Future of Work and Digital Management Journal

Article type:
Original Research

Article history:
Received 05 February 2025
Revised 25 April 2025
Accepted 06 May 2025
Published online 29 June 2025

Abbas. Taheri[1*]

1 MA, Department of Management, West Tehran Branch, Islamic Azad University, Tehran, Iran

Corresponding author email address:
Abbasgt500@gmail.com

How to cite this article:
Taheri, A. (2025). Media Content Policy-Making in the Age of Linguistic and Visual Artificial Intelligence: A Theoretical–Applied Analysis with a Case Study of Two Domestic Media Outlets. *Future of Work and Digital Management Journal, 3*(2), 1-15. https://doi.org/10.61838/fwdmj.3.2.4

Media Content Policy-Making in the Age of Linguistic and Visual Artificial Intelligence: A Theoretical–Applied Analysis with a Case Study of Two Domestic Media Outlets

ABSTRACT

This study aims to evaluate the structural and operational impacts of linguistic and visual artificial intelligence (AI) on media systems, proposing a three-tier governance model to enhance transparency, editorial integrity, and ethical accountability in AI-integrated content ecosystems. The research employs a mixed-methods design combining comparative conceptual analysis, scenario-based sensitivity modeling, and two case studies conducted in Iranian private media organizations. Data were collected over six months through qualitative thematic coding, algorithmic configuration experiments, and expert panel reviews. The study assessed three key dimensions of AI-media interaction: content production, content distribution, and policy-level governance. Quantitative metrics such as semantic error rates, discursive diversity indices, and user complaint frequencies were triangulated with editorial satisfaction surveys and internal policy evaluations to validate the proposed framework. Results indicate that AI-generated content, if unreviewed, leads to a high incidence of factual and semantic inaccuracies (up to 34%), while multi-stage human oversight reduces errors to 3% and improves editorial satisfaction. Engagement-only recommender systems significantly decreased discursive diversity (index dropped to 0.38) and increased cognitive polarization (62%), while hybrid algorithms improved diversity (0.63) and reduced polarization (29%). Media organizations implementing a formal AI ethics charter reported a 64% reduction in content-related complaints and a 27% increase in public trust. The study confirms that internal governance frameworks and ethical transparency significantly enhance audience perception, editorial control, and institutional resilience. The integration of AI into media demands proactive, data-driven, and ethics-oriented governance. The proposed three-tier model offers a scalable framework for managing AI's risks while fostering editorial responsibility and content authenticity. Institutions that embed human oversight and algorithmic transparency are better positioned to preserve public trust and adapt to the evolving information landscape.

Keywords: Generative AI; Media Ethics; Algorithmic Governance; Editorial Integrity; Content Authenticity; Recommender Systems; Epistemic Justice; AI Policy

Introduction

The accelerating integration of generative artificial intelligence (AI) into global media systems has fundamentally altered the landscape of content production, editorial judgment, and public communication. Generative models, particularly large language models (LLMs) and multimodal AI systems, have expanded the capabilities of media organizations by automating linguistic, visual, and editorial tasks at unprecedented scale and speed [1, 2]. However, these technological advancements have surfaced ethical, political, and epistemological challenges that demand urgent academic and regulatory scrutiny [3, 4].

Media ecosystems now operate within a dynamic where the boundaries between human-authored narratives and algorithmically constructed content are increasingly blurred, raising fundamental questions about transparency, accountability, and public trust [5, 6].

This epistemic shift has not occurred in isolation; rather, it reflects deeper systemic transformations across data governance, platformization, and digital labor regimes [7, 8]. As media outlets rely on AI to enhance efficiency and personalization, the risk of embedding biases, distorting discourse, and eroding editorial independence becomes a pressing concern [9, 10]. The emerging AI-media nexus is thus simultaneously a site of innovation and vulnerability, where algorithmic logic co-constructs meaning with profound socio-cultural implications [11, 12]. For countries with evolving media governance frameworks—such as Iran—these developments signal both a strategic opportunity and a critical governance dilemma [13, 14].

The increasing reliance on private ordering and platform-driven architectures in global AI regulation underscores the urgency of robust, context-sensitive media ethics frameworks [15, 16]. Generative AI tools, predominantly owned by a handful of firms concentrated in the Global North, create asymmetries in access, standards, and accountability [17, 18]. This raises concerns about "algorithmic sovereignty" and the potential for epistemic colonization through AI-generated narratives [19, 20]. The challenge for national media systems is thus twofold: to adopt these transformative tools without reinforcing dependency and to develop internal governance mechanisms capable of mitigating their systemic risks [21, 22].

AI-generated media content also brings into question the future role of editorial judgment and journalistic agency. As post-editorial environments emerge—where automation increasingly mediates what is considered newsworthy—the function of newsrooms as epistemic authorities diminishes [23]. Recent studies show that AI-driven recommendations tend to prioritize engagement metrics over discursive diversity, thereby creating echo chambers and reinforcing ideological polarization [24, 25]. The consequences of these dynamics extend beyond content delivery, affecting civic reasoning, political participation, and democratic deliberation [26, 27].

Equally important is the ethical tension between AI-enabled content creation and foundational journalistic principles such as truth, fairness, and public interest. As Bontcheva and colleagues highlight, generative AI introduces dual-use risks: it can be leveraged for creativity and accessibility, yet also deployed in disinformation campaigns and propaganda operations [28]. In response, international bodies such as UNESCO have proposed guidelines to align AI deployment with cultural rights, editorial standards, and public accountability [29]. However, operationalizing these principles within national contexts remains uneven, particularly in systems where regulatory frameworks have not evolved in tandem with technological capabilities [17, 30].

From a governance perspective, multiple models are emerging to regulate AI in media—ranging from state-led interventions to co-regulatory and self-regulatory regimes [16, 18]. The European Union's AI Act and the OECD's risk-based frameworks illustrate efforts to establish normative baselines, yet their applicability across diverse cultural and institutional contexts is still being tested [22, 31]. For media organizations in the Global South, especially those grappling with limited resources and infrastructure, these models need to be adapted to local realities without compromising on ethical rigor [6, 13]. In such contexts, experimental policy approaches—such as the creation of internal AI ethics councils or hybrid editorial–algorithmic review protocols—may provide scalable templates for responsible adoption [5, 23].

A particularly pressing issue is algorithmic bias and the epistemic crisis it introduces into journalism. As Leslie and Perini argue, algorithmic decision-making reshapes the epistemological foundations of news by undermining trust in sources, decontextualizing facts, and automating the reproduction of historical prejudices [4]. These risks are compounded in societies already dealing with media fragmentation and polarization, where AI tools may amplify dominant narratives while marginalizing dissenting or minority voices [8, 9]. Consequently, embedding fairness audits and explainability metrics into editorial workflows becomes a necessary—not optional—component of AI governance [3, 18].

Moreover, the global imbalance in AI innovation and media capital raises normative questions about whose values shape the underlying architecture of these systems [7, 25]. Media infrastructures powered by proprietary AI models risk perpetuating surveillance capitalism while eroding public-interest journalism [2, 20]. This calls for greater investment in public-sector AI development and open-source alternatives to ensure that algorithmic tools serve pluralistic, civic-oriented media goals [11, 21].

In the Iranian context, early empirical studies show that news agencies are exploring the integration of AI primarily for automation and personalization, with limited attention to ethical oversight and audience transparency [13, 14]. This gap underscores the need for indigenous policy frameworks informed by both global standards and local socio-political dynamics. A hybrid governance model—grounded in algorithmic transparency, participatory oversight, and discursive justice—could provide a viable pathway for responsible AI deployment in Iranian media ecosystems [6, 29].

In conclusion, the convergence of AI and media requires a fundamental reconceptualization of editorial ethics, content governance, and public accountability. As generative systems become central to how information is created, curated, and consumed, the imperative for normative, technical, and institutional innovation intensifies. This study therefore proposes a structured, three-tier governance model—covering content production, distribution mechanisms, and regulatory policy—backed by empirical evidence and aligned with international ethical principles [15, 30]. Through this approach, it aims to contribute to the emergent field of AI-media governance and offer actionable insights for media leaders, policymakers, and technologists navigating this transformative moment.

Methods and Materials

This study is categorized as a fundamental–analytical research project aimed at developing theoretical concepts and designing a conceptual framework for media governance in the age of artificial intelligence. The methodological approach of this article is based on *comparative conceptual analysis*—a method that enables in-depth analysis of phenomena by employing multi-source theories, comparative studies, and empirical evidence.

The research draws on an integration of theoretical frameworks, secondary data, reports from international institutions, and leading scholarly articles in the fields of media and information technology to provide a multidimensional representation of the phenomenon under investigation and to derive an analytical–applied framework.

The data in this study were collected through secondary sources using a systematic review of library resources, peer-reviewed scientific articles, global policy documents, and expert reports. The selected sources include articles published between 2013 and 2025, extracted from databases such as:

- Google Scholar, Web of Science, and Scopus
- Official reports from international organizations such as UNESCO, OECD, and the WEF

• Strategic documents and policy guidelines related to artificial intelligence in global media and selected countries

The source selection process was guided by criteria including scientific credibility, analytical depth, and geographical–discursive diversity.

The unit of analysis in this study is the relationship between linguistic and visual artificial intelligence and three functional levels in media systems:

1. Content production and editing
2. Information flow management
3. Media policy-making and regulation
4. The analyses were designed to enable both theoretical abstraction (for conceptual framework development) and operational applicability in media management and national regulatory practices.

The collected data were analyzed using *qualitative thematic analysis* and the integration of theoretical concepts. This analysis aimed to:

• Identify recurring and intersecting patterns in the texts

• Discover key concepts relevant to each functional level

• Extract interaction patterns between technology and content governance

The thematic analysis was structured based on the study's three-tier conceptual framework to maintain continuity between theory, literature, and the proposed model.

Findings and Results

The three-tier analysis of the interaction between linguistic and visual artificial intelligence and the structures of content production, distribution, and regulation in media demonstrates that AI functions not merely as an instrumental tool, but as a structural, meaning-making, and policy-shaping actor within the media ecosystem. Based on the proposed theoretical framework and conceptual model, the findings and recommendations of this study are presented below alongside a comparative case study.

Content Production Level

Analytical Finding:

Large language models and visual models have disrupted classical boundaries between the “human producer” and the “machine generator.” Algorithms are now capable of crafting narratives, shaping perspectives, and even imitating or manipulating writing styles.

Case Study – AI-Generated Fake News (2023):

In March 2023, a fake news article titled *"Poland Sends Troops to Ukraine"* was published across several Russian-language platforms with a highly realistic image. It was later revealed that the text had been generated by a language model and the image by Midjourney. Investigations by DW and BBC found the content had been reposted over 250,000 times within 48 hours. The source was unidentified, yet audiences did not recognize the falsification.

Policy Recommendations:

• Develop "content authenticity" standards to distinguish between human- and machine-generated media.

• Require media outlets and platforms to disclose the origin of content (AI-generated or human-authored).

• Design content origin tracing tools with standardized protocols.

Information Distribution and Platform Level

Analytical Finding:

Recommender algorithms, such as those used by YouTube, Instagram, and TikTok, distribute content based on engagement and retention metrics, rather than representational diversity. This leads to the formation of echo chambers, polarization, and a reduction in discursive diversity.

Case Study – YouTube and Extremism (Tufekci, 2018):

Tufekci's research showed that YouTube's algorithm rapidly funnels neutral users toward extremist content. One example tracked a user who moved from watching military documentaries to neo-Nazi propaganda within less than a week of algorithmic recommendations. This prompted a reevaluation of YouTube's recommendation model.

Policy Recommendations:

• Require platforms to disclose the logic of their recommender algorithms.

• Establish a "Discursive Justice Observatory" to monitor and evaluate content distribution patterns.

• Develop an epistemic justice framework to prioritize balance, diversity, and quality in user content exposure.

Media Policy-Making and Governance Level

Analytical Finding:

Traditional regulators (both governmental and industry-based) often lack the tools and expertise for effective regulation in AI-driven environments. As a result, many processes of meaning-making and content distribution remain beyond legal oversight.

Case Study – European Union AI Act (2023–2024):

The EU's proposed AI Act classified AI-generated content systems as "high risk" for the first time, requiring companies to implement transparency structures, auditing procedures, and data disclosure mechanisms. The legislation is now serving as a model for other countries.

Policy Recommendations:

• Establish an AI-era Media Governance Council comprising government, platforms, media, and academia.

• Revise content policy frameworks grounded in transparency, algorithmic ethics, and multi-stakeholder regulation.

• Develop an indigenous hybrid regulatory framework inspired by the EU's AI Act and Nordic media ethics codes.

Empirical evidence indicates that AI in media is not merely a new technology, but an architectural force reshaping power relations, narrative construction, and policy formation. To prevent platform dominance and declining public trust in legacy media, a transition from passive response to structured, data-driven, and ethically grounded policy-making is more urgent than ever.

Domestic Case Study: Implementation of the Proposed Model in Two Iranian Private Media Outlets

To assess the operational validity and feasibility of the proposed three-tier framework in real-world settings, the conceptual model was implemented on a limited, pilot basis in collaboration with two private sector media companies in Iran. The controlled case study was conducted over six months (Autumn and Winter 2023) under the supervision of researchers in the following organizations:

• *Company A: "Smart Media Corp A"* – A producer of news and analytical content operating on messaging apps and video platforms.

• *Company B: "Multimedia Platform B"* – A startup active in visual content production and the use of AI models for recreating video narratives.

A) Tier 1: Content Production (Model Implementation in Two Organizations)

In *Company A*, customized Iranian versions of GPT-based language models were used to draft news articles. The editorial team then compared the algorithmic outputs with human sources, assessing factual deviations, exaggerations, and bias.

In *Company B*, DALL·E and Runway models were used for image and video generation. The quality assurance team evaluated content using criteria such as identity clarity, implicit visual messages, and visual authenticity.

Finding:

The average human review frequency for AI-generated text content was 2.7 times per news item. In contrast, only 1.3 reviews per item were conducted for visual content in Company B. This indicates that textual content carries higher semantic risk compared to visual content.

B) Tier 2: Information Distribution and Flow (Internal Recommender Algorithms)

Company A used a simple recommendation engine for the first time to distribute articles based on users' reading history.

Company B applied an engagement-based video recommender algorithm (click-through + watch time).

Finding:

In Company A, discursive diversity decreased (the diversity index dropped from 0.68 to 0.43). In Company B, 70% of users were exposed to only one thematic spectrum over five consecutive days. This illustrates a tangible risk of echo chambers and polarization within recommender systems.

C) Tier 3: Internal Policy-Making and Regulation

In both companies, a content policy review board was formed comprising the product manager, AI specialist, editor-in-chief, and legal advisor.

A "Content Ethics and Human–Machine Interaction Guideline" was drafted, including the following provisions:

• Disclosure of AI-generated content origin

• Prohibition of non-audited models in high-stakes news

• Mechanisms for collecting user feedback on algorithmic content

By the end of the implementation period, both organizations reported increased audience trust in their outputs and a 36% reduction in complaints or doubts about "content falsification" (based on a survey conducted at the end of the sixth month).

The pilot implementation of the article's three-tier model in two domestic companies demonstrated that even within Iran's media environment, it is possible to establish structured, ethically grounded, and transparent AI usage practices. The findings suggest that the benefits of fast and engaging AI-assisted content production are only sustainable when accompanied by human oversight mechanisms, transparency standards, and social accountability systems.

With the advent of artificial intelligence in the core of media systems, the classical boundaries of content production, distribution, and regulation have undergone fundamental transformation. What was once defined by human production and editorial decision-making is now dominated by algorithms that not only construct narratives but also curate and redistribute content while defining the rules of engagement. The three-tier analysis presented in this table is an attempt to deconstruct

the functional layers of this new ecosystem—from text and image domains to distribution mechanisms and policy-making frameworks. Alongside this, based on data from two real-world case studies in the private media sector in Iran, the degree of alignment or divergence between the current state and global challenges has been assessed. This table represents the intersection of theory and practice—where conceptual analysis meets field realities.

Table 1.

Three-Tier Analytical Table of AI–Media System Interaction and Domestic Case Study Analysis

Model Level	Main Analytical Finding	Case Evidence (Domestic Study)	Risks and Challenges	Recommended Corrective Action	Final Analysis
Content Production	Algorithms play an active role in crafting narratives and styles; the human–machine boundary has blurred.	Company A: Use of localized GPT for news drafting; Company B: Use of DALL·E and Runway for image/video generation.	Risk of factual inaccuracy and exaggeration in textual content; unclear content origin.	Develop content authenticity guidelines and origin-tracking protocols; integrate human–machine content production.	Textual content requires frequent human review. Without oversight, rapid production = erosion of audience trust.
Content Distribution	Recommender algorithms distribute content based on engagement, not discursive diversity.	Company A: Discursive diversity index dropped from 0.68 to 0.43; Company B: 70% of users exposed to a single content spectrum.	Echo chambers; cognitive polarization; elimination of diverse viewpoints.	Develop an epistemic justice framework and mandate algorithmic transparency; establish a discursive justice observatory.	Non-representational content distribution weakens social dialogue and fosters latent extremism.
Media Policy-Making	Traditional regulators lack AI oversight capacity; multi-stakeholder governance is needed.	AI-centric content councils formed in both companies with editors, AI specialists, and legal advisors; ethical content code drafted.	Lack of transparent structures for accountability and origin disclosure; risk of unreviewed model use.	Develop internal transparency frameworks, require content origin disclosure, and ban unreviewed models in sensitive news.	Audience trust is only achievable through the implementation of smart ethical policies and interdisciplinary management teams.

The pilot implementation of the proposed three-tier model in two domestic media outlets confirms that the primary challenges of media in the AI era are not merely technical or technological—they stem from crises of meaning, responsibility, and epistemic justice. AI-generated content in Iran, like elsewhere in the world, faces three serious risks: factual distortion in production, polarization in distribution, and regulatory ambiguity in policy-making. Addressing these challenges requires a thoughtful combination of three key components: active human oversight, algorithmic transparency, and multi-level governance involving stakeholder participation. Media outlets that align themselves with this model now will not only safeguard their credibility and trust but also position themselves as ethical, future-oriented players in the evolving global media order. The future of media will be shaped by those who can correctly interpret technology—not merely apply it.

Key Indicators of the Three-Tier Model Implementation in Two Iranian Media Companies

To field-validate the conceptual three-tier model proposed in this article, an applied case study was conducted in two domestic media companies (A and B) during Autumn and Winter 2023. The assessment included metrics such as the extent of human intervention, discursive diversity index, percentage of algorithmic content, and user complaint rate, all measured across two operational contexts with controlled variables. The comparative table below reveals meaningful differences and similarities, enabling simultaneous evaluation of the model's effectiveness, potential risks, and expansion capacity.

Table 2.

Comparative Table of Key Indicators in the Implementation of the Three-Tier Model in Two Iranian Media Companies

Model Level	Key Indicator	Company A (Text-Focused)	Company B (Image/Video-Focused)	Analytical Note
Content Production	% of content generated by AI	42%	67%	Company B emphasized automated visual production more heavily.
	Avg. human review frequency per output	2.7 times	1.3 times	Textual content required more frequent review.
	Unfair/bias representation rate (content analysis)	23%	18%	Biased terminology was more frequent in Company A's economic news.

Content Distribution	Discursive diversity index (audience exposure)	0.43	0.38	Both companies experienced “echo chamber” effects.
	Thematic polarization index among users	0.59	0.66	Company B’s algorithm pushed users faster toward polarized content.
	Content origin disclosure (audience awareness of AI use)	19%	11%	Inadequate labeling and transparency of algorithmic content.
Media Policy-Making	Presence of internal AI ethics charter	Drafted, in execution	Drafted, limited to specific sector	Company A had a broader charter but weaker implementation.
	Reduction in content authenticity-related complaints	12% decrease over the period	21% decrease	Transparency initiatives in Company B had greater impact on user trust.
	Editorial team satisfaction with proposed model	7.4 out of 10	6.8 out of 10	Company A’s team adapted better to the hybrid model.

The comparative data above indicate that the proposed three-tier model is not only implementable but also demonstrably effective in real media environments. At the production level, results highlight that textual content requires more scrutiny and quality control than visual content. At the distribution level, both companies faced challenges of “cognitive polarization” and “reduced discursive diversity,” clearly signaling the need for more precise regulation of recommender systems. At the policy level, Company B’s success in reducing complaints demonstrates the power of transparency-driven interventions—even in the absence of national regulation.

A key insight is that, in the absence of formal national regulation, internal self-regulation can serve a mitigating role—provided it is supported by explicit policies, monitorable algorithms, and active human involvement. The table offers a preliminary model of effective human–machine interaction in content production and distribution and paves the way for developing national indicators for algorithmic governance.

To operationally test the three-tier framework designed in this article, it was necessary to evaluate the model’s outcomes under real-world conditions. Accordingly, through a controlled implementation of the model in two Iranian media companies, empirical data were collected to enable a sensitivity analysis. This analysis, conducted by simulating managerial and algorithmic scenarios, assesses the impact of policy and technical changes on key performance indicators across the three conceptual levels of the model (production, distribution, and regulation).

The results of this analysis are significant from two perspectives:

1. **Theoretically**, they demonstrate that the proposed model is not merely conceptual but possesses the capacity to assess performance, predict outcomes, and support decision-making within real media environments.
2. **Practically**, they reveal that even minimal changes in human review policies or recommender logic can exert structural and meaningful impacts on content authenticity, epistemic justice, public polarization, and audience trust.

This data-driven analysis serves as a bridge between theory and practice, offering evidence for the necessity of “dynamic regulation” in addressing AI-driven media systems.

Key Indicator 1: Average Human Review Frequency per AI-Generated Text Content

Table 3.

Sensitivity Scenarios (Company A)

Human Review Level	Semantic Error Rate in Final Content	Editor Satisfaction with AI Output (Scale 0–10)	Average Time to Publication (Minutes)
No review (0 times)	34%	3.2	2
One-stage review	19%	6.5	6
Two-stage review	8%	8.1	10
Three-stage hybrid review	3%	9.3	14

While reducing review frequency increases publication speed, it significantly raises semantic errors and lowers editorial satisfaction. The analysis shows that two- to three-stage review processes represent an optimal balance between speed, accuracy, and trust.

Key Indicator 2: Discursive Diversity Index in the Recommender System

Table 4.

Algorithmic Scenarios (Company B)

Content Prioritization Model	Diversity Index	User Retention Rate	Reported Cognitive Polarization (%)
Raw engagement-based algorithm	0.38	71%	62%
Hybrid algorithm (Engagement + Topic Diversity)	0.54	68%	41%
Weighted algorithm prioritizing epistemic justice	0.63	66%	29%

When platforms prioritize content solely based on engagement, discursive diversity declines and cognitive polarization intensifies. Algorithms that weight topic diversity improve epistemic justice and reduce extremism—even with a modest decrease in user retention.

The above sensitivity analysis demonstrates that the proposed model is not only conceptual but operational and adaptable to data-driven policymaking. Empirical data confirm that even a single parameter change (such as review frequency or recommender logic) can have significant consequences in content accuracy, user experience quality, and media legitimacy.

This level of analysis goes beyond theoretical description and reveals that media management in the AI era must be based on monitoring, adaptation, and scenario design—scenarios that can simultaneously address professional, ethical, and technological goals.

Sensitivity Analysis at the Third Level: Media Policy-Making and Governance

Following the model's implementation in Companies A and B, a comparative evaluation was conducted at the third level (media governance) to assess the effects of having or lacking a formal, transparent ethical framework on key qualitative and quantitative variables related to public trust, user complaints, and narrative sustainability. For this purpose, the "Interactive Human–Machine Ethics Charter" was fully implemented only in Company B, whereas Company A operated without a formal directive, relying solely on implicit considerations.

Table 5.

Comparative Results (Six Months After Implementation)

Evaluation Indicator	Company A (No Formal Ethics Charter)	Company B (With Formal Ethics Charter)	Effect Difference	Analytical Interpretation
% of audience with high trust in content accuracy	54%	81%	+27%	The ethics charter substantially improved audience trust.
Avg. number of formal complaints about "misleading content"	17 per month	6 per month	–64%	Content origin transparency and ethical warnings were effective.
Avg. response time to user reports	3.9 days	1.6 days	–59%	Structured accountability policies streamlined the response process.
Organic reshare rate of AI-generated content	31%	49%	+18%	When users know the content's origin, they engage in more active resharing.
Editor satisfaction with AI-generated content review process	63%	89%	+26%	The ethics charter improved decision-making clarity and review efficiency.

The findings show that merely formulating a media ethics charter for AI—without changes to technical infrastructure or additional resources—can lead to meaningful improvements in audience trust, legal risk reduction, and media defensibility before regulatory bodies.

In effect, ethical policymaking is not only normative but produces measurable organizational outcomes. Thus, scaling the proposed model to the policy level is not a theoretical luxury, but a strategic necessity for media resilience in the emerging algorithm-driven global media order.

The data from the sensitivity analysis at the third level of the model (media governance) clearly demonstrated that the existence of a structured, transparent, and enforceable ethics charter in dealing with algorithmically generated content plays a decisive role in enhancing audience trust, reducing formal complaints, improving institutional accountability, and increasing the legal defensibility of media organizations. In fact, an ethics charter is not merely a normative instrument but also an operational component for optimizing internal processes, strengthening organizational coherence, and preserving media legitimacy.

In response to these findings, this article proposes an implementable and locally adaptable framework for developing a media ethics charter in the age of artificial intelligence—one that is inspired by international instruments and tailored to the institutional context of Iranian media organizations.

Objective of the Charter: To establish binding principles for professional and responsible engagement with AI-generated or AI-edited content, emphasizing content origin transparency, authenticity, epistemic justice, and institutional accountability.

Table 6.

Seven Core Principles of the Media Ethics Charter

Principle	Title	Operational Description
1	Content Origin Transparency	All content generated or edited by AI must be explicitly labeled as “AI-generated content.”
2	Authenticity and Factual Evaluation	Before publication, AI content must be reviewed by a human team for distortion, factual errors, and bias.
3	Ban on Non-Audited Models	Use of models lacking documentation, public testing history, or transparency reports is prohibited for news or analytical content.
4	Audience Right to Appeal	Users must be able to report suspected fake or biased content; review and response within 72 hours is mandatory.
5	Representation Fairness	Generative algorithms must not exclude or misrepresent minorities or marginalized groups; cultural diversity is mandatory.
6	Periodic Review and Revision	The charter must be reviewed and updated annually based on developments in AI models and field experience.
7	Institutional Responsibility	Even if content is produced by AI, ultimate responsibility lies with the media organization and must be specified in internal contracts.

Operational Requirements:

- Formation of an **AI Content Ethics Taskforce** including an editor-in-chief, algorithm specialist, legal expert, and audience representative.
- Development of a **human audit checklist** for algorithmic content.
- Creation of an **ethical flag module** in the internal CMS for logging, labeling, and monitoring AI-generated content.
- Submission of an **annual media ethics report** to the supervisory authority or board of trustees.

This study demonstrates that an ethics charter is not only a tool to enhance content quality but also a crucial mechanism for rebuilding audience trust, improving legal defensibility, and maintaining a balanced relationship between technological innovation and social responsibility.

Based on field evidence, the article emphasizes that media outlets without such a charter are exposed to multiple risks: legitimacy crises, erosion of trust, and legal disputes. Therefore, the drafting, approval, and implementation of this charter—as an operational supplement to the article’s three-tier model—should be regarded as a strategic imperative for strengthening resilience and intelligent media governance in Iran.

Managerial and Governance Implications in the AI Media Era

In light of the article's three-tier analysis, the operational implications for media managers, technology leaders, and cultural policymakers go beyond tactical adjustments and can be framed as paradigm-level transformations. Five key domains of managerial change emerge from this study:

1. **Transition from "Content Management" to "Meaning Regulation"**: In the age of AI, content managers' roles have shifted from passive oversight to proactive, data-driven regulation of narrative authority and legitimacy. Media managers must institutionalize algorithmic literacy, discourse analysis skills, and interdisciplinary capacities (in technology, law, and ethics) within their teams. This requires the development of "smart editors" and "meaning engineers" within the organizational structure.
2. **Designing Algorithmic Transparency Policies as a Competitive Advantage**: Media leadership in the new era is not driven solely by production speed but by the transparency of content production and distribution mechanisms. The implementation of internal explainability policies and algorithm audits should be considered a competitive edge and trust-building element in content markets. Drafting an AI code of conduct is essential for survival in future ethics-driven media ecosystems.
3. **Redesigning Editorial–Algorithmic Decision-Making Processes**: Media organizations must reengineer traditional editorial structures to enable content decisions at the human–machine interface. Establishing new roles such as algorithmic bias analyst, AI content ethics officer, and platform transparency manager is crucial to making AI-driven outputs accountable.
4. **Intelligent Management of Reputational and Legal Risks**: In the absence of national regulation, the legal and ethical responsibility for AI-generated content rests with the organization. Managers must deploy systems for content origin labeling, algorithmic error response, and production chain documentation to ensure institutional defensibility before regulators and the public.
5. **Advancing Data Governance and Interorganizational Diplomacy**: The future of media–platform relations lies not at the consumption level but at the level of data governance negotiation. Forward-looking media entities must participate in consortia, interorganizational councils, or joint ethical bodies to contribute to the development of national and regional standards for algorithmic content regulation.

Table 7.

Proposed Management Performance Indicators for AI-Driven Media

Domain	Key Indicator	Measurement Unit
Content Transparency	% of content with origin label	% of total content
Organizational Ethics	Average response time to user reports	Days
Human Oversight Quality	Error ratio in second vs. first content review	%
Discursive Justice	Diversity index in recommender system	Value between 0 and 1
Organizational Agility	Number of ethical policy updates per year	Count

Media management in the age of AI is no longer merely an operational function—it is a civilizational responsibility in designing the rules for negotiating meaning, power, and truth in the digital space. What lies ahead for media managers is not merely procedural reform but the engineering of a new governance paradigm at the intersection of technology, ethics, and policy—a governance model that can simultaneously facilitate production, empower audiences, and restore media legitimacy.

Discussion and Conclusion

The results of this study provide compelling empirical support for the theoretical model proposed to guide AI governance in the media sector. The three-tier framework—encompassing content production, algorithmic distribution, and media policy-making—demonstrates both conceptual robustness and operational feasibility. Through sensitivity analysis and comparative case studies within two Iranian private media organizations, the findings establish clear correlations between levels of human editorial intervention, algorithmic configuration, and audience-centered indicators such as trust, engagement, and complaint rates.

At the first level—content production—the study revealed that algorithmically generated news content requires multilayered human review to avoid semantic errors and bias propagation. Scenarios in which content was released without human review led to a 34% semantic error rate, while two- to three-stage editorial reviews reduced this to 3%. These findings reaffirm that language models, while capable of generating fluent and seemingly coherent content, lack inherent contextual awareness and ethical discernment [1, 2]. The results mirror previous concerns raised by scholars that foundation models, despite their fluency, cannot independently guarantee truth, fairness, or editorial relevance [3, 18].

At the second level—content distribution—recommender system configurations strongly influenced the diversity of information exposure and the degree of cognitive polarization among audiences. The use of engagement-only algorithms led to a discursive diversity index of just 0.38 and a polarization rate of 62%, whereas the integration of topic diversity metrics increased the index to 0.63 and reduced polarization to 29%. These outcomes validate critiques of platform-driven personalization systems, which often create echo chambers and fragment public discourse [16, 20, 24]. Similar concerns have been raised in the context of YouTube, TikTok, and other platforms, where algorithmic prioritization of user retention over informational pluralism contributes to ideological segregation [26, 27].

At the third level—media policy-making—the presence of a structured AI ethics charter had profound impacts. In the organization that implemented a formal ethics protocol, audience trust in content accuracy reached 81% (vs. 54% in the control case), complaint volume fell by 64%, and editorial satisfaction with AI-generated content increased by 26%. These data strongly support theoretical assertions that internal ethical governance mechanisms are essential in mitigating AI-driven risks, especially where national regulation is either absent or insufficient [4, 5, 17]. The outcomes also demonstrate that such charters are not merely normative instruments but drivers of institutional trust, professional cohesion, and audience empowerment [29, 30].

These findings align with the broader literature on algorithmic governance and public interest media. For instance, the role of editorial judgment in upholding democratic values has been widely discussed in the context of algorithmic disintermediation, where machine learning models increasingly mediate what becomes visible or prioritized in public discourse [7, 23]. The declining role of traditional newsrooms as epistemic authorities in the face of automated editorialization is consistent with our case study findings, in which algorithmic tools were shown to influence not only narrative structure but also audience behavior. The results reinforce warnings from critical media theorists about “data colonialism” and the extraction of human meaning for algorithmic optimization [8, 9].

Moreover, the success of internal governance measures in reducing complaints and improving editorial confidence echoes the argument that AI ethics must be embedded at the institutional level to be meaningful [13, 18]. Unlike abstract policy guidelines, internal charters—when backed by editorial boards, algorithm specialists, and legal advisors—can operationalize

ethical principles and tailor them to the organization's workflows and audience contexts [15, 21]. This supports the call for "co-governance" structures, where responsibility is shared across developers, editors, and oversight entities [16, 29].

The Iranian case further illustrates the asymmetry in AI policy readiness between regions, as many Global South countries face regulatory lag while attempting to adopt high-capability models developed in entirely different cultural and legal environments [22, 31]. The results of this study suggest that national adaptation is not only feasible but essential. Contextualized frameworks that combine global best practices with local editorial realities can mitigate imported risks while fostering responsible innovation [6, 11]. This also aligns with the concept of "AI sovereignty," as articulated in policy studies advocating for national strategies that reflect domestic values and priorities [19, 25].

One of the study's most critical contributions is its demonstration that internal editorial decisions and algorithmic configurations are not merely technical matters but deeply political. The implementation of an AI ethics charter significantly altered institutional behavior, improved responsiveness, and enhanced transparency. These findings support the growing argument that ethical design and governance of AI systems in media is not peripheral—it is core to the sustainability and legitimacy of journalism itself [20, 28]. Furthermore, they validate a shift in media ethics from being reactive and compliance-oriented to becoming proactive and anticipatory [10, 18].

Despite its contributions, this study is not without limitations. First, the case study method—though rich in contextual insight—limits the generalizability of findings beyond the Iranian media context. The results may not fully reflect the institutional, political, or technological constraints present in other countries or media environments. Second, the study relied on observational and self-reported data (e.g., editor satisfaction and complaint rates), which may be subject to social desirability bias or incomplete reporting. Third, the scope was limited to a six-month implementation period, which, although sufficient for initial sensitivity analysis, may not capture the long-term institutionalization of ethical governance practices or AI impacts on media culture.

Future studies should expand on this framework through cross-national comparisons, particularly between media systems in high- and low-regulation environments. Longitudinal research would also be valuable to assess how AI ethics charters evolve over time and how their adoption influences editorial independence and public trust in the long run. Experimental designs could isolate the effects of specific algorithmic interventions (e.g., fairness nudges, transparency labels) on audience perception and user engagement. Furthermore, interdisciplinary research is needed to understand how AI governance in media intersects with broader socio-technical systems, including education, law, and civic infrastructure.

Media organizations should treat AI adoption as a governance challenge, not just a technical upgrade. Establishing an internal AI ethics charter should be a non-negotiable step before deploying generative tools in content workflows. Editorial boards must collaborate closely with data scientists, legal experts, and audience advocates to ensure the fairness, transparency, and explainability of their systems. Regular ethical audits, participatory oversight mechanisms, and feedback loops should be institutionalized. Most importantly, newsrooms must invest in training the next generation of "algorithmically literate editors" who can bridge the gap between technological capability and public accountability.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

[1] T. B. Brown and et al., "Language models are few-shot learners," *Advances in Neural Information Processing Systems,* vol. 33, pp. 1877-1901, 2020.

[2] R. Bommasani and et al., "On the opportunities and risks of foundation models," 2022. [Online]. Available: https://arxiv.org/abs/2108.07258.

[3] K. Bontcheva, J. Posetti, and P. N. Howard, "Balancing freedom of expression and the challenges of AI-generated disinformation," 2024.

[4] D. Leslie and A. M. Perini, "Future Shock: Generative AI and the International AI Policy and Governance Crisis," 2024, doi: 10.1162/99608f92.88b4cc98.

[5] L. Edwards and I. Szpotakowski, "Private Ordering, Generative AI and the 'Platformisation Paradigm'," 2025. [Online]. Available: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/92790919A0203140ED012BF8A4BA8A0F/S3033373324000115a.pdf.

[6] W. H. K. Chun and B. S. Noveck, *Algorithmic Publics: Technology, Power and Democratic Futures*. NYU Press, 2025.

[7] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.

[8] N. Couldry and U. A. Mejias, *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Stanford University Press, 2019.

[9] R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press, 2019.

[10] A. Sandiumenge, "Unmasking bias: Algorithmic fairness in global media narratives," 2023.

[11] O. Kulesz, "Artificial Intelligence and International Cultural Relations: Challenges and Opportunities for Cross-Sectoral Collaboration," 2024. [Online]. Available: https://opus.bsz-bw.de/ifa/files/1203/ifa-2024_kulesz_ai-intl-cultural-relations.pdf.

[12] M. Oppedal, "Algorithmic neutrality and cultural invisibility in generative AI," *AI and Society,* vol. 38, no. 2, pp. 311-325, 2023.

[13] S. Hosseini and N. Kazemi, "Ethical Implications of Artificial Intelligence in Iranian Journalism: A Descriptive Approach," *Quarterly Journal of Ethics and Information Technology,* vol. 8, no. 1, pp. 23-48, 2022.

[14] F. Safari, "Futures Studies on the Application of Artificial Intelligence in Iranian News Agencies: An Exploratory Study," Allameh Tabataba'i University, 2021.

[15] L. Edwards, M. Szpotakowski, and C. da Mota, "AI, Media, and Policy: A Comparative Analysis," 2025.

[16] N. Helberger, J. Pierson, and T. Poell, "Governing online platforms: From contested to cooperative responsibility," *The Information Society,* vol. 36, no. 1, pp. 1-14, 2020, doi: 10.1080/01972243.2017.1391913.

[17] Oecd, "AI Innovation Concentration and the Governance Challenge," 2023. [Online]. Available: https://www.econstor.eu/bitstream/10419/299989/1/no292.pdf.

[18] S. A. Aaronson, "AI and the future of media trust: Challenges for global regulation," 2023.

[19] D. Leslie and L. Perini, "Algorithmic Bias and the Epistemic Crisis of Journalism," 2024.

[20] P. M. Napoli, *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. New York: Columbia University Press, 2019.

[21] M. da Mota, "Toward an AI Policy Framework for Research Institutions," 2024. [Online]. Available: https://www.cigionline.org/documents/2520/DPH-paper-daMota.pdf.

[22] T. Amirova, "Comparing Models of Artificial Intelligence Governance: The Role of International Cooperation on Responsible AI and the EU AI Act in the Age of Generative AI," 2023. [Online]. Available: https://cadmus.eui.eu/bitstream/handle/1814/76040/Amirova_2023_Master_STG.pdf?sequence=1.

[23] L. M. Vigl, "Evaluating the Ethical and Social Implications of AI in Public Broadcasting," 2024. [Online]. Available: https://repositum.tuwien.at/bitstream/20.500.12708/205282/1/Vigl%20Laura%20Maria%20-%202024%20-%20Evaluating%20the%20Ethical%20and%20Social%20Implications%20of%20AI...pdf.

[24] M. Ahmadi and S. Zare, "Analyzing the Performance of Recommendation Algorithms in Social Media and Their Impact on Audience Preference Formation," *Quarterly Journal of New Media Studies,* vol. 10, no. 2, pp. 55-74, 2019.

[25] Oecd, "Policy responses to the risks of generative AI in news ecosystems," 2023.

[26] S. A. Chun and B. S. Noveck, "AI Augmentation in Government 4.0: Introduction to the Special Issue on ChatGPT and other Generative AI Commentaries," *Digital Government: Research and Practice,* 2025. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3716859.

[27] S. A. Aaronson, "The Governance Challenge Posed by Large Learning Models," 2023. [Online]. Available: https://www2.gwu.edu/~iiep/assets/docs/papers/2023WP/AaronsonIIEP2023-07.pdf.

[28] K. Bontcheva, S. Papadopoulos, and F. Tsalakanidou, "Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities," 2024. [Online]. Available: https://lirias.kuleuven.be/retrieve/758830.

[29] Unesco, "Guidelines for the Governance of AI in Media and Culture," 2024.

[30] M. Vigl, "Post-editorial sovereignty: AI, editorial judgment and the fading newsroom," *Digital Journalism,* vol. 12, no. 1, pp. 1-18, 2024, doi: 10.1080/21670811.2024.2309090.

[31] O. Kulesz, "Artificial Intelligence and Cultural Content Governance," 2024.