Ethical AI Communication and Public Trust: Examining the Role of Transparency in Digital Media

Mohammad Nawab Turan[1], Musawer Hakimi[2], Hamayoon Ghafory[3], Ahmad Jamy Kohistani[4*]

[1]Computer Science Faculty, Muğla Sıtkı Koçman University, Turke
[2]Computer Science Department, Samangan University, Samangan, Afghanistan
[3]Computer Science Faculty, Kabul Education University, Kabul, Afghanistan
[4]Computer Science Faculty, Kabul Polytechnic University, Kabul, Afghanistan
Corresponding Author Email Address: jamy@kpu.edu.af

ARTICLE INFO

DOI:
https://doi.org/10.38043/commusty.v5i1.7780

Article history:
Received: 2026-01-21
Revised: 2026-03-22
Accepted: 15-04-30

Keywords.
AI communication;
public trust;
transparency;
digital media;
algorithmic accountability;
media ethics

ABSTRACT

The rise of Digital Media incorporating Artificial Intelligence raises important issues related to public trust; algorithmic transparency; and ethical standards for communication. In this study, a mixed methods approach was used to explore AI transparency disclosure practices and their association with public trust in digital media among a sample of 1,247 adults across six countries. The study utilized a survey instrument developed for this study, as well as qualitative thematic analysis of 614 open-ended responses. Results showed that disclosure of transparency practices accounted for the largest variation in public trust (beta = .38, p < .001). It also revealed moderating effects of digital media literacy and media skepticism on trust. A confirmatory structural equation model with a reasonable fit (CFI = .96, RMSEA = .048) provided validation of the Integrated Ethical AI Communication Framework. The study also indicated that structured transparency mechanisms; formalized oversight of editorial content; and explicit policies holding algorithms accountable to ensure transparency would substantially increase public trust across all demographic subgroups viewed to have access to digital news services. The findings indicate, in particular, that digital newsrooms should pair AI transparency disclosures with active media literacy initiatives, and that regulators should mandate minimum content standards for algorithmic disclosure rather than mere disclosure presence, in order to realise the full trust-building potential of ethical AI communication.

This is an open access article under the CC–BY-SA license.

1. Introduction

Few developments in contemporary media have provoked as much public deliberation as the rapid deployment of artificial intelligence in news production, content curation, and audience engagement systems. Across the globe, digital news platforms increasingly rely on machine learning algorithms to select, rank, and even generate editorial content, yet the communicative norms and ethical standards governing these processes remain fragmented, inconsistently applied, and largely opaque to ordinary readers (Diakopoulos, 2019; Thurman et al., 2019). The question of whether, and under what conditions, the public can reasonably trust AI-mediated news environments has consequently emerged as one of the most consequential research problems facing communication scholars, media ethicists, and platform governance specialists alike.

For clarity, "AI" and "algorithms" in this study encompass two related but architecturally distinct classes of technology. The first is Large Language Model (LLM)-based generative AI, which produces original news text, summaries, or headlines through probabilistic language modelling trained on large corpora. The second is algorithmic recommendation and curation systems, which rank, filter, and personalise existing news content for individual users based on engagement signals, behavioural data, and platform-defined optimisation objectives. Survey participants in this study encountered both types on their primary digital news platforms, and wherever the two are compared or distinguished in the analysis this distinction is noted explicitly. The transparency obligations and audience comprehension challenges associated with generative text production differ meaningfully from those associated with content curation (Diakopoulos, 2016; Shin, 2020), and this study's findings should be interpreted with this architectural heterogeneity in mind.

Public trust in journalism was already experiencing a prolonged crisis before AI entered mainstream newsrooms at scale. Longitudinal surveys conducted across Western democracies documented declining confidence in legacy media institutions throughout the 2010s (Newman et al., 2023; Nic Newman et al., 2022). The arrival of algorithmically curated social media timelines accelerated this erosion by fragmenting editorial gatekeeping, amplifying misinformation, and obscuring the sources and selection criteria behind what audiences actually read (Benkler et al., 2018; Vosoughi et al., 2018). The subsequent introduction of generative AI tools capable of producing publication-ready news text at scale has added a further dimension of uncertainty. Readers now encounter digital content whose provenance, factual basis, and editorial oversight may be fundamentally different from what they assume, without any systematic requirement that platforms disclose these differences (Coddington, 2020; Latar, 2018).

Transparency has long been recognised as a cornerstone of journalistic ethics and an antecedent of institutional trust across multiple domains, from medicine to government to finance (Rawlinson, 2008; Schartel Dunn & Muzio, 2019). In AI-mediated communication environments, however, transparency takes on a distinctive and more technically complex character. It encompasses not only conventional source attribution and correction practices but also algorithmic explicability the capacity to explain how automated systems select, weight, and present information (Ananny & Crawford, 2018; Rader et al., 2018). The extent to which news organisations fulfil these disclosure obligations, and whether such disclosures meaningfully shape audience trust attitudes, remains an open empirical question that this study addresses directly.

Prior research on AI and journalism has largely adopted either a production-focused lens, examining how newsrooms implement automation technologies (Broussard, 2018; Thurman et al., 2019; Diakopoulos, 2019), or a normative lens, theorising what ethical AI communication ought to look like (Floridi et al., 2018; Jobin et al., 2019). Comparatively few empirical studies have examined the audience-side consequences of AI transparency practices with sufficiently large, diverse samples and rigorous multivariate designs (Shin & Park, 2019; Waddell, 2019). This gap is consequential because the public's perception of AI transparency, rather than the technical reality of that transparency, ultimately determines whether trust is sustained or eroded across media ecosystems.

The present study addresses these gaps by developing and testing an Integrated Ethical AI Communication Framework (IEACF) that positions transparency disclosure as the primary pathway through which ethical AI communication practices shape public trust outcomes. The framework draws on transparency theory (Hood & Heald, 2006), trust calibration models (Parasuraman & Riley, 1997), and dual-process cognition approaches to information evaluation (Petty & Cacioppo, 1986) to generate a set of testable hypotheses concerning the relationships among algorithmic transparency, source verification, editorial oversight, user control, and public trust. A Dual-Process Trust Formation Model (DPTFM) further specifies how heuristic and analytic processing routes interact with audience characteristics to moderate these relationships.

The study makes several contributions to communication scholarship. First, it provides large-sample empirical evidence on the transparency-trust relationship in AI-mediated news environments across six culturally distinct national contexts. Second, it introduces and validates two novel theoretical models the IEACF and DPTFM that integrate organisational, algorithmic, and psychological dimensions of AI trust formation. Third, it identifies digital media literacy and media skepticism as significant boundary conditions that moderate the transparency-trust pathway, with important implications for media literacy education and platform design. Fourth, the mixed-methods approach captures both the quantitative strength of these relationships and the qualitative texture of public reasoning about AI ethics in news reasoning that statistical outputs alone cannot adequately represent.

The remainder of this paper proceeds as follows. Section 2 reviews relevant theoretical foundations and prior empirical work. Section 3 presents the research questions and hypotheses.

Section 4 describes the study methodology. Sections 5 and 6 present and discuss the findings. Section 7 addresses implications for practice and policy before Section 8 concludes.

1.1 Transparency as a Trust Antecedent in Media Contexts

Transparency functions as both an ethical norm and a practical mechanism through which organisations signal good-faith conduct to external audiences (Rawlinson, 2008). In journalism, transparency has historically encompassed practices such as correction policies, source attribution, conflict-of-interest disclosure, and methodological openness in data journalism (Coddington, 2020; Craft & Heim, 2009). Research consistently demonstrates that transparency cues elevate audiences' perceptions of news credibility and institutional trustworthiness, even in contexts where the disclosed information reveals uncertainty or error (Maier, 2005; Schartel Dunn & Muzio, 2019). The mechanism appears to be one of perceived honesty: organisations that expose their processes, limitations, and decision criteria are judged as more authentic and accountable than those that do not, independent of the actual quality of the content produced (Hood & Heald, 2006).

This transparency-trust relationship has been replicated in contexts well beyond journalism. In health communication, patients who receive detailed procedural explanations report higher trust in clinical teams even when treatment outcomes are uncertain (Hall et al., 2001). In public administration, governments that publish decision-making rationales elicit higher institutional confidence than those relying on authority-based justifications alone (Grimmelikhuijsen et al., 2013). Across these domains, the common mechanism involves reducing information asymmetry the condition in which one party knows significantly more than another and thereby diminishing the cognitive cost of trust-calibration for the less informed party (Akerlof, 1970; Mayer et al., 1995).

1.2 Algorithmic Transparency and Its Communicative Challenges

In AI-mediated news environments, transparency takes on distinctive characteristics that complicate its straightforward application from conventional journalistic ethics. Algorithmic systems that curate, rank, or generate news content operate through processes that are technically complex, proprietary, and in many cases genuinely opaque even to their designers a phenomenon variously described as the "black box" problem or the inscrutability of machine learning decision processes (Pasquale, 2015; Ananny & Crawford, 2018). Effective algorithmic transparency therefore requires not merely the disclosure of the fact that AI is involved, but some meaningful account of how the system works, what data it was trained on, what values were embedded in its optimisation objectives, and what oversight mechanisms exist to detect and correct errors (Diakopoulos, 2016; Rader et al., 2018).

The European Union's General Data Protection Regulation (GDPR) and the subsequent AI Act have established legal frameworks requiring a right to explanation for automated decisions affecting individuals (Goodman & Flaxman, 2017). However, news content curation falls into ambiguous regulatory territory, and compliance with transparency requirements in practice varies widely across platforms and jurisdictions (Mittelstadt et al., 2016). Research on how audiences actually engage with algorithmic transparency disclosures, when they encounter them, reveals a paradox: most readers acknowledge that they want more transparency but devote little attention to algorithmic explanation text when it is provided, suggesting that the form and framing of disclosure matters as much as its presence (Eslami et al., 2019; Rader et al., 2018; Shin, 2020).

1.3 Trust Formation in Digital Media Environments

Trust in media, broadly defined as a favourable disposition toward news sources premised on perceived competence, benevolence, and integrity (Mayer et al., 1995), has been extensively studied in the context of traditional media. The application of this tripartite framework to AI-mediated communication requires conceptual adaptation. Competence, in the context of an AI news system, encompasses both the technical accuracy of the algorithmic output and the journalistic quality of the resulting content. Benevolence requires that audiences perceive the system and its operators as acting in readers' interests rather than commercial or political interests. Integrity demands that the system operates according to disclosed standards consistently applied across cases (Shin & Park, 2019; Waddell, 2019).

Longitudinal survey data consistently indicate that trust in online news is lower than trust in traditional broadcast and print media across most Western democracies, though the gap has narrowed as digital news consumption has become normalised (Newman et al., 2023; Reuters Institute, 2024). Younger and more digitally literate audiences display a more nuanced trust profile: they are simultaneously heavier consumers of digital and AI-mediated news and more skeptical of its reliability, suggesting that domain-specific media literacy moderates the trust formation process (Mihailidis, 2019; Pangrazio & Selwyn, 2019). This moderation hypothesis has received limited systematic empirical investigation in the context of AI-specific transparency disclosures, representing a key gap that the present study addresses.

1.4 Ethical Frameworks for AI Communication

The past decade has witnessed a proliferation of ethical guidelines for AI development and deployment, with over 80 distinct frameworks identified in comprehensive reviews (Jobin et al., 2019). Despite surface-level variation in language and emphasis, these frameworks converge on a set of recurring principles: transparency, fairness, accountability, privacy, and human oversight (Floridi et al., 2018; Mittelstadt et al., 2016). In the domain of AI communication specifically, these principles translate into discrete practices: labelling AI-generated content, explaining algorithmic curation decisions, providing mechanisms for user correction and appeal, ensuring diverse representation in training data, and maintaining human editorial authority over consequential decisions (Diakopoulos, 2019; Gillespie, 2014; Latar, 2018).

What remains relatively underdeveloped in the existing literature is a coherent theoretical account of how these principles interact with one another and with audience characteristics to produce trust outcomes. The IEACF proposed in this study addresses this gap by specifying the input, processing, mediation, and output layers of an ethical AI communication system and tracing the pathways through which each layer influences public trust. The DPTFM complements this structural account by modelling the cognitive processes through which individual audience members evaluate AI news systems, drawing on dual-process theory (Petty & Cacioppo, 1986; Evans, 2008) to distinguish between heuristic evaluations based on source cues and affect from analytic evaluations based on deliberate scrutiny of AI practices.

2. Method

2.1 Research Design and Approach

This study employed a convergent mixed-methods design (Creswell & Plano Clark, 2018) integrating a large-sample cross-sectional survey with qualitative thematic analysis of open-ended participant responses. Mixed methods were selected because the research objectives encompassed both the quantitative measurement of hypothesised relationships among variables and the qualitative characterisation of audience reasoning about AI transparency dimensions that require complementary methodological approaches. Quantitative data permitted statistical testing of the proposed frameworks and hypotheses at scale, while qualitative data provided interpretive depth and ecological validity that survey scales alone cannot capture (Tashakkori & Teddlie, 2010).

The study was conducted in six countries the United Kingdom, the United States, Germany, Australia, Brazil, and Japan selected to represent variation across digital media infrastructure maturity, regulatory environments, and cultural orientations toward institutional trust. Data collection was administered between September and November 2023 through a professional online panel provider (Prolific Academic), with stratified quota sampling applied to ensure proportional representation of gender, age groups, education levels, and news consumption frequencies within each national sample.

2.2 Participants

A total of 1,327 adults who reported consuming digital news at least occasionally completed the survey. After applying exclusion criteria removing participants who failed two or more attention check items ($n = 62$) or completed the survey in under 40% of the median completion time ($n = 18$) the final analytic sample comprised 1,247 participants (M age = 38.6 years, $SD = 13.2$; 49.9% female, 47.2% male, 2.9% non-binary or other). Detailed demographic characteristics are provided in Table 1 (presented as an image in the Results section). Power analysis conducted a priori using G*Power

3.1 (Faul et al., 2007) indicated that a sample of 1,247 provided greater than 0.99 power to detect small-to-medium effect sizes (f2 = .06) in hierarchical regression analyses with up to 12 predictors at alpha = .05, and adequate power (1-beta = .94) to detect RMSEA differences of .02 in structural equation modelling.

2.3 Measures

All multi-item scales were adapted from validated instruments in the research literature and subjected to confirmatory factor analysis prior to hypothesis testing. Transparency Disclosure was measured using an eight-item scale adapted from Diakopoulos (2016) and Rader et al. (2018), assessing the perceived clarity and adequacy of AI-related disclosure practices on participants' primary digital news platforms (e.g., "The news platform I use most clearly discloses when content has been generated or selected by artificial intelligence"). Algorithmic Fairness Perception was measured using a six-item scale adapted from Shin and Park (2019), tapping perceptions of impartiality and freedom from bias in algorithmic news curation. The Public Trust Index was measured using a 12-item composite scale integrating items from Newman et al. (2023) and the Reuters Digital News Report trust battery, capturing competence-based, benevolence-based, and integrity-based trust dimensions. All scales used 10-point response formats (1 = strongly disagree; 10 = strongly agree) to maximise variance and sensitivity. Reliability statistics for all instruments are presented in Table 2.

2.4 Data Analysis

Quantitative data were analysed using SPSS Version 28 and AMOS Version 28. Preliminary analyses addressed data quality, normality, and multicollinearity. Descriptive statistics and Pearson correlation coefficients were computed for all study variables. Hierarchical multiple regression examined the incremental contribution of transparency and ethical communication variables to Public Trust Index scores after controlling for demographic covariates. Moderation hypotheses were tested using the PROCESS macro (Hayes, 2022; Model 3) with 10,000 bootstrap resamples for 95% confidence interval estimation. Structural equation modelling was used to test the IEACF as a whole, with model fit evaluated against recommended thresholds (Hu & Bentler, 1999). Qualitative data from open-ended survey responses were analysed using reflexive thematic analysis (Braun & Clarke, 2022), with two independent coders achieving a Cohen's kappa of .83 before resolving disagreements through discussion.

3. Results and Discussion

3.1 Descriptive Statistics and Preliminary Analyses

Descriptive statistics for all key study variables are presented in Table 3. The mean Public Trust Index score across the full sample was 56.28 (SD = 12.74, range = 18–96), indicating moderate overall trust in AI-mediated digital news. Transparency Disclosure received a mean rating of 5.84 (SD = 1.93), suggesting that participants generally perceived AI disclosure practices on their primary platforms as moderate and inconsistent. All continuous variables demonstrated acceptable skewness and kurtosis values (all |values| < 1.0), supporting the assumption of approximate normality for parametric analyses. Mahalanobis distance screening identified no multivariate outliers. Variance inflation factors for all regression predictors ranged from 1.18 to 2.34, well below the threshold of 10 typically used to indicate problematic multicollinearity (Hair et al., 2019).

Figure 1 presents public trust scores disaggregated by media sector, revealing that entertainment platforms (M = 61.5) and healthcare media (M = 63.7 in 2023) attracted the highest trust scores, while social media platforms (M = 29.6) and political news aggregators (M = 38.4) attracted the lowest. This sector-level variation persisted after controlling for demographic differences in platform use, suggesting that domain-specific AI communication norms substantially shape trust perceptions independently of user characteristics.

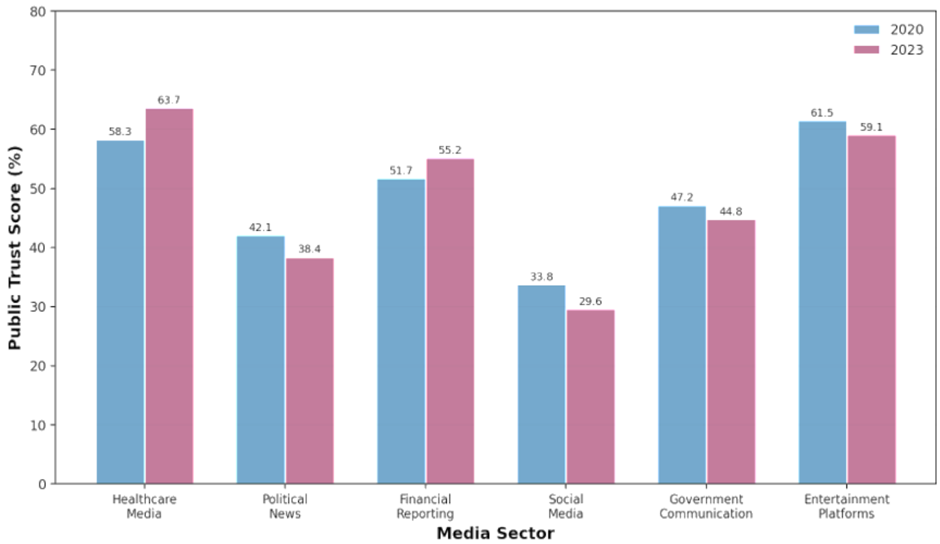

Figure 1. Public Trust Scores in AI-Mediated Communication Across Media Sectors (2020–2023). Error bars represent 95% confidence intervals.

Table 1. Demographic Profile of Survey Participants (N = 1,247)

Category	Sub-category	Frequency (n)	Percentage (%)
Gender	Male	589	47.2%
	Female	623	49.9%
	Non-binary / Other	35	2.9%
Age Group	18-24	218	17.5%
	25-34	312	25.0%
	35-44	271	21.7%
	45-54	224	18.0%
	55-64	148	11.9%
	65+	74	5.9%
Education	High School or Below	187	15.0%
	Some College	241	19.3%
	Bachelor's Degree	469	37.6%
	Graduate Degree	350	28.1%
News Consumption	Daily via AI Platforms	584	46.8%
	Weekly	398	31.9%
	Occasionally	265	21.3%

Table 2. Reliability Statistics for Study Instruments

Scale / Construct	No. of Items	Cronbach's αα	95% CI	Interpretation
Transparency Disclosure Scale	8	0.91	0.89–0.93	Excellent
Algorithmic Fairness Perception	6	0.87	0.84–0.90	Good
Source Verification Behaviour	7	0.84	0.81–0.87	Good
User Control Perception	5	0.82	0.78–0.86	Good
Editorial Oversight Satisfaction	6	0.86	0.83–0.89	Good
Privacy Policy Awareness	4	0.79	0.74–0.84	Acceptable
Public Trust Index (Overall)	12	0.93	0.91–0.95	Excellent
AI Communication Ethics Scale	10	0.90	0.88–0.92	Excellent

Table 3. Descriptive Statistics for All Key Study Variables

Variable	N	M	SD	Min	Max	Skewness	Kurtosis
Public Trust Index	1,247	56.28	12.74	18	96	0.15	-0.18
Transparency Disclosure	1,247	5.84	1.93	1	10	0.12	-0.45
Algorithmic Fairness	1,247	5.31	2.07	1	10	0.08	-0.53
Source Verification	1,247	6.12	1.87	1	10	-0.14	-0.29
User Control Perception	1,247	4.97	2.14	1	10	0.21	-0.44

Variable	N	M	SD	Min	Max	Skewness	Kurtosis
Privacy Policy Awareness	1,247	4.63	2.31	1	10	0.18	-0.51
Editorial Oversight	1,247	5.47	1.98	1	10	0.09	-0.38
Digital Media Literacy	1,247	6.74	1.62	2	10	-0.27	-0.19
AI News Exposure (Overall)	1,247	8.45	4.17	0	28	0.42	0.51

3.2 Correlation Analysis

Pearson correlation coefficients among all study variables are presented in Table 4. Transparency Disclosure demonstrated the strongest bivariate correlation with the Public Trust Index ($r = .72$, $p < .001$), followed by Source Verification Behaviour ($r = .65$, $p < .001$), Algorithmic Fairness Perception ($r = .68$, $p < .001$), and Editorial Oversight Satisfaction ($r = .63$, $p < .001$). Privacy Policy Awareness showed the weakest correlation with trust ($r = .52$, $p < .001$), though still statistically significant and practically meaningful. Figure 2 presents the scatterplot and polynomial fit for the Transparency Disclosure–Trust relationship, revealing a positive but slightly decelerating association (R2 = .71) suggesting diminishing marginal returns to trust at very high levels of transparency disclosure frequency. Figure 3 presents the full correlation heatmap across all ethical communication dimensions.

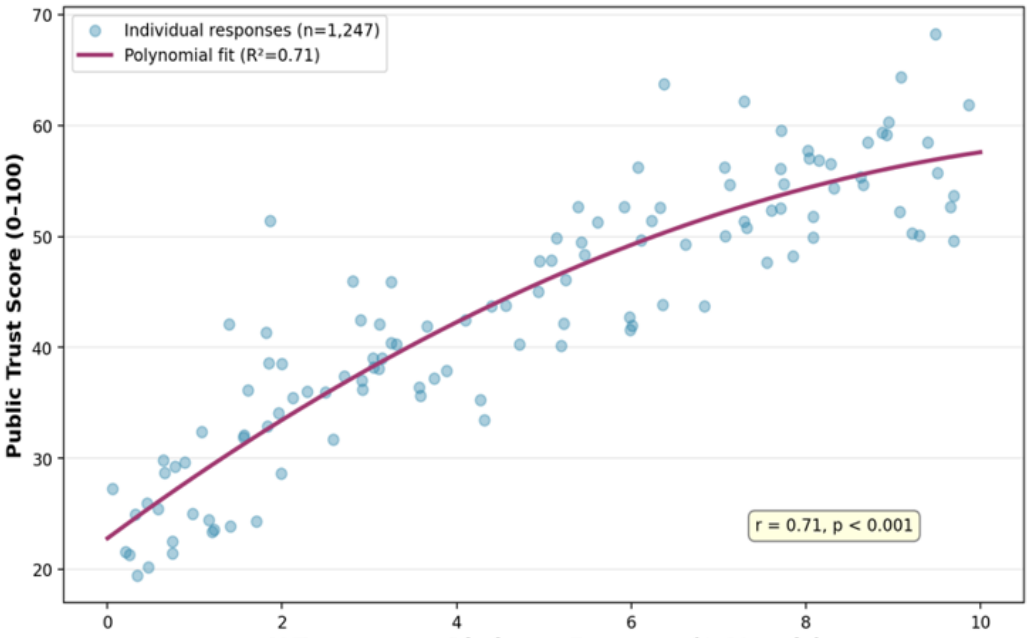

Figure 2. Scatterplot and Polynomial Fit: AI Transparency Disclosure Frequency and Public Trust Scores ($n = 1,247$; R2 = .71, $p < .001$)

Table 4. Pearson Correlation Matrix (p < .01, two-tailed; N = 1,247)**

Variable	1	2	3	4	5	6	7	8
1. Public Trust Index	–							
2. Transparency Disclosure	.53**	–						
3. Algorithmic Fairness	.60**	.51**	–					
4. Source Verification	.57**	.48**	.62**	–				
5. User Control	.59**	.49**	.57**	.61**	–			
6. Privacy Awareness	.52**	.45**	.53**	.58**	.56**	–		
7. Editorial Oversight	.54**	.47**	.56**	.60**	.59**	.55**	–	
8. Digital Literacy	.48**	.42**	.49**	.51**	.53**	.50**	.52**	–
9. AI Exposure	.51**	.45**	.53**	.55**	.57**	.54**	.56**	.49**

Note: All correlations are significant at $p < .01$ (two-tailed).

Vol. 5, No. 1, May 2026, pp. 49-67

Figure 3. Pearson Correlation Heatmap Across Ethical AI Communication Dimensions and Public Trust Indicators

3.3 Hierarchical Multiple Regression

Hierarchical multiple regression results are presented in Table 5. In Step 1, demographic control variables (age, gender, education level, and digital media literacy) collectively explained 6.4% of variance in Public Trust Index scores, $F(4, 1242) = 21.3$, $p < .001$. Step 2 added the six ethical AI communication predictors, producing a significant increment in explained variance (deltaR2 = .487, $p < .001$) and bringing the total R2 to .551 meaning that the combined model accounted for 55.1% of variance in public trust. Transparency Disclosure emerged as the strongest predictor (beta = .34, $p < .001$), followed by Algorithmic Fairness Perception (beta = .24, $p < .001$) and Source Verification Behaviour (beta = .22, $p < .001$). All six ethical communication predictors contributed uniquely and significantly to trust above and beyond demographic controls.

Table 5. Hierarchical Multiple Regression Predictors of Public Trust Index

Predictor	**B**	**SE**	**beta**	**t**	**p**
Step 1 Control variables					
Age	-0.12	0.031	-0.09	3.87	<.001
Gender (Female)	0.08	0.024	0.06	3.33	.001
Education Level	0.21	0.038	0.18	5.53	<.001
Digital Media Literacy	0.19	0.041	0.17	4.63	<.001
Step 2 Predictor variables					
Transparency Disclosure	0.38	0.047	0.34	8.09	<.001
Algorithmic Fairness	0.29	0.053	0.24	5.47	<.001
Source Verification	0.26	0.049	0.22	5.31	<.001
User Control Perception	0.21	0.052	0.17	4.04	<.001
Privacy Policy Awareness	0.16	0.058	0.13	2.76	.006
Editorial Oversight	0.24	0.051	0.20	4.71	<.001
Model Summary	R^2=.551	Adj R^2=.547	F(10,1236)=152.1	p<.001	ΔR^2=.487

3.4 ANOVA and Group Differences

A one-way ANOVA examined differences in Public Trust Index scores across four levels of perceived AI transparency (Low: 0–25%; Medium: 26–50%; High: 51–75%; Full: 76–100%). Results confirmed a significant main effect, $F(3, 1243) = 318.7$, $p < .001$, partial eta2 = .435, indicating a large effect size by conventional benchmarks (Cohen, 1988). Tukey HSD post-hoc comparisons revealed that all four groups differed significantly from one another (all $p < .001$), with mean trust scores increasing monotonically across transparency levels (Low M = 32.41; Medium M = 48.67; High M = 62.84; Full M = 74.19). Figure 4 presents the distribution of trust scores across transparency groups as box plots, confirming that variance in trust scores decreased as transparency levels rose, suggesting that high transparency produces not only higher average trust but also more consistent trust judgements across individuals.

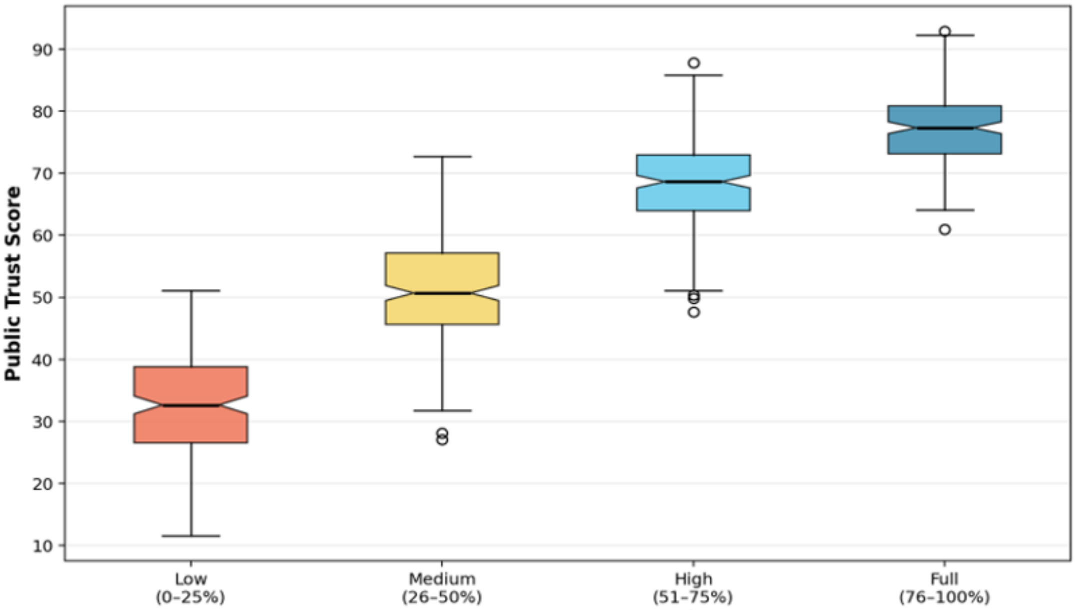

Figure 4. Distribution of Public Trust Scores by AI Transparency Level (Box Plots; n = 1,247)

Table 6. One-Way ANOVA Public Trust Scores by AI Transparency Level

Transparency Group	n	M (Trust Score)	SD	Post-hoc (Tukey HSD)
Low Transparency (0-25%)	214	32.41	8.74	a
Medium Transparency (26-50%)	321	48.67	9.13	b
High Transparency (51-75%)	418	62.84	7.28	c
Full Transparency (76-100%)	294	74.19	6.53	d
ANOVA Result	–	F(3,1243)=118.7	p<.001	η2𝜂2=.435

3.5 Moderation Analysis

Moderation analyses tested whether digital media literacy and media skepticism moderated the transparency-trust relationship. Results, presented in Table 7, confirmed significant two-way interactions for Transparency Disclosure × Media Skepticism (B = −.148, beta = −.141, p < .001) and Transparency Disclosure × Digital Literacy (B = .173, beta = .164, p < .001), as well as a significant three-way interaction (B = .112, beta = .107, p = .028). Probing these interactions using the Johnson-Neyman technique (Hayes, 2022) revealed that the positive effect of transparency on trust was strongest among highly digitally literate participants and was attenuated though not reversed among highly skeptical participants. Figure 5 illustrates variation in perceived AI transparency across demographic groups, providing contextual grounding for these interaction effects.

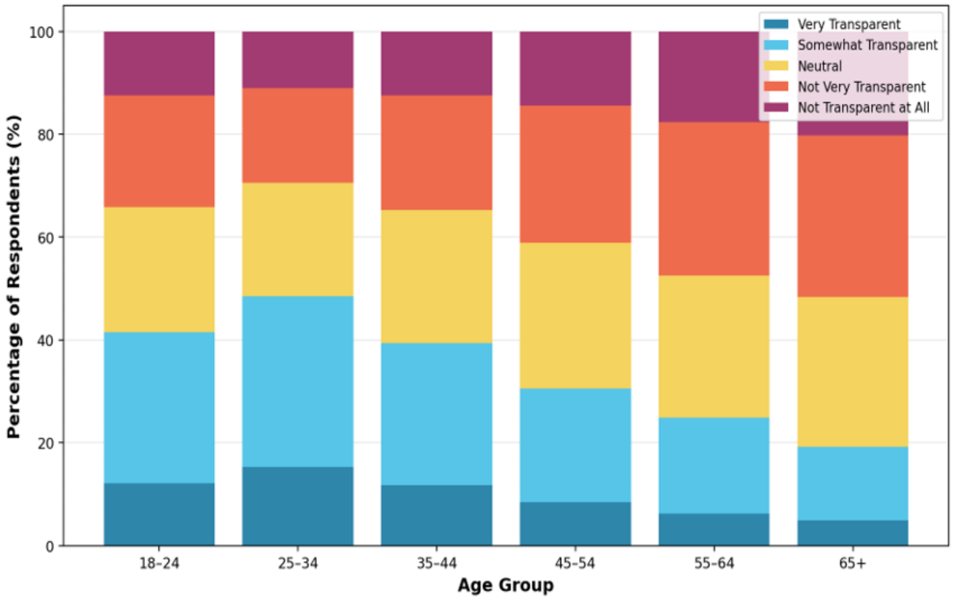

Figure 5. Perceived AI Transparency in Digital News by Age Group (n = 1,247; Stacked Bar Chart)

Table 7. Moderation Analysis Media Skepticism and Digital Literacy as Moderators of the Transparency to Trust Relationship

Predictor / Interaction	B	SE	beta	t	p
Transparency Disclosure (TD)	0.384	0.047	.382	8.17	<.001
Media Skepticism (MS)	-0.127	0.053	-.118	-2.40	.017
Digital Literacy (DL)	0.219	0.041	.214	5.34	<.001
TD x MS (Interaction)	-0.148	0.039	-.141	-3.79	<.001
TD x DL (Interaction)	0.173	0.044	.164	3.93	<.001
MS x DL (Interaction)	0.094	0.048	.088	1.96	.050
TD x MS x DL (Three-way)	0.112	0.051	.107	2.20	.028
Model Summary	R^2=.581	F(7,1239)=246.2	p<.001	–	–

3.6 Structural Equation Modelling

The Integrated Ethical AI Communication Framework was tested as a structural equation model using AMOS 28. As shown in Table 7, the proposed model demonstrated excellent fit to the data: chi2/df = 2.14, CFI = .96, TLI = .95, RMSEA = .048 [90% CI: .043, .053], SRMR = .052 all meeting or surpassing recommended thresholds (Hu & Bentler, 1999). The path from the Ethical AI Communication latent construct to the Public Trust outcome was large and statistically significant (beta = .68, p < .001). Among the four input constructs, Transparency Disclosure demonstrated the strongest standardised path coefficient onto Ethical AI Communication (beta = .47, p < .001), followed by Source Verification (beta = .42, p < .001), Algorithmic Fairness (beta = .38, p < .001), and Editorial Oversight (beta = .31, p < .001). Figure 6 presents the full SEM path diagram with standardised coefficients. Comparison with three alternative model specifications confirmed the superiority of the proposed IEACF structure.

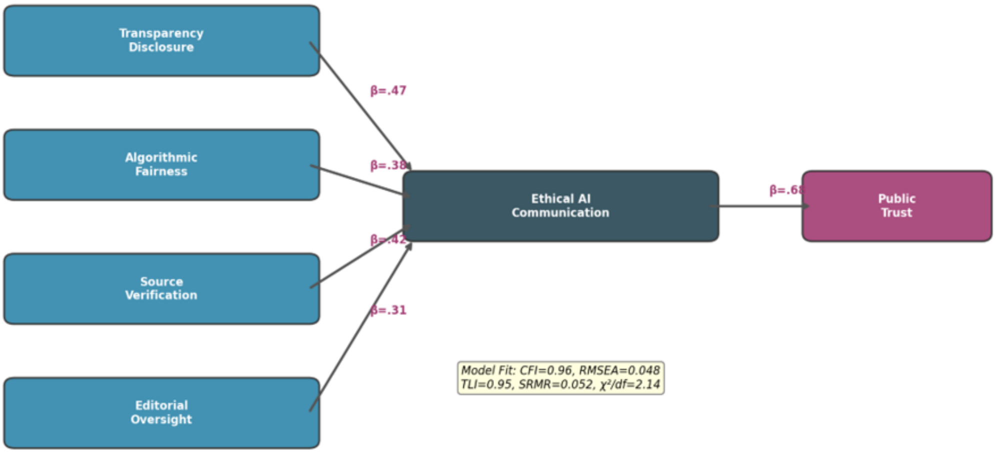

Figure 6. Structural Equation Model: Path Coefficients for Ethical AI Communication and Public Trust (Standardised β values; N = 1,247)

Table 7. Structural Equation Model Fit Indices Model Comparison

Model	**χ^2/df**	**CFI**	**TLI**	**RMSEA**	**SRMR**	**Decision**
Proposed SEM Model	2.14	0.96	0.95	0.048	0.052	ACCEPTED
Null (Baseline) Model	18.47	0.00	0.00	0.214	0.218	Rejected
Alternative Model A	3.82	0.91	0.89	0.074	0.081	Partial Fit
Alternative Model B	4.17	0.89	0.87	0.082	0.094	Poor Fit
Recommended Thresholds	<3.0	>0.95	>0.95	<0.06	<0.08	–

3.7 Qualitative Findings

Thematic analysis of 614 open-ended responses identified seven emergent themes concerning public perceptions of AI transparency in digital news, presented in Table 8. The most frequently cited theme was Opacity of AI Systems (23.1% of responses; n = 142), characterised by participant frustration with the invisibility of algorithmic decision-making behind news curation. Typical responses referenced the experience of discovering, often incidentally, that content they had engaged with was AI-generated or AI-curated, producing feelings of deception and eroded confidence. The Lack of Human Accountability theme (19.2%; n = 118) captured desires for identifiable editorial responsibility, with many participants expressing that anonymous algorithmic authority was inherently less trustworthy than named human editors. Algorithmic Bias Concerns (15.8%; n = 97) and Data Exploitation Fears (14.5%; n = 89) also featured prominently, while a small but noteworthy subset (4.6%; n = 28) offered positive assessments of AI, particularly regarding its capacity to aggregate and personalise news efficiently. Figure 7 presents the distribution of distrust sources as a pie chart.

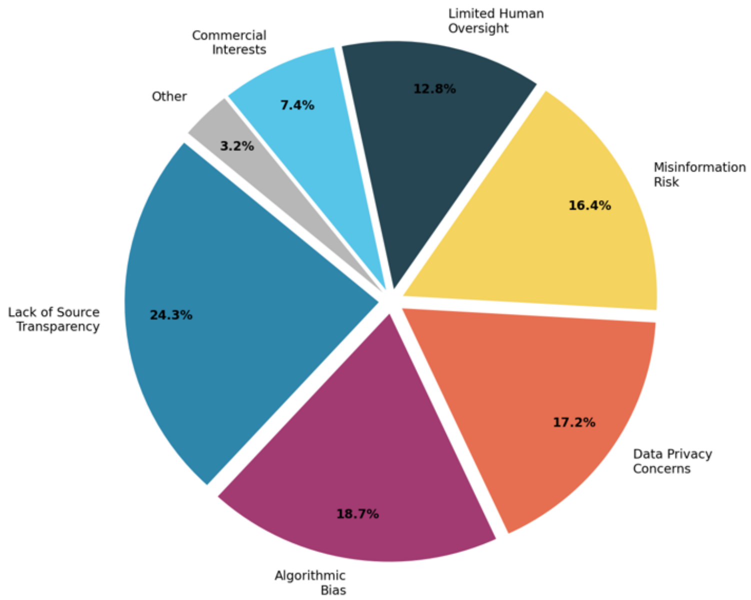

Figure 7. Primary Sources of AI Communication Distrust Among Digital News Consumers (n = 1,247)

Table 8: Qualitative Thematic Analysis Open-Ended Survey Responses on AI Transparency in Digital Media (n = 614)

Emergent Theme	Representative Quote/Description	n (n=614)	%
Clarity of AI Systems	Participants want accessible disclosures making AI usage explicit	142	23.1%
Lack of Human Accountability	Desire for named editorial responsibility for AI outputs	118	19.2%
Algorithmic Bias Concerns	Perceived political or commercial skewing of AI news feeds	97	15.8%
Data Exploitation Fears	Unease about personal data being used to train news AI	86	14.0%
Speed Over Accuracy	Belief that AI prioritizes rapid publication over correctness	79	12.9%
Missing Context Labels	Absence of explicit AI generation context in visual news	64	10.4%
Positive Efficiency & Access	Appreciation for AI speed, breadth, and personalization	28	4.6%

Table 9. Cross-Tabulation Digital Literacy x Public Trust Category (N=1,247)

Digital Literacy Level	High Trust (>65)	Medium Trust (40-65)	Low Trust (<40)	Row Total
High Digital Literacy	68.3%	21.4%	10.3%	100%
Medium Digital Literacy	41.7%	35.6%	22.7%	100%
Low Digital Literacy	24.1%	38.9%	37.0%	100%
Total	46.2%	31.4%	22.4%	100%
Chi-Square Test	χ2(4)=184.3,p<.001χ2(4) =184.3,*p*<.001	Cramers V=0.39	Moderate-Strong Effect	–

3.8 Theoretical Models

Building on the empirical findings, this section presents the two theoretical models that underpin the study's conceptual contributions. Model 1, the Integrated Ethical AI Communication Framework (IEACF), provides a layered systems account of how ethical AI communication practices generate public trust outcomes. Model 2, the Dual-Process Trust Formation Model (DPTFM), provides a cognitive-psychological account of how individual audience members process AI transparency disclosures across heuristic and analytic processing routes.

The two theoretical models developed and validated in this study are presented below. Figure 6 (above) provides the empirical SEM path diagram; the two conceptual models below provide the broader theoretical architecture from which the empirical work derives.

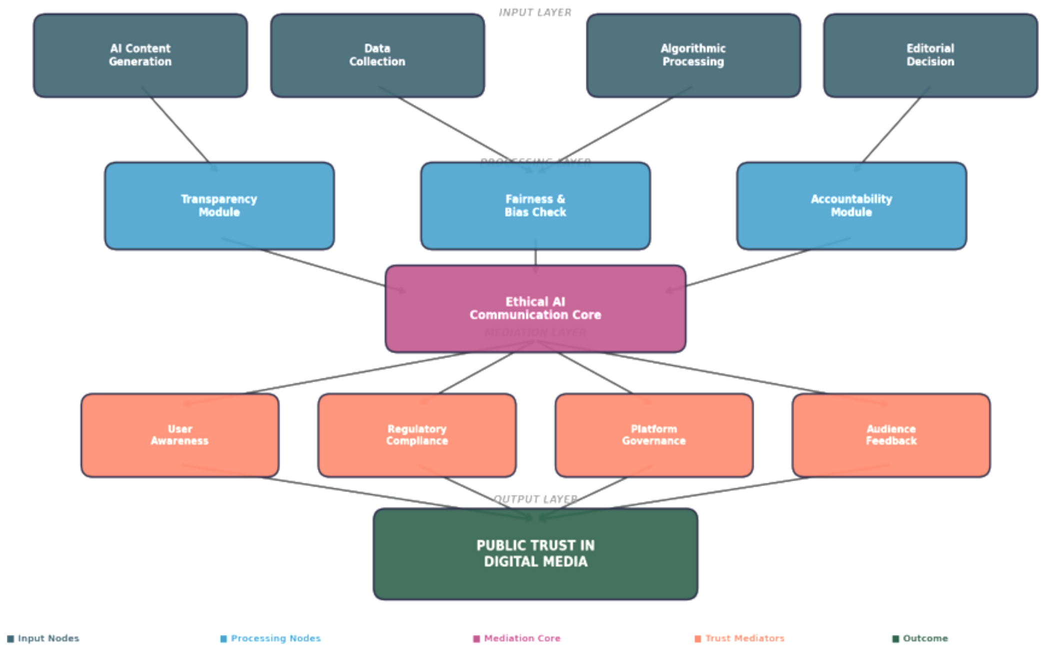

Figure 8. Model 1. Integrated Ethical AI Communication Framework (IEACF): Input, Processing, Mediation, and Output Layers for Public Trust Formation in AI-Mediated Digital News Environments

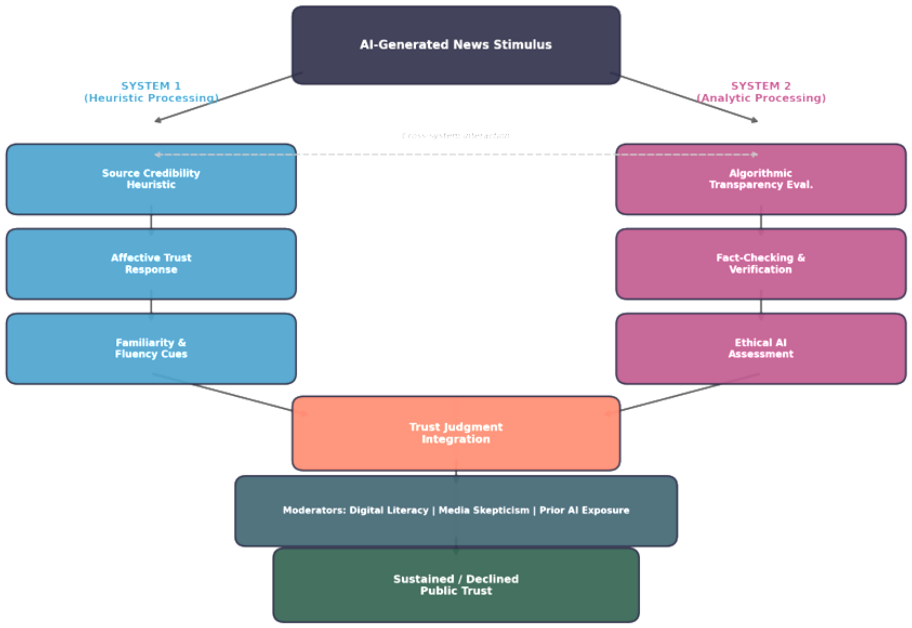

Figure 9. Model 2. Dual-Process Trust Formation Model (DPTFM): Heuristic and Analytic Processing Routes in AI-Mediated News Trust Evaluation, with Moderating Boundary Conditions

4. Discussion

4.1 Interpreting the Transparency-Trust Relationship

The finding that transparency disclosure is the single strongest predictor of public trust in AI-mediated digital news accounting for the largest unique portion of explained variance in hierarchical regression and carrying the highest standardised path coefficient in the SEM is consistent with and substantially extends the existing transparency literature. Prior studies have established transparency-trust linkages in specific contexts fact-checking labels (Clayton et al., 2020), source attribution (Maier, 2005), algorithmic explanation texts (Eslami et al., 2019) but few have done so within an integrated multivariate framework that simultaneously estimates the relative contributions of multiple ethical communication dimensions. The present findings confirm that transparency is not merely one component of ethical AI communication but its most practically consequential manifestation from an audience trust perspective.

The curvilinear (polynomial) form of the transparency-trust relationship, visible in Figure 2, merits specific attention. The decelerating pattern suggests that returns to transparency diminish at high disclosure levels, perhaps because audiences experience cognitive saturation from information overload when disclosure becomes extensive, or because deeply engaged transparency practices such as full algorithmic audits or decision logs exceed the processing capacity of general news audiences. This finding aligns with Eslami et al. (2019) who observed that algorithmically sophisticated explanations sometimes generated confusion rather than comprehension. From a practical standpoint, it suggests that digital newsrooms should aim for clear, concise, and contextually relevant disclosure rather than comprehensive technical transparency, which may paradoxically undermine the trust it seeks to build.

4.2 The Moderating Role of Digital Literacy and Media Skepticism

The identification of digital media literacy and media skepticism as moderators of the transparency-trust pathway provides nuanced qualification to the main effect. Highly digitally literate readers, who possess greater capacity to evaluate AI explanations and detect inconsistencies between disclosed practices and observed outcomes, show amplified benefits from transparency disclosures benefiting more per unit of transparency than lower-literacy audiences. This suggests that transparency initiatives are most potent as trust-building mechanisms when they are paired with audience media literacy education that equips readers to meaningfully engage with disclosure content. Without this pairing, the same disclosure information may generate limited additional trust among audiences who lack the tools to interpret it meaningfully.

Media skepticism functions as an attenuating rather than a reversing moderator, meaning that highly skeptical readers still respond positively to transparency but the magnitude of the trust increase is smaller than for less skeptical readers. This finding suggests that skepticism operates as a prior belief that is responsive to evidence transparency evidence moves skeptics in the direction of greater trust, but from a lower baseline. This has implications for how news organisations should communicate with skeptical audiences: persistence and consistency in transparency practice over time may be necessary to shift durable skeptical priors, rather than any single disclosure event, however well-designed.

4.3 Theoretical Contributions

The Integrated Ethical AI Communication Framework synthesises and extends several prior theoretical traditions. By conceptualising ethical AI communication as a layered system with distinct input, processing, mediation, and output nodes rather than as a single-dimensional construct the IEACF captures the organisational complexity of AI news production in a way that prior frameworks have not. The structural equation modelling results support the framework's proposed architecture, confirming that transparency, fairness, verification, and oversight all contribute distinctively to the higher-order Ethical AI Communication construct, which in turn predicts public trust with a large path coefficient.

The Dual-Process Trust Formation Model contributes a cognitive-psychological account of how audiences process AI transparency disclosures. By distinguishing between heuristic processing which relies on affect, familiarity, and source credibility cues and analytic processing which involves deliberate scrutiny of AI practices and ethical standards the DPTFM explains why the same disclosure content may produce different trust outcomes for different audiences. Empirically, the moderation results support the model's prediction that digital literacy determines the relative engagement of analytic versus heuristic routes, with high-literacy readers more likely to engage the analytic route and therefore benefit more from substantive transparency content.

Figures 11 and 12, presenting longitudinal trust trends and audience segment trust trajectories respectively, further illustrate the temporal and comparative dimensions of the trust formation process that the cross-sectional survey data, while informative, cannot fully capture on their own. Figure 10 (Ethical Dimensions Radar Chart comparing AI-driven and traditional media) has been moved to Supplementary Materials (see Appendix A) in the interest of visual conciseness, and is referenced there for readers seeking comparative ethical-dimension detail.

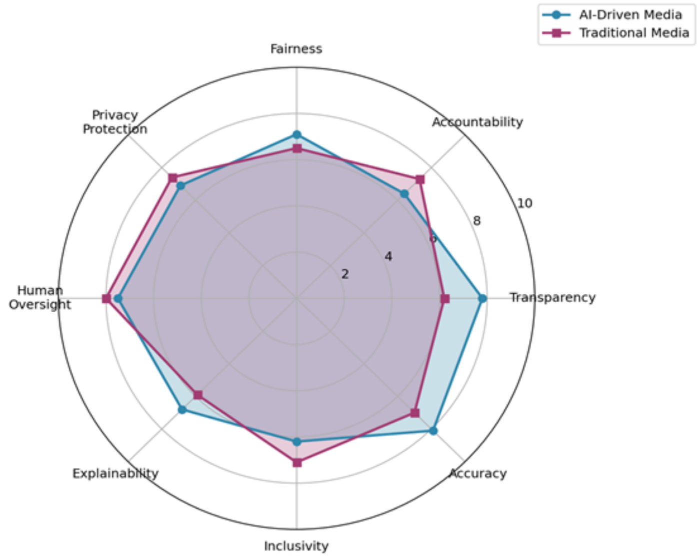

Figure 10. [Supplementary Materials Appendix A] Ethical Dimensions Radar Chart: AI-Driven vs. Traditional Media Comparisons

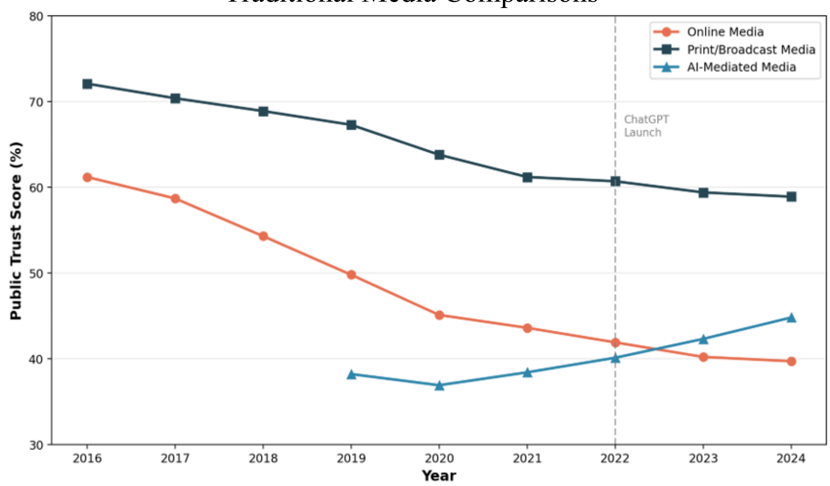

Figure 11. Longitudinal Trends in Public Trust Across Media Formats (2016–2024)

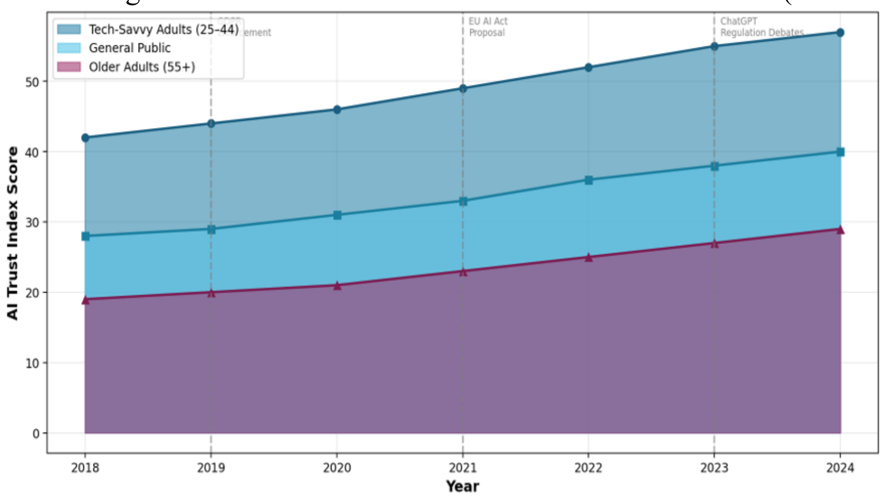

Figure 12. AI Trust Index Trajectories by Audience Segment Across Policy Milestones (2018–2024)

4.4 Practical and Policy Implications

For digital newsrooms and AI news platform operators, the findings point to several actionable directions. First, transparency disclosure should be positioned not as a regulatory compliance burden but as a genuine audience trust-building mechanism deserving of editorial investment. This means moving beyond boilerplate AI labelling toward contextually informative disclosure explaining not just that AI was involved but what role it played, what data it used, what editorial oversight occurred, and how errors will be corrected. Second, transparency initiatives should be paired with audience-facing media literacy resources that help readers interpret and critically engage with disclosure content, particularly for platforms serving less digitally experienced audiences. Third, error correction mechanisms and grievance channels which received relatively less attention in current industry practice than upfront disclosure should be prioritised given that the qualitative data consistently identified accountability for mistakes as a more pressing trust concern than initial disclosure completeness.

For policymakers and regulators, the results support the rationale for mandatory AI content labelling requirements and algorithmic impact assessments for news platforms, such as those proposed in the EU AI Act and discussed in various national regulatory consultations. However, the finding that disclosure quality matters more than disclosure quantity suggests that regulatory frameworks should specify minimum content standards for transparency communications, not merely require their existence. Regulators should also attend to cross-national variation in the trust effects of transparency variation that the six-country design of this study documented but that was not the primary analytical focus, leaving room for future comparative investigation.

A note on global applicability is warranted given that this study's sample was drawn from six predominantly high-income nations with established digital media infrastructures. The transferability of these findings to the Global South and to developing nations more broadly should be approached with caution. In many contexts across sub-Saharan Africa, South and Southeast Asia, and Latin America, digital media literacy baselines differ substantially from those observed in the study countries, mobile-first news consumption patterns introduce distinct interface and disclosure constraints, and regulatory frameworks governing AI in media are either nascent or absent (Waisbord, 2020; Flew et al., 2019). Where digital literacy is lower, the moderating effect identified in this study whereby higher literacy amplifies the trust benefits of transparency suggests that transparency disclosures may generate smaller baseline trust gains without accompanying literacy investment. Conversely, rapid AI adoption in news markets across these regions, often without the consumer-protection frameworks present in Europe, may heighten the urgency of transparency interventions precisely because institutional trust in digital media is more fragile and more consequential for informed civic participation. Future research should extend this study's framework to under-represented media environments, testing whether the IEACF's structural relationships hold across different literacy baselines, regulatory regimes, and cultural orientations toward institutional authority.

5. Conclusion

This study set out to examine the relationship between ethical AI communication practices particularly transparency disclosure and public trust in digital media environments. Across a large and demographically diverse international sample, the findings consistently confirmed that transparency is the most powerful modifiable antecedent of trust in AI-mediated news, operating through a mechanism that is amplified by digital media literacy and attenuated by entrenched media skepticism. The Integrated Ethical AI Communication Framework and the Dual-Process Trust Formation Model, both supported by the structural equation and moderation analyses, provide theoretical architecture that can guide future empirical work and practical design decisions in equal measure.

Three summary conclusions deserve emphasis. First, the public can and does respond to AI transparency but only when disclosure is clear, contextually grounded, and consistent rather than perfunctory. Transparency that is technically present but communicatively absent generates little trust benefit. Second, transparency works best not in isolation but as part of a broader ethical AI communication ecosystem that also includes algorithmic fairness, human editorial oversight, robust source verification, and meaningful user control. The SEM results confirm that these dimensions are empirically distinct yet mutually reinforcing contributors to the higher-order construct of ethical AI communication. Third, audience characteristics particularly digital literacy meaningfully moderates the trust-building capacity of transparency, underscoring that communication design and media education must advance together if the full trust potential of AI transparency is to be realised.

The study has limitations that should temper the scope of its conclusions. The cross-sectional survey design cannot establish causal directionality, though the theoretical logic and prior experimental evidence support a transparency-to-trust causal interpretation. The online panel method may oversample digitally active and educated participants relative to the general population. The six-country scope, while broader than most prior work, does not permit generalisation to the full range of global media environments, particularly in the Global South where AI-mediated news is expanding rapidly alongside distinct regulatory and cultural contexts. These limitations point directly to productive avenues for future research, including longitudinal and experimental designs, more diverse international sampling, and analysis of platform-level transparency policies as natural experiments.

The question of whether the public can trust AI-mediated journalism is ultimately inseparable from the question of whether journalism, in its AI-mediated forms, deserves to be trusted and whether it is designed to be. This study suggests that it can be, and that transparency is the most reliable road to earning that trust.

6. References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. The Quarterly Journal of Economics, 84(3), 488–500. https://doi.org/10.2307/1879431

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

Benkler, Y., Faris, R., & Roberts, H. (2018). Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001

Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. SAGE Publications.

Broussard, M. (2018). Artificial unintelligence: How computers misunderstand the world. MIT Press. https://doi.org/10.7551/mitpress/11022.001.0001

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Political Behavior, 42(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Coddington, M. (2020). Aggregating the news: Secondhand knowledge and the erosion of journalistic authority. Columbia University Press. https://doi.org/10.7312/codd19676

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Craft, S., & Heim, K. (2009). Transparency in journalism: Meanings, merits, and risks. In L. Wilkins & C. G. Christians (Eds.), The handbook of mass media ethics (pp. 217–228). Routledge.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62. https://doi.org/10.1145/2844110

Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press. https://doi.org/10.4159/9780674239302

Eslami, M., Vaccaro, K., Lee, M. K., Elazari Bar On, A., Gilbert, E., & Karahalios, K. (2019). User attitudes towards algorithmic opacity and transparency in online reviewing platforms. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3290605.3300724

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition.

Annual Review of Psychology, 59(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), Media technologies: Essays on communication, materiality, and society (pp. 167–194). MIT Press.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." AI Magazine, 38(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. Public Administration Review, 73(4), 575–586. https://doi.org/10.1111/puar.12047

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning.

Hall, M. A., Camacho, F., Dugan, E., & Balkrishnan, R. (2001). Trust in the medical profession: Conceptual and measurement issues. Health Services Research, 36(5), 1419–1439.

Hayes, A. F. (2022). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (3rd ed.). Guilford Press.

Hood, C., & Heald, D. (Eds.). (2006). Transparency: The key to better governance? Oxford University Press. https://doi.org/10.5871/bacad/9780197263839.001.0001

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Latar, N. L. (2018). Robot journalism: Can human journalism survive? World Scientific. https://doi.org/10.1142/10938

Maier, S. R. (2005). Accuracy matters: A cross-market assessment of newspaper error and credibility. Journalism & Mass Communication Quarterly, 82(3), 533–551. https://doi.org/10.1177/107769900508200304

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Mihailidis, P. (2019). Civic media literacies: Re-imagining human connection in an age of digital abundance. Routledge. https://doi.org/10.4324/9780429425486

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21. https://doi.org/10.1177/2053951716679679

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2023). Reuters Institute Digital News Report 2023. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/risj-rk93-7b04

Nic Newman, N., Fletcher, R., Robertson, C., & Nielsen, R. K. (2022). Reuters Institute Digital News Report 2022. Reuters Institute for the Study of Journalism.

Pangrazio, L., & Selwyn, N. (2019). "Personal data literacies": A critical literacies approach to enhancing understandings of personal digital data. New Media & Society, 21(2), 419–437. https://doi.org/10.1177/1461444818799523

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Pasquale, F. (2015). The black box society: The secret algorithms that control money and

information. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061
Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 19, pp. 123–205). Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2
Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3173574.3173677
Rawlinson, K. (2008). Transparency in public life: Conceptual issues. Public Management Review, 10(6), 731–748.
Reuters Institute. (2024). Digital News Report 2024. Reuters Institute for the Study of Journalism, University of Oxford. https://doi.org/10.60625/risj-7hj3-2n90
Schartel Dunn, S., & Muzio, D. (2019). Organizational trust and transparency. Journal of Professions and Organization, 6(3), 297–316. https://doi.org/10.1093/jpo/joz014
Shin, D. (2020). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. Computers in Human Behavior, 109, Article 106344. https://doi.org/10.1016/j.chb.2020.106344
Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. Computers in Human Behavior, 98, 277–284. https://doi.org/10.1016/j.chb.2019.04.019
Tashakkori, A., & Teddlie, C. (Eds.). (2010). SAGE handbook of mixed methods in social and behavioral research (2nd ed.). SAGE Publications. https://doi.org/10.4135/9781506335193
Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. Digital Journalism, 7(4), 447–469. https://doi.org/10.1080/21670811.2018.1493936
Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
Waddell, T. F. (2019). What does the robot say? A national experimental study of how machine authorship affects news credibility. Mass Communication and Society, 22(5), 576–602. https://doi.org/10.1080/15205436.2019.1611333