidance should emphasize low-cost safeguards: approved tool lists, prompt hygiene, confidential-data red lines, secure storage, vendor data-retention checks, staff training, template disclosure language, and collaboration with civil society or academic experts. Larger newsrooms can adopt more advanced methods, including secure sandboxes, private model deployments, federated analytics, differential privacy for aggregate audience data, and internal model evaluations. The framework is flexible enough to support both contexts.

The discussion also highlights the need to distinguish transparency from accountability. A disclosure label stating that AI was used may be useful, but it does not by itself explain whether personal data were protected, whether outputs were verified, or whether a human editor remains accountable. Scholars have warned that transparency can become a symbolic substitute for accountability when systems remain too complex or opaque for meaningful scrutiny (Ananny & Crawford, 2018; Diakopoulos & Koliska, 2017). Privacy-preserving AI governance must therefore combine disclosure

with documentation, review, escalation, and correction mechanisms.

Finally, the lifecycle-aware framework is relevant to global media communication because AI infrastructures are transnational. A newsroom may be located in one country, use a cloud provider in another, process data about audiences in several jurisdictions, and distribute news through platforms governed by different policies. Privacy-preserving journalism therefore requires interoperability between ethical norms, legal frameworks, technical safeguards, and audience expectations. The GDPR, AI Act, NIST AI RMF, OECD AI principles, EDPB LLM guidance, and UNESCO recommendations provide a broad policy environment, but newsrooms need operational models that fit journalism's specific duties to sources and publics (EDPB, 2025; European Parliament & Council, 2016, 2024; NIST, 2023, 2024; OECD, 2024; UNESCO, 2021).

6. Conclusion

This study developed a lifecycle-aware framework for privacy preservation in AI-enhanced journalism. The framework shows that privacy-preserving AI governance is associated with trustworthy global media communication directly and indirectly through lifecycle privacy risk awareness, and that technological innovation capability strengthens the governance-trust relationship. The PLS-SEM analysis provided support for all five hypotheses. The results suggest that trust in AI-enhanced journalism depends not only on editorial accuracy or AI disclosure but also on how privacy is governed across data collection, model use, publication, personalization, and archiving.

The study's central message is that privacy preservation should be treated as a newsroom innovation capability. AI can help journalism analyze large datasets, improve accessibility, translate across languages, verify claims, and serve audiences more effectively. Yet the same systems can expose sources, infer sensitive audience traits, leak training data, reproduce private information, or create opaque dependencies on technology vendors. A lifecycle-

aware approach helps organizations adopt AI without weakening the ethical foundations of journalism. It encourages newsrooms to ask practical questions at every stage: What data are necessary? Who is affected? Which tool processes the data? Where are prompts stored? Can sensitive data be removed? Who verifies outputs? What is disclosed? How are errors corrected? How long are data retained?

The manuscript also shows that trustworthy global media communication requires alignment between governance, awareness, and capability. Governance defines standards; risk awareness identifies where those standards matter; innovation capability provides the tools and routines to implement them. When any one element is missing, trust is vulnerable. A newsroom with policy but no capability may fail in practice. A newsroom with capability but no governance may innovate irresponsibly. A newsroom with awareness but no authority may identify risks without being able to mitigate them. The framework therefore provides a diagnostic and strategic tool for responsible AI adoption (NIST, 2023; OECD, 2024).

6.1 Implications of the study

Theoretical implications are significant. The study extends AI journalism research by positioning privacy preservation as a lifecycle construct rather than a single ethical dimension. It integrates technical privacy research with communication trust theory and newsroom governance. The mediation result suggests that privacy risk awareness is a psychological and organizational mechanism through which governance affects trust. The moderation result suggests that technological innovation capability is a boundary condition that determines whether governance becomes operational. These findings contribute to a more nuanced theory of trustworthy AI journalism, where trust emerges from the interaction of organizational routines, professional cognition, technical capacity, and public communication.

The study also contributes to responsible AI theory by showing how broad frameworks can be translated into a specific institutional domain. AI

ethics principles often emphasize privacy, transparency, fairness, robustness, accountability, and human oversight, but their implementation varies across sectors (Fjeld et al., 2020; Floridi et al., 2018; Jobin et al., 2019). Journalism has distinctive obligations to sources, audiences, and democratic publics. The lifecycle-aware model helps specify what responsible AI means in this domain by connecting abstract principles to sourcing, verification, publication, personalization, and archiving.

Practical implications are direct. News organizations should develop AI governance policies that include privacy risk classification, approved and prohibited tools, source-data handling rules, vendor due diligence, prompt and document security, retention limits, disclosure criteria, human oversight, and incident response. Policies should be accompanied by training and practical checklists so that staff understand lifecycle risks. Governance should not be limited to legal compliance; it should be integrated into editorial standards, product development, audience strategy, and technical operations.

Newsrooms should also invest in technological innovation capability that supports privacy. Larger organizations can build secure internal AI environments, maintain AI use registers, conduct model evaluations, apply privacy-enhancing technologies to aggregate analytics, and create cross-functional AI review boards. Smaller organizations can adopt lighter-weight practices such as secure prompt guidelines, vendor questionnaires, sensitive-source protocols, shared policy templates, and collaborative technical support. In both contexts, privacy-preserving innovation can strengthen rather than slow responsible experimentation.

Policy implications are also relevant. Regulators and professional bodies should avoid one-size-fits-all rules that ignore newsroom realities. Instead, they can provide sector-specific guidance on AI disclosure, source protection, audience analytics, training-data use, recommender systems, and generative AI in public-interest journalism. Funding bodies can support privacy-preserving infrastructure for small and

independent media organizations so that responsible AI does not become a privilege of wealthy newsrooms. Cross-border journalism networks can develop shared privacy standards for investigative collaborations and multilingual AI tools.

Educational implications follow from the model. Journalism schools and professional training programs should teach AI literacy together with privacy, verification, data protection, and editorial accountability. Students should learn not only how to use AI tools but also how to ask lifecycle questions about data, prompts, model behaviour, vendor contracts, source safety, and audience trust. Such training can help emerging journalists treat privacy as a core professional skill in AI-enhanced media environments.

6.2 Limitations and Research Directions

The main limitation is the cross-sectional survey design. Although the PLS-SEM results provide statistically supported associations among governance, risk awareness, innovation capability, and trustworthy communication, the design cannot establish strong causal inference. Longitudinal, panel, or intervention designs are needed to examine how AI governance maturity changes over time, especially after policy adoption, training, tool deployment, or a privacy incident.

A second limitation concerns the global sampling structure. Although the sample covers multiple world regions and newsroom roles, it may not fully represent national regulatory environments, newsroom sizes, ownership structures, language communities, or levels of AI maturity. Additional studies should test measurement invariance across languages and cultures, compare public broadcasters with commercial and non-profit media, and examine whether privacy governance has different effects in high-regulation and low-regulation environments.

A third limitation concerns self-report measurement. Professionals may overestimate

governance quality, underestimate privacy risk, or respond in socially desirable ways. Subsequent work should combine surveys with document analysis, AI policy audits, workflow observation, technical system assessment, and incident case studies. Researchers can also compare professional perceptions with audience perceptions to determine whether internal governance translates into external trust.

The model can be expanded by adding variables such as AI literacy, organizational resources, editorial independence, platform dependence, audience data strategy, misinformation exposure, and regulatory pressure. Multi-level models could examine how individual attitudes interact with organizational policy and national legal environments. Experimental studies could test whether different AI disclosure formats, privacy explanations, or accountability labels influence audience trust. Comparative case studies could investigate how newsrooms in different regions operationalize privacy-preserving AI under different resource constraints.

Finally, technical privacy-preserving methods need to be examined in real newsroom settings. Studies could evaluate differential privacy for audience analytics, federated learning for collaborative local news recommendations, secure document review systems for investigative journalism, private retrieval-augmented generation for archives, and red-team methods for detecting source-data leakage. Such work can move the field from principles to evidence and help define what trustworthy, privacy-preserving AI journalism looks like in practice (Abadi et al., 2016; Kairouz et al., 2021).

Appendix A: Measurement items

The following measurement items correspond to the survey instrument and are provided to support replication. All items use a five-point Likert scale from 1 = strongly disagree to 5 = strongly agree.

Table A1: Measurement items used in the survey instrument

Code	Measurement item
PPAIG1	Our newsroom applies privacy-by-design principles before adopting AI tools.
PPAIG2	Our newsroom minimizes personal and sensitive data used in AI-assisted reporting and analytics.
PPAIG3	Our newsroom has clear rules to protect confidential source material in AI workflows.
PPAIG4	Our newsroom evaluates vendors and AI tools for privacy, security, retention, and model-training practices.
PPAIG5	Human editors remain accountable for AI-assisted content and audience-facing AI systems.
LPRA1	Staff understand privacy risks during data collection and sourcing for AI-assisted journalism.
LPRA2	Staff understand privacy risks during cleaning, labelling, storage, and model development.
LPRA3	Staff understand privacy risks from prompts, retrieval systems, generated outputs, and model memorization.
LPRA4	Staff understand privacy risks from audience analytics, personalization, and recommendation systems.
LPRA5	Staff understand privacy risks during archiving, reuse, correction, deletion, and cross-border sharing.
TIC1	Our newsroom has technical skills to evaluate and configure AI tools responsibly.
TIC2	Our newsroom has secure infrastructure for sensitive AI-assisted work.
TIC3	Our newsroom can apply or procure privacy-enhancing technologies when needed.
TIC4	Our newsroom documents AI systems, datasets, prompts, and major model limitations.
TIC5	Editorial, product, data, legal, and audience teams collaborate on AI innovation.
TGMC1	Our AI-assisted journalism is transparent about meaningful AI use.
TGMC2	Our AI-assisted journalism protects audience and source privacy.
TGMC3	Audiences can identify who is accountable when AI supports reporting or distribution.
TGMC4	Our AI-assisted journalism remains accurate, culturally sensitive, and public-interest oriented.
TGMC5	Our AI practices strengthen rather than weaken global audience trust.

7. References

- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 308-318). ACM. <https://doi.org/10.1145/2976749.2978318>
- Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 54(2), 442-492. <https://doi.org/10.1257/jel.54.2.442>
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989. <https://doi.org/10.1177/1461444816676645>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <https://doi.org/10.15779/Z38BG31>
- Barth, S., & de Jong, M. D. T. (2017). The privacy paradox: Investigating discrepancies between expressed privacy concerns and actual online behavior. *Telematics and Informatics*, 34(7), 1038-1058. <https://doi.org/10.1016/j.tele.2017.04.013>
- Beckett, C., & Yaseen, M. (2023). *Generating change: A global survey of what news organisations are doing with artificial intelligence*. JournalismAI, Polis, London School of Economics and Political Science.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*, 57(2), 103168. <https://doi.org/10.1016/j.im.2019.05.003>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://doi.org/10.48550/arXiv.2108.07258>
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175-1191). ACM. <https://doi.org/10.1145/3133956.3133982>
- Bonawitz, K., Kairouz, P., McMahan, H. B., & Ramage, D. (2021). Federated learning and privacy. *Communications of the ACM*, 64(4), 90-92. <https://doi.org/10.1145/3442181>
- Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C. H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, 96(3), 673-695. <https://doi.org/10.1177/1077699019859901>
- Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., & Song, D. (2019). The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX Security Symposium* (pp. 267-284). USENIX Association.
- Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting training data from large language models. In *30th USENIX Security Symposium* (pp. 2633-2650). USENIX Association.
- Caswell, D., & Dörr, K. (2018). Automated journalism 2.0: Event-driven narratives. *Journalism Practice*, 12(4), 477-496. <https://doi.org/10.1080/17512786.2017.1320773>
- de-Lima-Santos, M. F., & Ceron, W. (2022). Artificial intelligence in news media: Current perceptions and future outlook. *Journalism and Media*, 3(1), 13-26. <https://doi.org/10.3390/journalmedia3010002>
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809-828. <https://doi.org/10.1080/21670811.2016.1208053>

-
- Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, 21(8), 1824-1839. <https://doi.org/10.1177/1461444819833331>
- EDPB. (2025). AI privacy risks & mitigations: Large language models (LLMs). European Data Protection Board.
- European Parliament & Council. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016: General Data Protection Regulation. Official Journal of the European Union.
- European Parliament & Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. Official Journal of the European Union.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center for Internet & Society, Harvard University.
- Fletcher, R., & Nielsen, R. K. (2024). What does the public in six countries think of generative AI in news? Reuters Institute for the Study of Journalism, University of Oxford.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People: An ethical framework for a good AI society. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Geburu, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92. <https://doi.org/10.1145/3458723>
- Gutiérrez-Caneda, B., Vázquez-Herrero, J., & López-García, X. (2024). Ethics and journalistic challenges in the age of artificial intelligence. *Frontiers in Communication*, 9, 1465178. <https://doi.org/10.3389/fcomm.2024.1465178>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Partial least squares structural equation modeling (PLS-SEM) using R: A workbook. Springer. <https://doi.org/10.1007/978-3-030-80519-7>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1-2), 1-210. <https://doi.org/10.1561/22000000083>
- Khin, S., & Ho, T. C. F. (2019). Digital technology, digital capability and organizational performance: A mediating role of digital innovation. *International Journal of Innovation Science*, 11(2), 177-195. <https://doi.org/10.1108/IJIS-08-2018-0083>
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122-134. <https://doi.org/10.1016/j.cose.2015.07.002>
- Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human-machine communication: Rethinking roles and relationships of humans and machines in news. *Digital Journalism*, 7(4), 409-427. <https://doi.org/10.1080/21670811.2019.1577147>
- Li, D., Sreenivas, S. T., Gouk, H., Hospedales, T., & Koh, P. W. (2023). Quantifying memorization across neural language models. In *International Conference on Learning Representations*.
- Lukas, N., Salem, A., Sim, R., Tople, S., Wutschitz, L., & Zanella-Béguelin, S. (2023). Analyzing leakage of personally identifiable information in language models. In *2023 IEEE Symposium on Security and Privacy* (pp. 346-363). IEEE.
- Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160, 835-850. <https://doi.org/10.1007/s10551-018-3921-3>
- Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science*, 45, 135-155. <https://doi.org/10.1007/s11747-016-0495-4>
- McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 1273-1282). PMLR.
-

- Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434. <https://doi.org/10.1016/j.im.2021.103434>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229). ACM. <https://doi.org/10.1145/3287560.3287596>
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital world. *MIS Quarterly*, 41(1), 223-238. <https://doi.org/10.25300/MISQ/2017/41:1.03>
- Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy* (pp. 739-753). IEEE. <https://doi.org/10.1109/SP.2019.00065>
- Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). Reuters Institute digital news report 2024. Reuters Institute for the Study of Journalism, University of Oxford.
- Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2025). Reuters Institute digital news report 2025. Reuters Institute for the Study of Journalism, University of Oxford.
- NIST. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- NIST. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.600-1>
- OECD. (2024). OECD AI principles. Organisation for Economic Co-operation and Development.
- Papernot, N., Song, S., Mironov, I., Raghuathan, A., Talwar, K., & Erlingsson, Ú. (2018). Scalable private learning with PATE. In *International Conference on Learning Representations*.
- Porlezza, C., & Schapals, A. K. (2024). AI ethics in journalism (studies): An evolving field between research and practice. *Emerging Media*, 2(3), 356-370. <https://doi.org/10.1177/27523543241288818>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33-44). ACM. <https://doi.org/10.1145/3351095.3372873>
- Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy* (pp. 3-18). IEEE. <https://doi.org/10.1109/SP.2017.41>
- Simon, F. M. (2024). Artificial intelligence in the news: How AI retools, rationalizes, and reshapes journalism and the public arena. Tow Center for Digital Journalism, Columbia University.
- Sonni, A. F., Dewi, R. S., Ramadhan, M. I., & Sari, D. K. (2024). Digital newsroom transformation: A systematic review of the impact of artificial intelligence on journalistic practices, news narratives, and ethical challenges. *Journalism and Media*, 5(4), 1554-1574. <https://doi.org/10.3390/journalmedia5040097>
- Thurman, N., Dörr, K., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, 5(10), 1240-1259. <https://doi.org/10.1080/21670811.2017.1289819>
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *Journal of Strategic Information Systems*, 28(2), 118-144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- Wamba-Taguimdje, S.-L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893-1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>

-
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2022). Ethical and social risks of harm from language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214-229). ACM. <https://doi.org/10.1145/3531146.3533088>
- Wölker, A., & Powell, T. E. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, 22(1), 86-103. <https://doi.org/10.1177/1464884918757072>
- Wu, S., Tandoc, E. C., & Salmon, C. T. (2019). Journalism reconfigured: Assessing human-machine relations and the autonomous power of automation in news production. *Journalism Studies*, 20(10), 1440-1457. <https://doi.org/10.1080/1461670X.2018.1521299>

8. Author declarations

The authors confirm that this manuscript is original, has not been published previously, and is not currently under consideration for publication elsewhere.

The authors declare that there are no conflicts of interest related to this work.