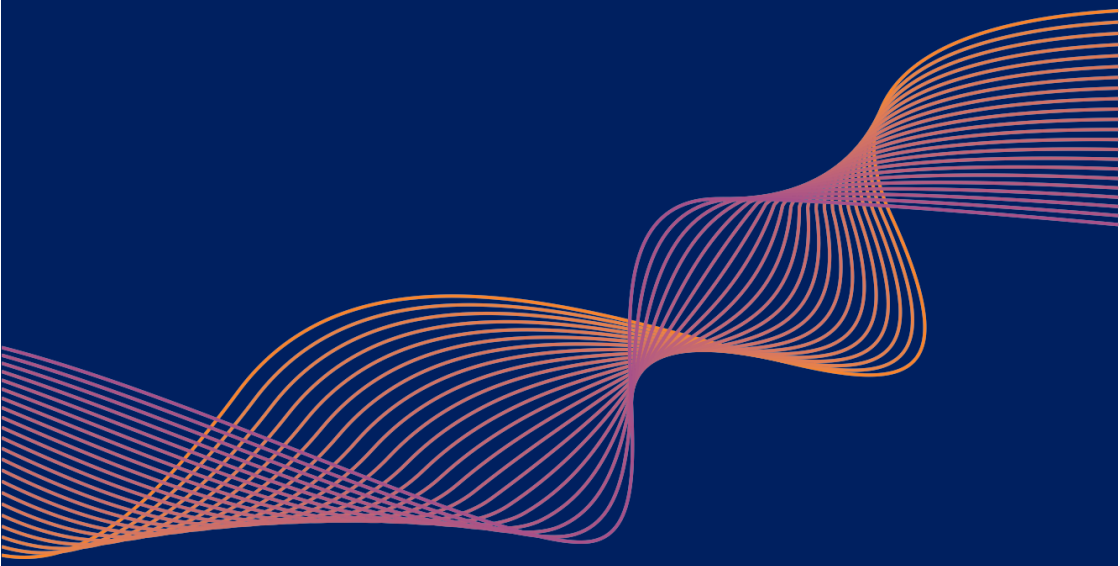
Next-Gen Intelligence: Media, Digital Marketing and AI



Ms. Richa Mahajan, Mr. Nitesh Kumar Khandelwal, Dr. Urvashi Sharma, Dr. Shipra Dua Piplani, Dr. Rajshree Chatterjee (Editors)

RESEARCH COLLAB.

Next-Gen Intelligence: Media, Digital Marketing and AI

Ms. Richa Mahajan
Assistant Professor, Department of Journalism & Mass Communication, Jagannath University, NCR.

Mr. Nitesh Kumar Khandelwal
Assistant Professor, Digital Marketing, Jagannath Community College, Delhi.

Dr. Urvashi Sharma
Associate Professor, Department of Journalism & Mass Communication, Jagannath University, NCR.

Dr. Shipra Dua Piplani
Associate Professor, Department of Journalism & Mass Communication, Jagannath University, NCR.

Dr. Rajshree Chatterjee
Associate Professor, Department of Journalism & Mass Communication, Jagannath University, NCR.

RESEARCH COLLAB.

Published by
Research Collab. 2026
India |

Networks: United Kingdom | South Africa | Cameroon | Tunisia | Morocco

Registration No.: AS-22-0003855
Bhaskar ID: IN-0426-9442QY
Official Website: www.researchcollab.in
Publication Portal: https://publication.researchcollab.in
Email: editor@researchcollab.in
© 2026 Research Collab. All rights reserved except where otherwise stated under the applicable Creative Commons license.

ISBN: 978-81-180547-0-2

E-ISBN: 978-81-180547-1-9

https://doi.org/10.5281/zenodo.23308752

Copyright © Richa Mahajan, Nitesh Kumar Khandelwal, Urvashi Sharma, Shipra Dua Piplani, and Rajshree Chatterjee

Citation: Mahajan, Richa; Khandelwal, Nitesh Kumar; Sharma, Urvashi; Piplani, Shipra Dua; and Chatterjee, Rajshree (Eds.). (2026). Next-Gen Intelligence: Media, Digital Marketing and AI. Research Collab.(https://doi.org/10.5281/zenodo.23308752)

This book is published online under a fully Open Access program and is licensed under the ***Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).*** Under the terms of this license, third parties may copy, share, adapt, and redistribute the material in any medium or format for non-commercial purposes, provided that appropriate credit is given to the author(s), a link to the license is provided, and any changes made are clearly indicated. The views, opinions, findings, and conclusions expressed in this publication are those of the respective author(s) and do not necessarily reflect those of the publisher or editors. The publisher, editors, and authors shall not be held responsible for any errors, omissions, or consequences arising from the use or application of the information contained in this book. No warranty, express or implied, is provided regarding the accuracy, completeness, or reliability of the content.While every reasonable effort has been made by the publisher, editors, and authors to ensure that the information presented is accurate and not misleading at the time of publication, no representation or warranty is made that all information, including material obtained from third-party sources, has been independently verified. The publisher remains neutral with regard to jurisdictional claims in published maps, institutional affiliations, and territorial boundaries. The publisher, editors, and authors have made every reasonable effort to identify and acknowledge all copyright holders for any material reproduced in this publication. If any copyrighted material has been used without proper acknowledgment or if any copyright holder has been inadvertently omitted, please contact the publisher. Appropriate corrections or acknowledgments will be incorporated in future editions or reprints.

Next-Gen Intelligence: Media, Digital Marketing and AI

PREFACE

Artificial Intelligence is no longer a distant technological possibility; it has become an integral part of the contemporary digital ecosystem. Its influence is increasingly visible across journalism, media production, photography, education, scholarly communication, digital marketing, consumer behaviour, governance, business management, and everyday social interaction. Against this rapidly evolving backdrop, *Next-Gen Intelligence: Media, Digital Marketing and AI* brings together diverse scholarly perspectives to examine how artificial intelligence is transforming communication, organisations, markets, institutions, and society.

As editors of this volume, we have sought to bring together contributions that reflect the multidisciplinary nature of artificial intelligence and its expanding relationship with contemporary media and digital practices. This volume moves beyond a purely technological discussion by examining AI as a social, organisational, communicative, educational, and ethical force, while also considering the opportunities, limitations, and challenges associated with its adoption.

The opening chapters explore the governance and policy dimensions of artificial intelligence, including ethical AI governance, digital governance in India, algorithmic accountability, transparency, privacy, human oversight, and the need to balance technological innovation with constitutional and democratic values. The discussion subsequently extends to journalism and communication, examining AI-assisted news production, audience engagement, citizen journalism, agentic AI in newsrooms, digital news credibility, and the evolving relationship between journalists, algorithms, and audiences.The volume also examines the transformation of visual communication. Discussions on AI and photography explore the transition from conventional camera-based practices to computational and generative imaging, raising important questions about authenticity, authorship, verification, and public trust. Similarly, the examination of deepfakes and misinformation highlights the growing challenges of distinguishing credible digital content from manipulated or synthetically generated media.

Education and scholarly communication constitute another significant dimension of this volume. Contributions addressing digital literacy, AI readiness among teachers, generative AI in journalism education, educational technology, and AI-assisted research and writing examine how educational institutions and researchers are responding to emerging technological realities. These discussions

emphasise the importance of critical thinking, academic integrity, human judgement, transparency, and the responsible use of artificial intelligence.

Furthermore, the volume explores the relationship between AI, society, consumers, and business. Discussions on digital culture, Generation Z and sustainable fashion, AI-driven consumer behaviour, personalisation, human oversight in AI-enabled targeting, managerial decision-making, strategic innovation, digital transformation, sustainability, branding, and competitive advantage demonstrate that AI adoption involves much more than the introduction of new technologies. Its effectiveness depends on human capabilities, organisational culture, leadership, data practices, ethical considerations, and responsible governance.

A defining strength of this volume is its interdisciplinary approach. The contributors include academics, researchers, professionals, and students who examine the changing digital environment from diverse disciplinary and practical perspectives. Their contributions collectively demonstrate that the future of artificial intelligence cannot be understood through technological developments alone. It must also be examined through the perspectives of communication, media ethics, education, management, public policy, consumer behaviour, digital culture, and human values.

This volume is intended for researchers, academicians, media professionals, students, educators, managers, policymakers, and readers interested in understanding the evolving relationship between artificial intelligence and the contemporary digital ecosystem. We hope it will serve as a useful reference for academic discussions and further research on responsible innovation, AI literacy, digital transformation, and the future of communication and business.

We express our sincere appreciation to all the contributors for sharing their research, insights, and perspectives. We are equally grateful to everyone who supported the editorial process and contributed to the development of this volume. Bringing together diverse perspectives on such a rapidly evolving subject has been both a significant responsibility and a valuable intellectual experience.

We hope that *Next-Gen Intelligence: Media, Digital Marketing and AI* encourages informed, critical, and constructive discussion rather than technological enthusiasm alone. Artificial intelligence will continue to evolve, but its impact on society will depend on the choices made by individuals, institutions, industries, educators, communicators, and policymakers. The central

challenge is therefore not merely to understand what AI can do, but also to consider how it can be developed and used with accuracy, accountability, creativity, inclusion, and human responsibility at its core.

The Editors

Ms. Richa Mahajan
Mr. Nitesh Kumar Khandelwal
Dr. Urvashi Sharma
Dr. Shipra Dua Piplani
Dr. Rajshree Chatterjee

TABLE OF CONTENTS

SI. No.	Name of Author(s)	Paper Title	Page No.
1	Varundeep Kaur	AI Governance Beyond Compliance: Ethical Challenges and Adaptive Governance of AI-Powered Digital Platforms	1–18
2	Ms. Pooja Sharma	Artificial Intelligence and Digital Governance in India: Opportunities, Ethical Challenges and Policy Directions	19–34
3	Mehak Arora	Artificial Intelligence in Journalism: Transforming News Production and Audience Engagement	35–49
4	Abhishek Chakraborty	Consumer Behaviour of Generation Z Towards Sustainable Fashion Products: An Exploratory Study in the Indian Context	50–61
5	Saburi Puri	Deep Fakes and Misinformation: Challenges for Digital Communication	62–76
6	Mr. Abhishek Kumar	Digital Literacy and AI Readiness among School Teachers	77–91
7	Jyanti Kumari	From Camera to Algorithm: The Transformation of Photography in the Age of AI	92–106
8	Sridevi B. and Jayaprakash M.	From Manuscript to Machine: AI-Assisted Research, Writing, and the Reimagining of Scholarly Communication in the Digital Ecosystem	107–119
9	Shaivi Singh	Technology, Society, and Digital Culture	120–130
10	Dr. Sandhya Godara	AI-Driven Consumer Behaviour: The Role of Data Analytics and Personalisation in Digital Business	131–147
11	Kashish Sahni	AI-Driven Citizen Journalism: Transforming	148–161

		Political Mobilization in India	
12	Dr. Priya Agarwal and Dr. Manjari Agarwal	EdTech Revolution in India: A Digital India Perspective towards Viksit Bharat 2047	162–171
13	Mr. Jwala Krishna Joshi and Ms. Anuja Dayal	Sustainable HRM, Green Marketing and Corporate Sustainability: An Integrated Business Perspective	172–184
14	Dr. Sheweta Gaur	Generative AI in Journalism Education: Transforming Teaching, Learning, and Professional Skills in the Digital Age	185–199
15	Pawan Raj	From Gandhian Journalism to the Contemporary Indian Digital News Feed: Reading Gandhiji's Editorial Record Against the Evidence on Algorithmic News Curation in India	200–213
16	Ms. Saumya Grover	Perceived Credibility of News Content on Instagram Reels among Generation Z: A Conceptual Perspective	214–233
17	Chandramouly Pandey	Integrating Agentic AI in the Newsroom Processes: Re-imagining News Delivery in India's Pluralistic Society	234–245
18	Haseeba Sayyed	Human Oversight and Control in the Era of AI-enabled Targeting	246–253
19	Dr. Pratibha Saini	From AI Adoption to Organizational Performance: The Mediating Role of AI-Driven Managerial Decision-Making	254–267
20	Simran Thakur	AI-Driven Business Management: Strategic Innovation, Digital Transformation, and Competitive Advantage	268–283
21	Sourav Sen and Monali Debnath	Computational Transparency in the AI Era: Ethical Divides in Digital Research Oversight	284–296
22	Yuvika and Bhavishya	The Impact of Social Media Branding and Digital	297–312

		Communication in the AI Era	
23	Saksham Khurana and Honey Gupta	Artificial Intelligence in Journalism and Mass Communication: The Future of News, Automation, Ethics, and Audience Interaction	313–325

Chapter 1

AI Governance Beyond Compliance: Ethical Challenges and Adaptive Governance of AI-Powered Digital Platforms

Varundeep Kaur, CIO, AIPPL
varundeepkaur@gmail.com

Abstract

Artificial intelligence is no longer limited to discrete applications such as prediction, automation, or decision support. It is increasingly embedded in digital platforms that determine what users see, what content receives attention, which products are recommended, how advertisements are targeted, and how online interactions are moderated. This expansion creates a governance problem that is broader than model accuracy or regulatory compliance. The ethical consequences of an AI-enabled platform can emerge from the interaction of models, data, business objectives, human decisions, platform design, and user behaviour, making accountability difficult to locate within any single system or function.

The chapter examines these challenges by bringing together scholarship on AI ethics, responsible AI, algorithmic accountability, platform governance, recommender systems, content moderation, and generative AI with international and Indian policy frameworks. The analysis identifies three related gaps in existing approaches. The first is vertical: how ethical principles are translated into organisational decisions and controls. The second is horizontal: how responsibility is managed when several AI systems, functions, vendors, and stakeholders interact. The third is temporal: how governance adapts when data, users, objectives, technologies, or consequences change after deployment.

To address these gaps, the chapter proposes the Ethical AI Governance Continuum (EAGC), comprising five dimensions, Purpose, Foundation, Decision, Consequence, and Adaptation. The framework extends existing risk and compliance approaches by connecting purpose, organisational decisions, post-deployment consequences, and adaptation within one governance cycle, while also making interaction across systems explicit.

Keywords: Artificial Intelligence, AI Governance, Responsible AI, Digital Platforms, Platform

Governance, Algorithmic Accountability

1. Introduction

Artificial intelligence has become an integral part of the digital platform infrastructure and in many cases users are unaware of the same.

Search engines rank information algorithmically and social-media platforms decide which posts appear in a user's feed. Recommendation systems shape which videos, products, or news stories get attention. Advertising systems select which messages are shown to particular audiences. Content moderation increasingly mixes human judgement with automated classification and ranking. Generative AI is everywhere, be it search, customer interaction, advertising, content creation, or any other platform functions.

These concerns are not entirely new. Algorithmic decision-making has long raised questions about bias, transparency, privacy, autonomy, and responsibility (Mittelstadt et al., 2016). The more recent development is the scale and interconnectedness of AI-enabled systems. Platform governance research has shown that algorithmic moderation and recommendation are not simply technical decisions; they can alter the broader rules, visibility structures, and participation conditions through which platforms govern users (Gorwa et al., 2020; Gillespie, 2022; Wright, 2022).

This can lead to problems for conventional governance approaches. Assessment of an individual model can be on its own terms however it can’t be understood in isolation. In order to understand it fully, it is important and required to know how and with what systems and in which manner it interacts with. When multiple systems operate together, it becomes really hard to govern. Commercial objectives often influence technical choices however the consequences of the deployment become visible only later.

The chapter considers the problem without assuming that existing governance frameworks have failed. The NIST AI Risk Management Framework, OECD AI Principles, UNESCO Recommendation on the Ethics of Artificial Intelligence, European Union AI Act, and India's Responsible AI initiatives each address important aspects of the problem. Rather, they serve different purposes. The question is whether there remains a conceptual space for a framework that connects purpose, organisational decision-making, system interaction, consequences, and adaptation.

The argument here is more modest. There still seems to be room for a framework that connects these pieces. The Ethical AI Governance Continuum (EAGC) is developed as a conceptual

framework for this purpose. It is not presented as a replacement for regulatory or risk-management frameworks. Instead, it is intended to connect areas that are often treated separately and to provide a basis for empirical research into adaptive AI governance.

2. From AI Ethics to AI Governance

Despite the fact that AI ethics and AI governance are concerned with different questions, they have an overlap. Ethics deals with what ought to be acceptable; governance with how an organisation turns that judgement into decisions, controls, responsibilities, and action. The distinction becomes clearer when considering common principles such as fairness, transparency, privacy, accountability, and meaningful human review.

There is now considerable international agreement that these principles matter. Jobin et al. (2019), for example, identified substantial convergence across global AI ethics guidelines around transparency, justice and fairness, non-maleficence, responsibility, and privacy, while also finding significant variation in how these principles are interpreted. Floridi and Cowls (2019) similarly developed a framework centred on beneficence, non-maleficence, autonomy, justice, and explicability.

The difficulty is not a shortage of ethical principles. The harder part begins when one of those principles has to guide a real organisational choice. Consider transparency. A governance team may require an AI system to be explainable, but that requirement still leaves several practical questions unresolved. What must be explained? To whom? At what stage? In what language? What happens if the explanation reveals a limitation that changes the decision to deploy the system?

3. What Existing Governance Frameworks Already Do

A useful starting point for developing another conceptual framework is to acknowledge what existing frameworks already accomplish. The NIST AI Risk Management Framework (AI RMF 1.0) provides a structured approach to managing AI risks across the lifecycle and organises activities around the functions Govern, Map, Measure, and Manage (Tabassi, 2023). It is deliberately designed as voluntary, rights-preserving, non-sector-specific guidance rather than as a regulatory instrument

The European Union AI Act adds a binding regulatory layer through a risk-based classification of AI systems and corresponding obligations (European Parliament & Council of the European Union, 2024).

India's policy direction has similarly moved beyond broad statements of principle. NITI Aayog's Responsible AI work identifies principles including safety and reliability, equality and inclusivity, privacy and security, transparency, accountability, and reinforcement of positive human values (NITI Aayog, 2021a, 2021b). More recent IndiaAI material also places emphasis on documentation, explainability, accountability, human supervisory control, and lifecycle responsibilities.

4. The Digital Platform Changes the Governance Problem

Digital platforms complicate AI governance because their outcomes are rarely produced by a single algorithm. A social-media environment, for example, may include systems for content recommendation, advertising, fraud detection, account security, content moderation, ranking, search, and generative assistance. These systems may be managed by different teams and may use different data, objectives, vendors, and monitoring processes.

Platform-governance research helps explain the practical significance of this problem. Gorwa et al. (2020) show that automated content moderation introduces challenges that cannot be reduced to classification accuracy. Their analysis highlights opacity, fairness, and the risk that automation may obscure the political and normative choices embedded in moderation. Gillespie (2022) extends this discussion by showing that reducing the visibility of content can itself function as a form of moderation. Recommendation and moderation, therefore, cannot always be treated as separate governance areas simply because different product teams manage them. Gillespie (2022) extends this discussion by showing that reducing the visibility of content can itself function as a form of moderation.

Recommendation and moderation, therefore, cannot always be treated as separate governance areas simply because different product teams manage them.

Wright (2022) makes a related argument in his study of automated moderation, showing that automation can alter the broader governance regime of a platform by changing visibility and scale.

Model risk

The possibility that an individual AI system produces an incorrect, biased, unsafe, or otherwise problematic result.

System risk

The possibility that interactions among AI-enabled components create consequences that are not visible

when each component is assessed separately.

Platform Environment risk

The possibility that the wider network of platforms, vendors, users, regulators, and institutions produces outcomes that cannot be attributed to one organisation or model. The distinction does not mean that model-level controls are unimportant. They remain necessary. The concern is that they may not be sufficient.

5. The Research Gap

The literature points to three gaps. They overlap, but they are not quite the same problem.

5.1 The Vertical Gap: Principles to Decisions

Existing ethics frameworks establish principles, while risk-management frameworks provide processes for identifying and managing risks. What remains less clear is how organisations translate ethical principles into everyday decisions about whether, where, and how AI should be used. The resulting gap sits between principle and organisational choice.

For example, an organisation may endorse fairness but still have to decide which fairness criterion should influence a recommendation system. It may endorse transparency but have to determine how much information a user needs before the explanation becomes meaningful. It may support meaningful human review while allowing the AI system to shape the decision so strongly that the human reviewer effectively becomes an approver.

What matters for governance is whether those principles actually influence decisions.

5.2 The Horizontal Gap: Responsibility Across Systems

Governance of AI systems is typically organized around specific systems, owners and phases of a lifecycle. Digital platforms however do not easily fit into these structures. Responsibility for a platform may be distributed across product, data scientists, infrastructure providers, marketing, content moderation, vendors, legal and senior management. Thus the interaction between these individuals and departments may be as important as the performance of individual models. Research on platform governance reaches similar conclusions.

5.3 The Temporal Gap: Governance After Deployment

Most governance frameworks recognise the importance of lifecycle management. Yet the practical

challenge of post-deployment adaptation deserves greater conceptual attention. Data, users, business objectives, models, connected systems, and regulation can all change. Behavioural responses may also differ from what designers expected. A governance decision made at deployment may therefore become inappropriate later without the system itself having technically failed.

This is the temporal gap: how governance changes when the context around an AI system changes. These three gaps provide the conceptual basis for the EAGC: translation from principles to decisions, coordination across interacting systems, and adaptation after deployment.

6. Where EAGC Fits

Ethical AI Governance Continuum (EAGC) is not meant to compete with existing standards, regulations, or lifecycle frameworks. Its contribution is narrower: it provides an integrated conceptual lens for examining three questions that tend to become separated in practice:

1. How are ethical principles translated into organisational decisions?
2. How does responsibility operate when AI systems and organisational functions interact?
3. How does governance change after deployment?

The distinction can be stated more precisely through the following comparison.

7. Ethical Challenges in AI-Powered Digital Platforms

7.1 Bias Is Not Only a Data Problem

Bias is often considered to be a training-data problem, however that is only a part of the story. Bias can arise from definition of the problem, which outcomes are selected, how labels are generated, how errors are distributed, and how a model is used after deployment. A historically accurate model can reproduce patterns that an organization may consider unacceptable.

This matters significantly when it comes to scale of the digital platforms, a small performance difference can result into a massive ripple effect. The right governance question is therefore not simply whether a model is statistically fair however it is whether the chosen definition of fairness is appropriate for the decision and context. For India, this question has an additional dimension. Language, geography, digital literacy, and social context can affect how people interact with AI systems and how accurately those systems interpret them.

7.2 Transparency Has to Be Meaningful

Transparency is sometimes reduced to documentation. Documentation matters, but it is only one part of meaningful transparency. A model card may be useful to a technical team but meaningless to a consumer. A technical explanation may satisfy an audit requirement without helping an affected person understand why a decision was made. The relevant issue is therefore who needs to understand what, and for what purpose.

This becomes especially important when AI is used in ranking, recommendation, advertising, moderation, or other platform functions where the user may not even know that an algorithm has influenced what they see. Gorwa et al. (2020) demonstrate why this matters in content moderation: automation can make already complex platform decisions more difficult to understand and audit.

7.3 Privacy Is Increasingly About Inference

AI changes the privacy problem because organisations can infer things from data that users never explicitly provided. The governance question extends beyond:

What data was collected? It also includes:

What was inferred, how was it used, and could the individual reasonably have anticipated that use? This is especially relevant to digital marketing, where behavioural data can be used to construct increasingly detailed profiles. India's Digital Personal Data Protection Act, 2023 establishes a legal framework for processing digital personal data and recognises both individual data-protection interests and the need to process data for lawful purposes (Government of India, 2023). (meity.gov.in) Statutory compliance does not settle the wider governance question. Organisations still need to consider whether an inference is proportionate to the stated purpose and whether its downstream consequences are understood.

7.4 Recommendation Can Become Governance

Recommendation is often treated as a commercial optimisation problem: a platform seeks to predict what a user will click, watch, purchase, or read. However, recommendation also controls visibility. Gillespie (2022) argues that reducing the visibility of content should itself be understood as a form of content moderation. This questions the fundamental distinction between recommendation and governance.

Generative AI can make content creation, personalisation, translation, and interaction much faster, which means that it can also make influential or misleading content easier to produce and that too at scale. Klenk (2024) highlights manipulation as a significant ethical concern associated with

generative AI and calls for design-oriented research into the conditions under which such systems can influence people. For digital marketing, the distinction between personalisation and manipulation becomes increasingly difficult.

Personalization can improve relevance, however manipulation begins to become a concern when systems exploit vulnerabilities, hides the true intent, or optimize behavior beyond reasonable user understanding and comprehension. This does not mean that AI-driven influence is inherently unethical however it does suggest that the degree and nature of influence should form part of governance assessment.

7.5 Content Moderation and Human Agency

Automated moderation is a good example of why a purely technical view of AI governance can be limiting. Gorwa et al. (2020) found that algorithmic moderation can introduce new problems of opacity and fairness. More recent research has also emphasised the role of human agency in platform moderation, particularly where moderators exercise contextual judgement within complex governance networks (Zhao & Zhang, 2025).

That is why the phrase human-in-the-loop needs more scrutiny than it usually receives. A person who simply reviews an AI-generated recommendation without sufficient context or authority may not provide meaningful oversight. The quality of oversight depends on whether the human reviewer has:

- relevant competence;
- sufficient information;
- time to exercise judgement;
- authority to override the system;
- and clear accountability for the decision.

The issue is consequently not the presence of a human but the quality of human agency within the decision process.

8. India: Why Context Matters to AI Governance

India offers an especially useful setting for examining adaptive AI governance. Its digital environment combines rapid adoption with substantial linguistic, social, geographic, and economic diversity. The issue is not simply market size. A governance framework developed in a relatively

homogeneous environment may assume that users interact with technology in similar ways, have comparable levels of digital literacy, and are represented adequately in the underlying data. Those assumptions become more difficult to sustain in India's context.

The country's work on multilingual AI makes the point quite clearly. Project BHASHINI, for example, explicitly seeks to strengthen language technologies across Indian languages and includes community participation in the development and validation of language data. Its programmes also identify under-resourced Indian languages as an area requiring focused research and collaboration. (bhashini.gov.in; bhashini.gov.in) Language is a governance issue because it affects more than accessibility.

Hence, linguistic inclusion becomes part of responsible AI rather than simply a user-experience concern. India's digital public infrastructure (DPI) also changes the scale at which AI-enabled systems can operate. The IndiaAI Mission is developing national infrastructure around compute, datasets, foundation models, future skills, applications, startups, and safe and trusted AI. The scale creates both opportunity and governance responsibility.

As AI gets integrated into both public and private digital infrastructure, the errors and exclusions get amplified to a greater extent that they would have in a insolated environment or application. A well-governed system may improve access and inclusion at considerable scale.

The Indian case hence supports context-sensitive governance. Universal principles remain useful, but their implementation needs to consider he social and technological environment in which AI operates. Within EAGC, that context affects the Foundation, Decision, Consequence, and Adaptation stages rather than simply remaining outside of the governance cycle.

9. The Ethical AI Governance Continuum

EAGC is organised around five connected dimensions:

Purpose → Foundation → Decision → Consequence → Adaptation

The sequence begins before model development, moves through the point at which AI influences organisational decisions, examines what happens after deployment, and then feeds those observations back into governance.

9.1 Purpose: Why Is AI Being Used?

Governance often starts after the organisation has already decided to use AI. EAGC begins one step

earlier. The purpose stage asks:

- What problem is AI intended to solve?
- Why is AI appropriate?
- What alternatives were considered?
- Who benefits?
- Who might bear the risks?
- What would happen if AI were not deployed?

This matters because technological capability can create pressure to automate simply because automation is possible. Purpose clarity provides a basis for determining whether the proposed use is legitimate, proportionate, and aligned with organisational objectives.

9.2 Foundation: What Supports the System?

Once the purpose is clear, attention turns to the foundation:

- data provenance;
- data quality;
- representativeness;
- privacy;
- model choice;
- assumptions;
- validation;
- security;
- documentation;
- and known limitations.

The purpose stage therefore asks whether the organisation should use AI; the foundation stage asks whether the proposed system is capable of supporting that purpose responsibly. A mismatch between the two can create risk even when the model performs well technically.

9.3 Decision: How Does AI Influence Judgement?

AI does not always make the final decision. It may recommend, rank, classify, generate, prioritise, filter, or flag information for human consideration, and these different forms of influence create different governance requirements. An AI system that recommends ten options to a human is different

from one that automatically selects one option. The degree of human discretion changes the governance requirement.

The Decision dimension examines:

- where AI has authority;
- where humans retain authority;
- what information the human receives;
- whether overrides are possible;
- and whether the organisation has defined escalation mechanisms.

9.4 Consequence: What Happens in Practice?

The consequence stage asks what actually happened after the system entered use. The organisation should examine:

- intended outcomes;
- unintended effects;
- user complaints;
- differential impacts;
- changes in behaviour;
- misinformation;
- privacy incidents;
- trust;
- and effects on groups not directly represented in the original design process.

This dimension is particularly important for platforms because the consequences of ranking, recommendation, and moderation can emerge from user interaction with the wider system rather than from an isolated model output.

Wright's (2022) analysis of automated moderation illustrates why this broader perspective matters: automation can change what becomes visible within a platform's governance system even when the underlying algorithm is accurate.

9.5 Adaptation: When Should Governance Change?

The final dimension accepts a fairly obvious point that governance discussions sometimes overlook: conditions change. Reassessment may be required when:

- its data changes;
- its user population changes;
- its business objective changes;
- a new model is introduced;
- another AI system is connected;
- a material incident occurs;
- regulation changes;
- or evidence of unintended consequences emerges.

Adaptation therefore includes monitoring, reassessment, incident response, policy revision, model modification, restriction, suspension, replacement, and retirement. Deployment is therefore a point in the cycle, not its end.

10. What EAGC Adds to Existing Governance Approaches

EAGC is not another five-principle ethics model. Its contribution lies in the connections between stages and in the feedback between deployment experience and the assumptions that preceded deployment. Existing frameworks already provide mechanisms for risk identification, documentation, monitoring, meaningful human review, transparency, and accountability. EAGC brings those mechanisms into a sequence that begins with purpose and ends with organisational learning.

Three features deserve particular attention.

First, purpose precedes controls.

A governance process should not assume that every technically feasible use of AI is justified. The purpose stage creates a point at which the organisation can question the use case itself.

Second, consequences are treated as governance evidence.

A system should not be regarded as adequately governed merely because it passed pre-deployment controls. Evidence from actual use can change the assessment.

Third, adaptation closes the loop.

The organisation is expected to learn from deployment and reconsider earlier assumptions where necessary. This makes EAGC complementary to rather than competitive with frameworks such as NIST AI RMF.

11. From Individual Model Risk to Platform Environment Governance

EAGC also extends the earlier distinction between model risk, system risk, and wider platform risk. Suppose a platform uses:

- one model to identify user interests;
- another to rank content;
- another to generate advertising copy;
- another to detect policy violations;
- and a fifth system to measure engagement.

Each system may have its own governance process, yet the combined platform may create a feedback loop. User behaviour generates data. The data influences recommendations. Recommendations influence behaviour. Behaviour generates new data. Advertising responds to the new profile. Engagement metrics then become evidence for further optimisation. No single model necessarily creates the entire outcome. Part of the governance problem therefore exists between systems.

12. Theoretical Logic of the Research Propositions

The six propositions follow from the relationships proposed by the framework. They are intended to make the argument testable; they are not presented as validated causal claims. The logic is:

Governance condition → organisational mechanism → expected outcome The propositions put those relationships into a form that can be tested later.

Proposition 1: Purpose Clarity

Mechanism: A clearly defined purpose establishes boundaries around what the AI system is expected to achieve and what uses fall outside its legitimate scope. This reduces ambiguity during design and deployment and creates a reference point against which later changes can be assessed.

Proposition 1: Greater clarity of AI purpose is associated with lower likelihood of technology-driven deployment in which the availability of AI becomes the primary justification for its use.

The proposition can be examined through organisational case studies comparing AI initiatives with different levels of purpose definition.

Proposition 2: Governance Integration

Mechanism: When ethical considerations are embedded within product, risk, technology, and business decisions, they are more likely to influence actual behaviour than when they remain separate compliance requirements.

Proposition 2: The integration of responsible-AI considerations into core organisational decision processes is positively associated with governance effectiveness, particularly where those considerations influence deployment, escalation, and resource decisions.

This proposition follows from the gap between stating a principle and making it part of an operating decision.

Proposition 3: Meaningful Meaningful Human Review

Mechanism: Human involvement can only affect outcomes when reviewers possess the competence, information, and authority necessary to challenge AI-generated recommendations.

Proposition 3: The effectiveness of meaningful human review is positively associated with the competence, information access, and decision authority available to human reviewers. This proposition could be tested through studies of moderation, recruitment, financial decision-making, or customer-service applications.

Proposition 4: System Interaction

Mechanism: When AI systems interact, risks can move across system boundaries. A control applied to one model may therefore fail to address a consequence created by the interaction between multiple models.

Proposition 4: Greater interaction among independently governed AI systems is associated with a greater need for governance mechanisms that examine dependencies, risk transfer, and emergent effects across system boundaries.

The proposition is especially relevant to digital platforms, where several forms of AI-mediated influence may coexist.

Proposition 5: Post-Deployment Adaptation

Mechanism: AI systems operate within changing environments. Monitoring provides information about actual outcomes, allowing organisations to revise assumptions and controls.

Proposition 5: The quality and continuity of post-deployment monitoring are positively associated

with an organisation's ability to detect governance drift and revise AI controls under changing operational conditions. Longitudinal studies of deployed AI systems could examine this relationship.

Proposition 6: Trust and Accountability

Mechanism: Technical performance does not fully determine how stakeholders perceive an AI-enabled decision. Perceptions of accountability, transparency, and organisational responsibility may influence whether stakeholders regard the use of AI as legitimate.

Proposition 6: Perceived organisational accountability and meaningful transparency are positively associated with stakeholder trust in AI-enabled digital services, controlling for perceived system performance. This proposition opens a connection between AI governance and research in digital trust, consumer behaviour, and platform legitimacy.

13. Implications for Digital Marketing

AI now sits in many parts of marketing, including segmentation, recommendation, targeting, content generation, campaign optimisation, customer service, and behavioural prediction. The governance issue is not whether these applications should exist. Many provide clear value. A more difficult question concerns the boundary between relevance and influence. A recommendation can help a consumer find something useful. The same capability can be used to identify when a consumer is particularly susceptible to a message. Generative AI can make campaign content more relevant, but it can also make highly personalised persuasion inexpensive and scalable.

At the Consequence stage, it should monitor not only conversion but also complaints, exclusion, unintended targeting, and other effects. At the Adaptation stage, the organisation should be willing to change or stop practices when evidence suggests that the original assumptions no longer hold. This approach does not require marketing to become less data-driven. It requires marketing to recognise that optimisation itself can have governance consequences.

14. Implications for Media and Journalism

Media organisations face a related problem, although the stakes are somewhat different. AI can affect commercial outcomes, but it can also shape the information people encounter. AI is increasingly relevant to transcription, translation, search, recommendation, content production, fact-checking, and audience analysis. Recommendation makes this governance issue especially visible. When platforms prioritize certain stories, they shape what people see without actually deleting any content. Gillespie's

(2022) analysis of reduction as a form of content moderation is relevant here because it demonstrates that controlling how visible something is can be used to govern it.

15. Implications for Boards and Senior Management

AI governance is increasingly an organisational governance issue. Boards do not need to understand the mathematics of every model. They do need to understand how AI affects organisational risk, reputation, customers, employees, and strategic decisions. A board-level discussion can begin with seven questions:

16. A Governance Maturity Heuristic Through the EAGC Lens

EAGC can also be used as a simple maturity guide, provided the levels are treated as a heuristic rather than a mandatory sequence.

17. Practical Governance Questions by EAGC Dimension

The framework can be translated into practical questions without turning it into another compliance checklist. The table is deliberately framed around questions rather than controls. Its purpose is to support judgement without turning EAGC into another tick-box exercise.

18. Future Research Agenda

The framework also leaves several questions open for empirical research.

18.1 Testing EAGC as a Governance Maturity Framework

Future studies could develop measurement scales for the five dimensions and examine whether organisations with stronger EAGC characteristics experience fewer AI-related incidents or greater stakeholder trust.

18.2 Cross-Platform Research

Comparative studies could examine how social-media, e-commerce, financial, educational, and media platforms govern AI differently.

18.3 AI-to-AI Governance

Research should examine what happens when independently governed AI systems exchange data, recommendations, classifications, or decisions. This may become increasingly important as agentic

systems and AI-enabled workflows become more interconnected.

19. Conclusion

AI governance is becoming more difficult as AI gets deeper into products, platforms, workflows, and routine decisions. Earlier the governance questions were largely if the organizations had identified major AI related ethical risks however now the additional question is more operational one – if governance is effective once AI gets embedded or integrated into products, platforms, workflows, and decisions that interact with one another.

Existing frameworks provide important pieces of this architecture. NIST offers risk-management guidance; OECD and UNESCO provide normative foundations; regulation such as the EU AI Act establishes binding requirements; and India's Responsible AI initiatives are developing principles and operational guidance appropriate to the country's context. None of these approaches needs to be displaced by the argument advanced here.

References

- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics, 5*, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w
- European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1
- Gillespie, T. (2022). Do not recommend? Reduction as a form of content moderation. *Social Media + Society, 8*(3), 1–13. https://doi.org/10.1177/20563051221117552
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society, 7*(1). https://doi.org/10.1177/2053951719897945
- Government of India. (2023). *The Digital Personal Data Protection Act, 2023 (No. 22 of 2023).* Ministry of Law and Justice.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

- Klenk, M. (2024). Ethics of generative AI and manipulation: A design-oriented research agenda. *Ethics and Information Technology, 26,* Article 9. https://doi.org/10.1007/s10676-024-09745-x
- Kokshagina, O., Reinecke, P. C., & Karanasios, S. (2023). To regulate or not to regulate: Unravelling institutional tussles around the regulation of algorithmic control of digital platforms. *Journal of Information Technology*, *38*(2), 123–140. https://doi.org/10.1177/02683962221114408

- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679
- NITI Aayog. (2021a). *Responsible AI: Approach document for India, Part 1: Principles for responsible AI.*
- NITI Aayog. (2021b). *Responsible AI: Approach document for India, Part 2: Operationalizing principles for responsible AI.*
- Organisation for Economic Co-operation and Development. (2024). *OECD AI principles.* https://oecd.ai/en/ai-principles
- Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Chapter 2

Artificial Intelligence and Digital Governance in India: Opportunities, Ethical Challenges and Policy Directions

Ms. Pooja Sharma, Assistant Professor
Jagannath Community College
pooja071999sh1@gmail.com

Abstract

Artificial Intelligence is reforming governance by enhancing administrative efficiency, refining public service delivery, and aiding evidence-based policymaking. In India, AI has become a significant constituent of the digital governance ecosystem, accompanying initiatives such as Digital India, Aadhaar, DigiLocker, CoWIN, and other digital public substructures. While AI offers transformative opportunities in areas including healthcare, education, agriculture, law enforcement, and public administration, its rapid incorporation into governance also raises intricate ethical, legal, and political concerns. Problems relating to algorithmic bias, data privacy, scrutiny, transparency, accountability, digital segregation, and cybersecurity challenge the values of democratic governance and citizens' rights. This chapter critically inspects the evolving role of AI in India's governance agenda by analysing both its developing potential and the ethical dilemmas connected with its distribution. Drawing upon present-day policy initiatives, nationwide AI strategies, and worldwide governance frameworks, the study evaluates India's preparedness to normalize AI while promoting innovation and comprehensive development. It contends that effective AI governance needs a balanced policy approach that fit in technological development with constitutional values, ethical precautions, institutional responsibility, and citizen involvement.

Keywords: Artificial Intelligence, Digital Governance, AI Ethics, Public Policy, Ethics

1. Introduction

1.1 Background

Artificial Intelligence (AI) has become a transformative technology in public governance, enabling governments to analyse large datasets, automate processes, support evidence-based policymaking, and improve service delivery. Unlike earlier e-governance approaches focused primarily on digitisation, digital governance integrates AI, big data, cloud computing, and other technologies to facilitate proactive and adaptive administration (Margetts & Dorobantu, 2019; Russell & Norvig, 2021).

1.2 Context of Digital Governance in India

India's digital governance ecosystem, supported by initiatives such as Digital India, Aadhaar, DigiLocker, UPI, and CoWIN, has significantly reshaped citizen state interactions. AI increasingly complements this infrastructure through applications in healthcare, education, agriculture, law enforcement, and public administration. NITI Aayog's *National Strategy for Artificial Intelligence* promotes an "AI for All" approach, while the IndiaAI Mission further supports responsible AI development and public-sector adoption (NITI Aayog, 2018; MeitY, 2023).

1.3 Problem Statement

Despite these opportunities, AI-enabled governance raises concerns relating to algorithmic bias, privacy, transparency, accountability, cybersecurity, and digital exclusion. These challenges are particularly significant in India, where technological innovation must remain consistent with constitutional rights, equality, social justice, and democratic accountability (Floridi & Cowls, 2019; UNESCO, 2021).

1.4 Research Gap

Existing scholarship often examines AI, digital governance, and public administration separately. Limited interdisciplinary analysis addresses their intersection within India's constitutional and democratic framework.

1.5 Purpose and Objectives of the Chapter

This chapter critically examines AI's opportunities and challenges within India's digital governance ecosystem and evaluates policy directions for transparent, accountable, ethical, and citizen-centric AI governance.

1.6 Scope

The chapter focuses on AI-enabled governance in India's public sector, particularly healthcare, education, agriculture, law enforcement, and public administration, with selective international comparisons.

1.7 Organization of the Chapter

The chapter examines conceptual foundations, contemporary literature, India's AI governance ecosystem, ethical and legal challenges, policy frameworks, international approaches, and recommendations before concluding with key pathways for responsible AI governance in India.

2. Conceptual and theoretical foundations.

2.1 Understanding Artificial Intelligence

Artificial Intelligence (AI) is a field of computer science that involves the simulation of human intelligence and abilities in computer systems, such as learning, reasoning, problem-solving, and decision-making (Russell & Norvig, 2021). In the governance sector, AI can be used for data analysis, service delivery, and evidence-based policymaking, yet it is essential to have good data quality, institutional capacity, human oversight, and accountability.

2.2 Digital Governance

Digital governance goes beyond e-governance and combines AI, big data, cloud computing, and digital infrastructure with citizen-centric administration (Margetts & Dorobantu, 2019). This shift is exemplified by initiatives such as India's Digital India, Aadhaar, DigiLocker, and CoWIN, where AI is playing a more significant role in enhancing administrative effectiveness and data-driven decision-making.

2.3 AI Governance

AI governance refers to legal, ethical, institutional and regulatory frameworks that shape AI design, deployment, monitoring and oversight (OEC, 2019). It covers issues of reliability, explainability, privacy, fairness, human rights, and institutional accountability.

2.4 Ethical AI

The principles of ethical AI are designed to ensure that AI systems are consistent with human dignity, democratic principles, and human rights (Floridi & Cowls, 2019; UNESCO, 2021). To avoid automated decisions that are discriminatory or opaque, it is important to have fairness, transparency, explainability, accountability, and human oversight.

2.5 Democratic Governance and Constitutional Values

Governance through AI should be accountable, transparent, participatory, responsive and governed by the rule of law (United Nations, 2015). In India, constitutional equality, fundamental rights, procedural fairness and privacy are important parameters, bolstered by the recognition of privacy as a fundamental right in Justice K. S. Puttaswamy (Retd.) v. Union of India (2017).

2.6 Theoretical Frameworks

2.6.1 Good Governance Theory

AI must help enhance democratic legitimacy, as Good Governance is built on the principles of accountability, transparency, participation, effectiveness, equity, and responsiveness (World Bank, 2017).

2.6.2 New Public Governance

The emphasis of New Public Governance is on the collaborative nature of public governance, involving government, private organisations, civil society and citizens (Osborne, 2010), which is relevant to multi-stakeholder AI governance.

2.6.3 Algorithmic Governance

Algorithmic governance is a field that investigates the impact of algorithms on policy and administrative decisions, and raises issues related to opacity, bias, accountability, and democratic oversight (Danaher et al., 2017).

2.6.4 Responsible AI Framework

Responsible AI is integrated throughout the AI lifecycle, covering all stages from design and development to deployment and use, with a focus on fairness, transparency, accountability, privacy, robustness, and human-centred design (OECD, 2019; UNESCO, 2021).

2.6.5 Digital Era Governance

The concept of Digital Era Governance offers insights into how digital technologies are transforming governance, encompassing integrated services, citizen engagement, and data-driven governance (Dunleavy et al., 2006).Digital Era Governance provides a framework for understanding how digital technologies are transforming governance, including the use of integrated services, citizen engagement, and data-driven governance (Dunleavy et al., 2006).

2.7 Proposed Analytical Framework

This chapter brings these discussions together and explores the intersection of technological innovation, ethical governance, institutional accountability and democratic values in the context of AI. It assesses governance in terms of efficiency and service delivery, constitutional protection, ethical safeguards, regulatory oversight, transparency, and citizen participation.

3. Literature Review

3.1 Evolution of AI Governance Scholarship

The scholarship of the AI governance movement is evolving and changing. While early AI governance scholarship focused on technical issues of algorithmic performance, it has now turned to multi-disciplinary questions of ethics, law, public policy, institutional accountability, and democratic legitimacy. AI governance is now increasingly being understood in the literature as a framework for responsible use, based on public value and constitutional values (OECD, 2019; UNESCO, 2021). Smoothing the path for the future of public administration.

3.2 AI and Public Administration.

AI's potential to enhance administrative efficiency, predictive policymaking, resource allocation, and evidence-based governance has been highlighted by research (Margetts & Dorobantu, 2019; Wirtz et al., 2019). But experts warn that efficiency is not enough for transparency and accountability. AI should therefore complement, not supplant, administrative judgment, and build institutional capacity and digital literacy (Sun & Medaglia, 2019).

3.3 AI in Public Service Delivery

There are important areas of AI application identified in literature: healthcare, education, agriculture, welfare, and urban governance. AI has the potential to increase access and increase resource efficiency, but access to AI, literacy and institutional capacity may be unequal, which could lead to an unequal distribution of benefits. Therefore, inclusive digital governance is still a vital need (World Bank, 2021).

3.4 Ethical Debates

The key ethical issues include algorithmic bias, privacy, transparency, explainability, accountability, and human oversight (Floridi & Cowls, 2019; UNESCO, 2021). Bias is now understood as a symptom of larger institutional inequities not just a technical flaw, and opaque automated decisions are problematic for their lack of fairness and democratic legitimacy.

AI is being explored in the context of Digital India, Digital Public Infrastructure, IndiaAI Mission, and changing data protection laws in the Indian scholarship landscape. Issues of particular concern are digital exclusion, algorithmic discrimination, privacy, cybersecurity, institutional capacity and regulatory fragmentation. This literature provides support for the approach to the evaluation of AI governance through the democratic and constitutional framework of India.

The researcher has synthesized the literature in the following manner:

Overall, there is a recognition that AI can help to enhance governance, but that technological capability is not enough to guarantee legitimate outcomes. The world of literature is becoming more unified, with Responsible AI combining innovation with ethical protections, institutional responsibility, and multi-stakeholder governance.

Research Gaps

There are significant gaps in the integration of AI, digital governance, public administration, ethics, and constitutional values into a unified framework in India. There is still little empirical evidence on the long-term cross-sectoral outcomes, and less research on institutional preparedness, regulatory capacity, citizen trust, digital inclusion, and participatory governance**.**

4. Artificial Intelligence in India's Digital Governance Ecosystem

4.1 Evolution of India's Digital State

India's digital governance ecosystem has evolved from the digitisation of administrative processes towards an integrated Digital Public Infrastructure (DPI) supporting large-scale public service delivery. While early e-governance initiatives primarily sought administrative efficiency and electronic service delivery, contemporary digital governance increasingly relies on interoperable platforms, data-driven decision-making, and AI-enabled systems. This represents a transition from technology-assisted administration towards intelligent governance, in which AI functions as a strategic enabler of proactive and evidence-based governance.

The Digital India Programme, launched in 2015, established digital infrastructure, electronic service delivery, and citizen empowerment as important governance priorities (Ministry of Electronics and Information Technology [MeitY], 2023). Initiatives such as Aadhaar, DigiLocker, Unified Payments Interface (UPI), Government e-Marketplace (GeM), and CoWIN have further strengthened India's digital governance architecture. These platforms provide infrastructure and data capabilities that can support predictive analytics, policy evaluation, and evidence-based governance.

However, India's digital transformation involves more than technological modernisation. Digital systems increasingly influence policy formulation, implementation, monitoring, and citizen–state interaction. Within this environment, AI has become an important component of India's governance ecosystem, complementing digital initiatives while creating new opportunities and governance challenges.

4.2 AI Across Government Sectors

AI's value in governance lies primarily in strengthening, rather than replacing, institutional and administrative capacities across policy sectors.

Healthcare. AI can support disease surveillance, medical imaging, clinical decision support, resource planning, and predictive public health interventions. During the COVID-19 pandemic, CoWIN demonstrated the capacity of digital infrastructure to support nationwide vaccination management and large-scale coordination. Although CoWIN was not itself an AI system, its data infrastructure illustrates the foundations upon which AI-enabled public health analytics can be developed (World Bank, 2021). However, AI-based recommendations should complement professional and clinical judgement, particularly in ethically sensitive or uncertain situations.

Education. AI-enabled educational technologies can support personalised learning, adaptive instruction, automated assessment, and early identification of students requiring additional assistance. However, unequal digital access, infrastructure, and digital literacy may limit these benefits, particularly among rural and disadvantaged populations. AI should therefore complement rather than replace the pedagogical and social functions of teachers.

Agriculture. AI applications in agriculture include crop monitoring, pest detection, weather forecasting, precision irrigation, yield prediction, and geospatial analysis. These tools can support timely agricultural advice, resource management, and disaster preparedness. Their effectiveness, however, depends on reliable connectivity, accurate local data, institutional coordination, and farmers' awareness and capacity to use digital services.

Public Administration. AI can strengthen public administration through decision-support systems, document processing, grievance management, workflow automation, and analysis of administrative data. Such applications can help governments identify policy trends, monitor programmes, and improve resource allocation (Wirtz et al., 2019). Nevertheless, AI should augment administrative capacity rather than replace public officials, since contextual judgement, ethical reasoning, and discretion remain essential where decisions affect citizens' rights and interests.

Law Enforcement. AI is increasingly used in predictive analytics, crime mapping, facial recognition, and digital forensics. These applications may improve investigative efficiency and

resource allocation but also create significant risks. Predictive systems can reproduce historical biases embedded in policing data, potentially undermining fairness and public trust. Law-enforcement applications therefore require particularly strong safeguards, transparency, accountability, and appropriate oversight (UNESCO, 2021).

4.3 Benefits of AI Integration

AI can generate significant benefits for India's digital governance when supported by appropriate institutional and ethical safeguards. First, automation can improve administrative efficiency by reducing routine workloads and procedural delays, allowing public officials to focus on complex decisions requiring human judgement. Second, predictive analytics can strengthen evidence-based policymaking by identifying trends, anticipating service requirements, and evaluating programme outcomes.

Third, AI can improve public service delivery through more accurate targeting of welfare programmes, responsive citizen services, and continuous monitoring. Fourth, data-driven systems can support transparency through performance monitoring and digital record management. However, transparency depends on explainability and effective accountability mechanisms.

AI can also strengthen governmental capacity to address complex policy challenges by integrating diverse datasets into analytical frameworks. These benefits should not be understood as technologically automatic. Their effectiveness depends upon institutional competence, regulatory oversight, data quality, ethical safeguards, and citizen trust. Thus, efficiency must remain balanced with constitutional protections, democratic accountability, and inclusive governance.

4.4 Contribution to Sustainable Development Goals

AI-driven digital governance can help achieve the Sustainable Development Goals (SDGs) through the implementation of inclusive public policy. AI can contribute to SDG 3 (Good Health and Well-being) by better healthcare planning, to SDG 4 (Quality Education) by personalised learning and to SDG 2 (Zero Hunger) by precision agriculture and better food-security planning. AI-enabled public administration can also help achieve SDG 16 (Peace, Justice and Strong Institutions), through enhanced institutional effectiveness, public service delivery and evidence-informed policy-making (United Nations, 2015).

But, technology adoption is not enough for sustainable development. Digital exclusion, respect for fundamental rights and equitable access to public services are all issues that need to be tackled by AI-governance. While India's DPI offers a valuable framework for harnessing the power of AI in governance, the sustainability and democratic nature of its impact in the future will hinge on the delicate equilibrium between technological advancement, ethical considerations, institutional accountability, and citizen-centric governance.

5. Ethical, Legal and Democratic Challenges

The integration of Artificial Intelligence (AI) in the digital governance ecosystem in India offers great potential but also poses ethical, legal and democratic issues with regards to the legitimacy of governance, accountability of institutions and citizens' rights. Issues such as algorithmic bias, data privacy, transparency, accountability, cybersecurity, digital exclusion, and the compatibility of AI-facilitated governance with constitutional values are key concerns.

5.1 Algorithmic Bias

Bias can be propagated or exacerbated by AI systems, as they are based on historical data. Data that is unrepresentative, poor design, or institutional inequalities can create bias (Barocas & Selbst, 2016). These risks can impact access to welfare benefits, health care, recruitment, policing, and resource distribution in India, including disparities by socioeconomic, geographic, linguistic, gender, caste, and disability status.

Algorithmic bias is thus not only a technical issue. It can be a symptom of wider social and institutional environments, necessitating representative datasets, human oversight, ongoing auditing and processes for citizens to challenge automated decisions (Floridi & Cowls, 2019).

5.2 Data Privacy

With a lot of data, there are privacy, consent, security, surveillance and function creep concerns with AI-enabled governance. The digital platforms in India handle substantial personal data on identity, healthcare, education, welfare, and financial services. In the case of Justice K. S. Puttaswamy (Retd.) v. Union of India, 2017, privacy was recognized as a fundamental right, and coupled with the Digital Personal Data Protection Act, 2023, it offers a valuable legal framework. But the protection also

needs to be privacy by design, data minimisation, purpose limitation, secure storage, and transparent data sharing.

5.3 Transparency and Explainability

AI systems are “black boxes” that are hard to understand, as automated recommendations and decisions are made. There is, therefore, a need for explainability in the context of procedural fairness, judicial review, accountability and public trust (OECD, 2019). Transparency should not be limited to technical documentation, but should also include the disclosure of AI's goals, constraints, data sources, and governance mechanisms, allowing for proper democratic oversight.

5.4 Institutional Accountability

The use of AI introduces complexities into traditional administrative duties, as decisions can affect government bodies, technology companies, developers, data sources, and automated systems. The use of AI should therefore complement, and not supplant, administrative accountability (Wirtz et al., 2019). The human officials should continue to be responsible for decisions that impact citizens' rights and access to services. The responsibilities of institutions must be clearly defined, independent oversight must be in place, algorithmic impact assessments must be accessible, and grievance procedures must be available.

5.5 Cybersecurity Risks

Relying on the interconnected digital infrastructure makes one more vulnerable to cyberattacks, ransomware, data breaches, and manipulation of AI systems. While AI can bolster fraud prevention and network security, it can also be used to create more advanced fraud schemes and misinformation campaigns. Cyber infrastructure security, regular audits, incident response plans, and institutional cyber resilience are thus part of responsible AI governance.

5.6 Digital Divide and Inclusion

Vulnerable groups may not reap the benefits of AI-driven governance due to unequal access to the internet, digital literacy, infrastructure, language accessibility, and socioeconomic factors. These differences can be mitigated through inclusive algorithms, multilingual interfaces, accessible

platforms, and capacity building initiatives that are targeted. Digital inclusion is thus not just a technical goal, but a democratic and social-justice goal.

5.7 Constitutional and Democratic Concerns.

The governance system under AI should be transparent, accountable, participatory, respect fundamental rights, and adhere to the rule of law. In India, these requirements are linked with constitutional obligations to equality, liberty, dignity and social justice. The question now isn't so much whether AI should be part of governance, but how to govern it so it doesn't take away democratic accountability and still allow for innovation. This needs to be supported by legal protection, ethical guidelines, institutional control, and active citizen involvement.

6. Policy Analysis and Global Comparisons

6.1 India's AI Policy Landscape

India has adopted an incremental, policy-driven approach to AI governance, combining innovation initiatives with emerging ethical and legal safeguards. NITI Aayog's *National Strategy for Artificial Intelligence* (2018) introduced the vision of “AI for All,” identifying sectors such as healthcare, agriculture, and education as priorities (NITI Aayog, 2018). The IndiaAI Mission further supports AI infrastructure, research, datasets, and public-sector adoption.

The Digital Personal Data Protection Act, 2023 provides an important legal foundation for addressing privacy concerns associated with AI-enabled governance. However, the absence of a dedicated risk-based AI law and fragmented institutional responsibilities highlight the need for stronger regulatory coordination and accountability.

6.2 Comparative International Experience

International frameworks demonstrate varied approaches to responsible AI governance. The OECD principles emphasise human-centred values, transparency, robustness, and accountability, while UNESCO's framework adopts a broader human-rights and social-inclusion approach (OECD, 2019; UNESCO, 2021). The European Union's risk-based AI Act provides stronger regulatory obligations for high-risk systems, whereas Singapore and the United Kingdom illustrate more flexible, sector-

oriented approaches. These models demonstrate that effective AI governance can take different institutional forms while maintaining common commitments to accountability, transparency, and public trust.

6.3 Lessons for India

For India, international experience suggests three evidence-based priorities: operationalising ethical principles through mechanisms such as algorithmic impact assessments and independent audits; strengthening coordination among government, regulators, academia, industry, and civil society; and ensuring that AI governance protects equality, privacy, democratic participation, and social justice. India's framework should therefore balance technological innovation with constitutional safeguards, institutional accountability, and inclusive development.

7. Policy Recommendations and Future Directions

7.1 Building Responsible AI Governance

How can we create responsible AI governance? How to establish responsible AI governance?

The Responsible AI approach should be adopted in India, which involves integrating fairness, transparency, explainability, accountability, privacy and human oversight into the AI lifecycle (OECD, 2019; UNESCO, 2021). Algorithmic impact assessment should be conducted prior to high-impact deployments in the public sector, and independent audits and regular evaluation for risks of discrimination, privacy, cybersecurity and human rights.

7.2 Institutional Reforms

There is a need for better coordination between government departments, regulators, academia, industry and civil society in order to ensure effective governance of AI. Capacity building should thus provide civil servants, policymakers and regulators with the knowledge of AI, data governance, digital ethics and algorithmic accountability. This institutional capacity is essential for both successful implementation of AI and for the necessary monitoring.

7.3 Legal and Regulatory Strengthening

The legal and regulatory framework is strengthened. India needs to create a principles-based and adaptive regulatory structure with enforceable protections. The requirements for transparency, documented decision making, human review and grievance redress should apply to public sector AI systems that have a substantial impact on citizens' rights and access to services. These protections can enhance **equality and fairness in procedures, and accountability of administration.**

7.4 Citizen-Centric Governance

AI governance should not just be about citizens as users of AI-enabled services, but also as stakeholders. Public consultation, digital literacy, data rights awareness, and easy access to information about algorithmic decision-making can enhance public trust. In India's diverse social and linguistic landscape, addressing digital disparities by enhancing connectivity and ensuring the adoption of user-centred, inclusive AI systems is especially crucial.

7.5 Future Research Directions

Further studies are needed to empirically evaluate the impact of AI in the fields of healthcare, education, agriculture, law enforcement, and public administration in the long term. There is also a need for a greater focus on the constitutional and public-policy aspects of generative AI, explainable AI and automated decision-support systems. Research should explore the impacts of AI on institutional trust, democratic accountability, citizen engagement and social inclusion, to help develop governance models that are technologically innovative while also providing democratic and ethical protections.

8. Conclusion

The integration of Artificial Intelligence (AI) into India's digital governance ecosystem represents a significant transformation in public administration and democratic governance. As demonstrated in this chapter, AI can enhance administrative efficiency, evidence-based policymaking, public service delivery, and responsiveness across sectors including healthcare, education, agriculture, law enforcement, and public administration. However, its effectiveness cannot be assessed solely through technological or operational outcomes. AI-enabled governance requires robust legal, institutional, and ethical safeguards that protect citizens' rights and democratic values.

The chapter demonstrates that algorithmic bias, privacy, transparency, explainability, accountability, cybersecurity, and digital inclusion are fundamental governance concerns affecting institutional legitimacy and public trust. By integrating perspectives from political science, public administration, digital governance, AI ethics, and public policy, the chapter contributes an interdisciplinary understanding of AI governance in India.

The chapter concludes that India's future AI governance should emphasise Responsible AI, ethical design, transparent decision-making, human oversight, institutional accountability, and citizen participation. Such an approach can help India harness AI's transformative potential while strengthening constitutional values, democratic governance, and inclusive digital development.

References:

- Barocas, S., & Selbst, A. D. (2016). *Big data's disparate impact*. *California Law Review, 104*(3), 671–732.
- Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., Morison, J., Murphy, M. H., O'Brolcháin, F., Schafer, B., & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society, 4*(2).
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). *Digital era governance: IT corporations, the state, and e-government*. Oxford University Press.
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).
- Government of India. (2023). *Digital Personal Data Protection Act, 2023*.
- Justice K. S. Puttaswamy (Retd.) & Anr. v. Union of India & Ors., (2017) 10 SCC 1.
- Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature, 568*(7751), 163–165.
- Ministry of Electronics and Information Technology. (2023). *Annual report 2022–2023*. Government of India.
- NITI Aayog. (2018). *National strategy for artificial intelligence #AIForAll*. Government of India.
- OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD Publishing.

- Osborne, S. P. (Ed.). (2010). *The new public governance? Emerging perspectives on the theory and practice of public governance*. Routledge.
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368–383.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
- United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations.
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615.
- World Bank. (2017). *World development report 2017: Governance and the law*. World Bank.
- World Bank. (2021). *World development report 2021: Data for better lives*. World Bank.

Chapter 3

Artificial Intelligence in Journalism: Transforming News Production and Audience Engagement

Mehak Arora

Assistant Professor, BBK DAV College for Women, Amritsar

Abstract

Artificial Intelligence (AI) is transforming the field of journalism by reshaping news production, distribution, and audience engagement. AI-powered technologies such as machine learning, natural language processing, automated content generation, predictive analytics, and recommendation algorithms are enabling media organizations to produce news more efficiently while delivering personalised content to diverse audiences. Newsrooms increasingly rely on AI for tasks including news gathering, fact-checking, data analysis, transcription, content moderation, and audience analytics, allowing journalists to focus on investigative reporting and storytelling. Despite these advancements, the integration of AI raises significant ethical and professional concerns. Issues such as algorithmic bias, misinformation, deepfakes, transparency, accountability, data privacy, and the potential displacement of media professionals challenge the credibility and integrity of journalism. This chapter examines the transformative role of AI in modern journalism by analysing its impact on editorial practices, newsroom workflows, audience behaviour, and public trust. It further explores how media organisations can balance technological innovation with ethical responsibility through human oversight, media literacy, and responsible AI governance. By presenting contemporary examples and emerging trends, the chapter highlights both the opportunities and challenges of AI-driven journalism. It concludes by proposing strategies for integrating AI responsibly to strengthen journalistic values, enhance audience engagement, and ensure the sustainability of credible journalism in the rapidly evolving digital media landscape.

Keywords: Artificial Intelligence, Journalism, Digital Media, News Production, Audience Engagement

1. Introduction

Journalism has always been closely connected with technological change. From the printing press

and telegraph to radio, television, satellite communication, the internet, smartphones, and social media, every major technological development has influenced how news is gathered, produced, distributed, and consumed. Artificial Intelligence represents another significant stage in this continuing transformation. Unlike earlier technologies that primarily improved the speed or reach of communication, AI is increasingly participating in activities that were traditionally associated with human cognitive work, including language processing, information analysis, content generation, recommendation, and pattern recognition.

The use of Artificial Intelligence in journalism is not entirely new. News organisations have used automated systems and algorithms for years to process large datasets and generate routine reports. The Associated Press, for example, began using automation for corporate earnings reports in 2014. By 2015, it reported that automated systems were producing more than 3,000 earnings stories each quarter, substantially increasing the volume of coverage while freeing journalists from repetitive data-processing tasks (Associated Press, 2015).

What has changed significantly in recent years is the emergence of generative AI. Generative AI systems can produce text, images, audio, video, summaries, translations, headlines, and other forms of content. This has expanded the role of AI from automated data processing to more sophisticated forms of editorial assistance. Journalists can now use AI tools to transcribe interviews, summarise documents, analyse large datasets, translate material, generate preliminary headlines, identify patterns, and assist with research.

The development presents an important opportunity for news organisations. Journalism operates under constant pressure to produce accurate information quickly while managing limited financial and human resources. AI can help reduce routine workloads and allow journalists to spend more time on activities that require human judgment, such as source development, field reporting, investigative journalism, interpretation, and storytelling.

At the same time, AI introduces serious challenges. Journalism is fundamentally based on credibility, verification, independence, accountability, and public trust. If AI-generated material contains factual errors, fabricated sources, biased interpretations, or misleading information, the consequences can be significant. Generative AI systems may produce convincing but incorrect information, commonly referred to as hallucinations. Algorithms can reproduce biases contained

within their training data. Personalisation systems can influence which stories audiences see, while automated recommendation systems may prioritise engagement over public interest. The transformation is therefore not simply about replacing human journalists with machines. It is more accurately understood as a transformation of the relationship between journalists, technology, news organisations, and audiences. The central question is not whether journalism will use AI, because AI adoption is already occurring. The more important question is how journalism can use AI without compromising its professional and democratic responsibilities.

2. Understanding Artificial Intelligence in Journalism

Artificial Intelligence refers broadly to computational systems capable of performing tasks that normally require aspects of human intelligence. These tasks may include learning from data, recognising patterns, understanding language, making predictions, and generating content.

Several branches of AI are particularly relevant to journalism are as follows:

2.1 Machine Learning

Machine learning allows systems to identify patterns within large datasets and make predictions or classifications. In journalism, machine learning can assist with analysing audience behaviour, detecting unusual patterns in data, identifying potential news events, and organising large collections of documents.

2.2 Natural Language Processing

Natural Language Processing, or NLP, enables computers to process human language. News organisations can use NLP for transcription, translation, sentiment analysis, document classification, keyword extraction, summarisation, and automated text generation.

2.3 Generative AI

Generative AI refers to systems capable of producing new content based on patterns learned from existing data. In journalism, these systems can help generate drafts, summaries, headlines, captions, scripts, and other forms of editorial material.

2.4 Computer Vision

Computer vision enables machines to analyse images and videos. News organisations can use computer vision for image classification, object identification, video indexing, archive management, verification, and automated tagging.

2.5 Recommendation Algorithms

Recommendation systems analyse user behaviour and preferences to determine which stories or content should be displayed. They play an increasingly important role in digital journalism because audiences often encounter news through platforms, search engines, social networks, and personalised feeds rather than directly visiting a newspaper homepage.

These technologies are interconnected. A modern digital newsroom may use several AI systems simultaneously, with one tool processing transcripts, another analysing data, another generating a summary, and another recommending the finished story to audiences.

This development has led to what Diakopoulos (2019) describes as the increasing importance of computational systems in journalistic decision-making. Algorithms are no longer simply technical tools operating behind the newsroom. They increasingly influence what information is identified as news, how it is presented, and which audiences receive it.

3. AI and the Transformation of News Production

One of the most visible effects of AI has been the transformation of the newsroom workflow.

Traditional news production generally involves several stages: identifying a story, gathering information, interviewing sources, verifying facts, writing the report, editing the material, producing multimedia elements, publishing, and distributing the story.

AI can now assist at almost every stage.

During news gathering, AI systems can scan large volumes of information, including government

documents, press releases, social media posts, financial reports, and public records. This can help journalists identify potential stories that might otherwise remain hidden within large datasets.

During research, AI can summarise lengthy reports and organise information. However, summaries must be treated as preliminary research rather than authoritative representations of the original documents. Journalists remain responsible for reading important primary sources and checking the accuracy of AI-generated summaries.

During transcription, speech recognition tools can convert interviews, press conferences, and meetings into text within minutes. This can save considerable time, particularly for television, radio, and digital journalists working under tight deadlines.

During translation, AI can help newsrooms produce content in multiple languages. This can be particularly useful in multilingual societies such as India.

During writing, AI can assist with headlines, summaries, basic structured reports, and alternative versions of stories. However, the journalistic value of a story still depends on reporting, context, evidence, and editorial judgment.

The Associated Press has demonstrated how automation can be used to increase the volume of routine journalism. Its automated earnings project showed that technology could handle repetitive data-based reports while allowing journalists to focus more on enterprise reporting and stories requiring interpretation (Associated Press, 2015).This model illustrates an important principle: AI can be most useful when it handles repetitive tasks rather than replacing the core responsibilities of journalism.

4. AI-Assisted News Gathering and Research

Journalists increasingly work with enormous quantities of information. Government reports, court documents, corporate filings, academic papers, social media posts, datasets, video recordings, and public databases can contain valuable information, but manually examining all of these sources is time-consuming.

AI can help journalists process this information more efficiently.

For example, a journalist investigating government expenditure could use AI-assisted systems to identify unusual changes in spending patterns. A business journalist could analyse thousands of financial documents. A sports journalist could use structured data to generate basic match statistics. An environmental journalist could examine large datasets relating to temperature, rainfall, pollution, or land use. AI can also help identify connections across documents. This is particularly relevant for investigative journalism, where important information may be distributed across thousands of pages.

However, AI-assisted research creates a new responsibility. Journalists must distinguish between information discovery and information verification.AI can identify a potentially relevant document, but the journalist must still examine the original document. AI can suggest that two events are connected, but the journalist must independently establish the relationship. AI can identify a possible source, but the journalist must assess that source's credibility. This distinction is crucial because generative AI systems are designed to produce plausible outputs, not necessarily verified facts. A response can sound confident and professional while still being incorrect.Therefore, AI should be considered a research assistant rather than an autonomous research authority.

5. Automated Journalism and the Changing Newsroom

Automated journalism refers to the use of computer systems to produce news content with limited direct human writing. It is particularly effective when information is available in a structured format.Examples include sports scores, weather updates, financial results, election statistics, and routine business reports. Carlson (2015) argues that automated journalism raises questions about the changing boundaries between human and machine labour in news production. When a computer transforms structured data into a readable news report, the traditional assumption that journalism requires a human writer becomes more complicated.

However, automated journalism works best in specific circumstances. Structured information can be converted into predictable narratives relatively easily. Complex political developments, investigative stories, human-interest reporting, and sensitive social issues require much greater contextual understanding. This is why the future newsroom is unlikely to consist entirely of automated systems. Instead, it is more likely to involve a combination of journalists and AI tools. The Associated Press's experience provides an example. Its automation projects have focused on repetitive tasks, allowing journalists to spend more time on reporting and higher-value journalism (Associated Press,

2022).This suggests that AI may not necessarily eliminate journalism jobs in the short term. Instead, it may change the nature of journalistic work. The journalist of the future may spend less time typing routine reports and more time verifying information, interpreting data, developing sources, producing multimedia stories, investigating complex issues, and supervising automated systems.

6. AI and Personalised News Distribution

The digital environment has changed how audiences receive news. In the traditional newspaper model, editors decided the order and prominence of stories. In digital environments, algorithms increasingly determine which stories appear on an individual's screen. Personalization can improve audience experience. A reader interested in sports can receive more sports coverage. Another reader interested in international affairs can receive more international stories. AI can analyse reading patterns and recommend content based on previous behaviour. Reuters Institute research shows that publishers increasingly view AI-powered personalisation and recommendation as important areas of newsroom development. Its 2025 industry survey found that 80% of surveyed publishers considered personalisation and recommendation important for the year ahead, while 96% identified back-end efficiency as important (Newman, 2025).

Personalisation can therefore help news organisations connect audiences with relevant content.However, it also creates concerns.If algorithms continually recommend stories similar to what a person has already consumed, audiences may receive a narrow range of perspectives. This can contribute to filter bubbles and ideological fragmentation.There is also the question of whether public-interest journalism should be governed by audience preferences alone. Important stories are not always the stories that audiences actively choose. Investigative reports, public health information, environmental issues, elections, and policy developments may be socially important even if they do not generate the highest engagement.

7. AI and Audience Engagement

Audience engagement has become a central concern for digital journalism. News organisations compete not only with other news organisations but also with entertainment platforms, influencers, creators, podcasts, social networks, and messaging applications.

News organisations can use AI-powered chatbots to allow users to ask questions about a news topic. AI can summarise long articles, translate stories, generate audio versions, create personalised newsletters, and help users navigate large archives. The Reuters Institute's 2025 Digital News Report found that audiences showed interest in several AI-supported news features, including summaries, translations, recommendations, and chatbots. However, audiences remained more comfortable when humans remained involved in the production process (Newman, 2025). The 2025 Generative AI and News report further found that 19% of respondents had seen AI-generated bullet-point summaries and 16% had encountered AI chatbots while using news websites or apps.

These features can improve accessibility

A reader with limited time may prefer a short summary. A person with a visual impairment may prefer an audio version. A student may want a simplified explanation of a complex policy story. A multilingual audience may benefit from translation. However, audience engagement should not be reduced to clicks, likes, or time spent on a page. Journalism has a social responsibility that extends beyond engagement metrics.An AI system designed only to maximise engagement could favour sensationalism, outrage, or emotionally provocative content. Such a system might increase traffic while weakening the quality of public discourse.Therefore, responsible AI-driven audience engagement must balance user convenience with journalistic values.

8. AI, News Trust, and Credibility

Trust is one of the most important issues in contemporary journalism. Audiences expect journalists to provide accurate, verified, and independent information. AI creates both opportunities and risks for this relationship.

On one hand, AI can support accuracy by identifying inconsistencies, checking large datasets, detecting duplicated content, and helping journalists locate relevant sources. On the other hand, AI can generate false information, reproduce bias, and create convincing fabricated content. Recent audience research shows that people remain cautious about AI-generated news. The Reuters Institute's 2025 report found that only 12% of respondents were comfortable with news produced entirely by AI, compared with 62% for news produced entirely by human journalists. Comfort increased when humans remained involved in the process (Simon et al., 2025). This finding has

important implications. It suggests that audiences do not necessarily reject AI. Rather, they are uncomfortable with removing humans from editorial responsibility.The same report found that people generally viewed AI as capable of making news cheaper and more up-to-date, but they also expected it to make news less transparent and less trustworthy.

This indicates that news organisations need to communicate clearly about how they use AI.Transparency can help establish a social contract between news organisations and audiences.

9. Human Oversight and Editorial Judgment

One of the strongest arguments for responsible AI journalism is the principle of human oversight. A journalist must decide whether a source is credible, whether an allegation deserves publication, whether information is in the public interest, whether a photograph is ethically appropriate, and whether a particular detail could unnecessarily harm an individual.

These are not simply computational questions. A machine can identify that a story is receiving high engagement, but it cannot independently determine whether that story serves the public interest. An algorithm can identify a pattern in data, but a journalist must determine whether the pattern is meaningful.

The Associated Press's current newsroom standards illustrate this approach. AP states that AI can assist with specific activities such as research, summarization, transcription, translation, headlines, and grammar, but editorial judgment, verification, and accountability remain the responsibility of journalists (Associated Press, 2026). Reuters has similarly stated that its journalists use generative AI in reporting, writing, editing, production, and publishing, while disclosing cases where AI plays a primary or substantial role in creating content (Reuters, 2024). These approaches suggest a model in which AI assists journalism but does not become the final editorial authority.

10. Algorithmic Bias and Journalism

Bias is one of the most complex challenges associated with AI. AI systems learn from data. If the training data contains social, cultural, political, gender, racial, linguistic, or economic biases, algorithms can reproduce or amplify those patterns.In journalism, algorithmic bias can influence which stories are recommended, which topics are classified as important, which sources are

prioritised, and how audiences are segmented.

For example, a recommendation system might learn that sensational political content receives high engagement. If the algorithm optimises primarily for engagement, it may continue recommending similar material. Bias can also appear in automated language systems. Some languages and dialects may receive less accurate transcription or translation than widely represented languages. This issue is particularly relevant in India because of its linguistic diversity. AI systems designed primarily around English or other globally dominant languages may perform differently when handling Indian languages, regional dialects, or mixed-language communication. Journalists and news organisations should therefore test AI systems across different demographic and linguistic groups. They should also maintain mechanisms through which users and journalists can report algorithmic errors. Bias cannot always be completely eliminated, but it can be identified, monitored, disclosed, and reduced.

11. Misinformation, Deepfakes, and Synthetic Media

Generative AI has made it easier to create synthetic content. Images, audio, and videos can now be generated or manipulated at increasingly realistic levels. For journalism, this creates two major problems. First, AI can be used to produce misinformation and disinformation. Second, genuine journalistic material can be falsely labelled as AI-generated. The second problem is sometimes described as the "liar's dividend." As audiences become aware that synthetic media exists, individuals accused of wrongdoing may claim that authentic evidence is fake. Journalists therefore need new verification skills. AI-assisted verification tools can examine images, videos, audio files, metadata, visual inconsistencies, and other indicators. However, detection systems are not perfect. Journalists should combine technological tools with traditional verification techniques such as source confirmation, reverse image searching, location verification, comparison with original footage, and contacting people directly involved in an event. The development of AI therefore increases rather than reduces the importance of journalism's traditional verification function.

12. Privacy and Data Protection

AI journalism also raises questions about privacy. News organisations collect large amounts of audience data. This can include reading history, search behaviour, location information, device information, subscription data, and interaction patterns. AI systems can analyse this data to

personalise content. However, excessive data collection can create privacy risks. There is a fundamental ethical question here: how much should a news organisation know about its audience? Personalisation can be useful, but users should understand what information is collected and how it is being used. Data should not be collected simply because technology makes it possible.

Privacy concerns also arise when journalists use AI tools to process confidential documents or interview transcripts. Uploading sensitive information to an external AI system could create risks if appropriate safeguards are not in place. News organisations therefore need clear internal policies covering confidential sources, unpublished material, personal data, and sensitive journalistic documents.

13. Impact of AI on Journalistic Employment

The possibility of job displacement is one of the most discussed concerns surrounding AI. Some routine journalistic tasks can already be automated. This naturally creates anxiety among journalists whose work involves repetitive reporting, transcription, translation, data processing, or basic content production. However, technological change does not necessarily mean that journalism itself will disappear.

Historically, journalism has adapted to technological developments. Desktop publishing changed newspaper production. Digital cameras changed photojournalism. Content management systems changed newsroom workflows. Social media changed distribution. AI is likely to create a similar transformation.

Some roles may decline, while new roles emerge. These may include AI editors, data journalists, verification specialists, audience analysts, prompt specialists, computational reporters, AI policy managers, and digital forensic journalists. The journalist's professional identity may therefore shift from being primarily a content producer to being a content producer, investigator, verifier, editor, and technology supervisor.

This makes reskilling essential. Journalism education should introduce students to AI tools while continuing to teach traditional reporting. Technical knowledge without journalistic ethics is insufficient. Similarly, traditional journalism skills without digital literacy may leave graduates unprepared for modern newsrooms.

14. AI and Journalism Education

Journalism schools and universities have a major role in preparing future professionals. Students should understand how AI systems work at a basic conceptual level. They do not necessarily need to become programmers, but they should understand concepts such as machine learning, generative AI, algorithmic bias, hallucination, data privacy, synthetic media, and content provenance.

Practical training should include:

1. AI-assisted research
2. AI-supported transcription
3. Data analysis
4. Fact-checking
5. Synthetic media verification
6. AI-assisted translation
7. Automated content generation
8. Prompt design
9. Algorithmic accountability
10. AI ethics

At the same time, students must continue developing traditional skills such as interviewing, source verification, news writing, editing, photography, video production, investigative journalism, media law, and communication ethics.

15. The Future Relationship Between Journalists and AI

The future of journalism is unlikely to be completely human or completely automated. It is more likely to be collaborative. AI is particularly strong at speed, scale, pattern recognition, repetitive processing, and information organisation. Human journalists remain stronger in areas requiring empathy, contextual understanding, ethical reasoning, source relationships, creativity, skepticism, and public accountability.

A journalist could use AI to transcribe a two-hour interview, identify key themes, compare statements with previous public records, translate selected sections, and suggest possible follow-up

questions. The journalist would then verify the information, conduct additional interviews, decide what is newsworthy, and produce the final story. Such a workflow can make journalism more efficient without removing the human element. The future journalist may therefore be less concerned with competing against AI and more concerned with learning how to work effectively with it.

16. Conclusion

Artificial Intelligence is changing journalism at a speed that few previous technologies have matched. It is influencing how journalists gather information, analyse data, write stories, edit content, distribute news, and interact with audiences. From automated earnings reports and transcription to personalised recommendations and conversational news interfaces, AI has moved from being an experimental technology to becoming an increasingly important part of the modern newsroom.

The benefits are significant. AI can reduce repetitive work, improve efficiency, support multilingual communication, process large datasets, increase accessibility, and help journalists produce more content for increasingly diverse audiences. It can allow journalists to spend more time on investigative reporting, original storytelling, and complex public-interest issues. However, technological efficiency cannot be treated as a substitute for journalistic responsibility.

The major challenges include misinformation, algorithmic bias, hallucinations, privacy concerns, deepfakes, lack of transparency, audience manipulation, and possible employment disruption. These challenges become particularly serious because journalism is not simply a commercial activity. It performs an important democratic function by providing citizens with information required to understand society and participate in public life. Current audience research reinforces the importance of human involvement. People are generally more comfortable when AI supports journalists than when AI replaces them. The Reuters Institute's findings show a clear preference for human-led journalism and continued concerns about transparency and trust when AI is used extensively (Simon et al., 2025). The appropriate response is therefore neither complete rejection of AI nor unquestioning technological enthusiasm. Journalism needs a balanced approach based on human oversight, transparency, verification, ethical governance, and continuous professional education. AI should handle what machines are good at: processing large amounts of information, identifying patterns, automating repetitive tasks, assisting with transcription, translation, and personalisation,

and supporting newsroom efficiency. Journalists should continue doing what humans are uniquely positioned to do: asking difficult questions, building relationships with sources, understanding social context, exercising ethical judgment, investigating wrongdoing, interpreting complex events, and holding powerful institutions accountable.

References

- Associated Press. (2015). *A leap forward in quarterly earnings stories*. The Associated Press. (The Associated Press)
- Associated Press. (2015, June 15). *Automated earnings stories multiply*. The Associated Press. (The Associated Press)
- Associated Press. (2022, January 5). *Helping local newsrooms use AI*. The Associated Press. (The Associated Press)
- Associated Press. (2022, May 26). *Applying automation in the newsroom*. The Associated Press. (The Associated Press)
- Associated Press. (2023, July 13). *AP, OpenAI agree to share select news content and technology in new collaboration*. The Associated Press. (The Associated Press)
- Associated Press. (2026, July 23). *AP updates newsroom standards for artificial intelligence*. The Associated Press. (The Associated Press)
- Beckett, C. (2019). *New powers, new responsibilities: A global survey of journalism and artificial intelligence*. London School of Economics and Political Science.
- Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.
- Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism, 3*(3), 416-431. https://doi.org/10.1080/21670811.2014.976412
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
- Graefe, A. (2016). *Guide to automated journalism*. Tow Center for Digital Journalism, Columbia University.
- Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human-

machine communication: Rethinking roles and relationships of journalists and machines. *Digital Journalism*, *7*(4), 409-427.

- Newman, N. (2025). *Journalism, media, and technology trends and predictions 2025*. Reuters Institute for the Study of Journalism, University of Oxford. (reutersinstitute.politics.ox.ac.uk)
- Newman, N. (2025). *Overview and key findings of the 2025 Digital News Report*. Reuters Institute for the Study of Journalism, University of Oxford. (reutersinstitute.politics.ox.ac.uk)
- Reuters. (2024, October 30). *Reuters and AI*. Reuters. (Reuters)
- Reuters. (2026, August 4). *Reuters launches industry-leading live transcriptions on real-time video on Reuters Connect, extending AI capabilities for agency customers*. Reuters. (Reuters)
- Simon, F. M., Nielsen, R. K., & Fletcher, R. (2025). *Generative AI and news report 2025: How people think about AI's role in journalism and society*. Reuters Institute for the Study of Journalism, University of Oxford. (reutersinstitute.politics.ox.ac.uk)
- Thurman, N., Dörr, K. N., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, *5*(10), 1240-1259. https://doi.org/10.1080/21670811.2017.1289819

Chapter 4

Consumer Behaviour of Generation Z Towards Sustainable Fashion Products: An Exploratory Study in the Indian Context

Abhishek Chakraborty[1]
Assistant Professor
Jagannath University, Haryana

Abstract

The fashion industry has grown rapidly over the last few decades, but this growth has also created serious environmental and social concerns, including textile waste, excessive resource consumption, carbon emissions and concerns related to labour practices. Sustainable fashion has therefore emerged as an important approach to reducing the negative effects associated with conventional fashion consumption. Generation Z is particularly relevant to this discussion because of its strong digital presence, growing purchasing influence and exposure to environmental and social issues. However, being aware of sustainability does not necessarily mean that consumers consistently purchase sustainable products.

This study examines the consumer behaviour of Generation Z towards sustainable fashion products in the Indian context. The research focuses on awareness, environmental concern, attitudes, purchase behaviour, product quality, price, social media influence and perceived barriers. A quantitative and descriptive research design was adopted. Primary data were collected through a structured online questionnaire from 25 Generation Z respondents aged between 18 and 27 years using convenience sampling. The responses were analysed using Microsoft Excel through frequency and percentage analysis. The findings indicate that respondents demonstrate a generally positive attitude towards sustainable fashion and possess considerable awareness of its environmental and ethical dimensions. Most respondents had purchased sustainable fashion products at least once, but regular purchasing was comparatively lower. Product quality and durability emerged as important considerations, while social media played a noticeable role in awareness and purchase decisions. Although many respondents were willing to pay somewhat more for sustainable products, price and availability continued to act as barriers. The study highlights the continuing gap between positive attitudes and actual purchasing behaviour and suggests that affordability, accessibility, product quality and

transparent communication are important for encouraging more consistent sustainable fashion consumption among Generation Z in India.

Keywords: Generation Z, sustainable fashion, consumer behaviour, purchase intention, environmental awareness, social media, Indian fashion market

1. Introduction

Fashion is closely connected with everyday life, identity, culture and self-expression. At the same time, the rapid expansion of fashion production and consumption has created environmental and social challenges. The growth of fast fashion has encouraged consumers to purchase garments more frequently and to replace them quickly. This pattern contributes to increasing textile waste and places pressure on natural resources and production systems. Sustainable fashion has developed in response to these concerns and promotes more responsible approaches to the design, production, distribution and consumption of fashion products.

Sustainable fashion can be understood not simply as the use of environmentally friendly materials, but as a broader approach that considers environmental protection, ethical labour, responsible sourcing, durability and responsible consumption. Niinimäki (2010) notes that clothing has a strong relationship with consumer identity and values, making sustainable fashion more complex than a purely environmental choice. Consumers may use clothing to communicate who they are and what they value, including their ethical and environmental beliefs.

Consumer behaviour is therefore central to the transition towards sustainable fashion. Even when consumers understand the environmental consequences of fashion, their final purchasing decisions can be shaped by practical considerations such as price, quality, style, convenience and availability. Joergens (2006) identified this tension in ethical fashion consumption, showing that consumers may express concern about ethical issues while still prioritising personal needs when making actual purchases. Similarly, McNeill and Moore (2015) found that consumers may experience a conflict between their sustainability values and their desire for fashionable, convenient and affordable clothing.

Generation Z has become an important consumer group in this context. Generally associated with individuals born from the late 1990s to the early 2010s, this generation has grown up with digital

technologies and social media. They are regularly exposed to information about climate change, ethical consumption, environmental responsibility and social issues. Their digital behaviour also makes them an important audience for fashion brands and influencers.

However, awareness alone may not be sufficient to create sustained purchasing behaviour. This study therefore examines how Generation Z consumers in India perceive sustainable fashion and how awareness, environmental concern, product attributes, price and social media influence their purchasing behaviour. The study also considers the barriers that prevent consumers from purchasing sustainable fashion products more frequently.

2. Review of Literature

2.1 Sustainable Fashion and Consumer Behaviour

Sustainable fashion consumption involves more than choosing an environmentally friendly garment. It includes the way consumers select, purchase, use, maintain and eventually dispose of clothing. Niinimäki (2010) explains that eco-fashion is connected with consumer identity, personal values and ethical commitment. This suggests that sustainable fashion choices can be influenced by both environmental beliefs and the symbolic meaning of clothing.

Joergens (2006) found that consumers often support ethical fashion in principle but may hesitate when ethical products require them to compromise on price, style or convenience. This observation remains important because fashion is a highly personal category in which consumers frequently balance ethical values against aesthetic and functional expectations.

McNeill and Moore (2015) similarly found that consumers may have different motivations for engaging with sustainable fashion. Some may be primarily concerned with personal benefits, while others may be influenced by social or environmental considerations. Sustainable fashion consumption is therefore not driven by a single motivation.

2.2 Environmental Awareness

Environmental awareness is frequently identified as an important factor in sustainable consumption. Consumers who understand the environmental consequences of fashion production may be more

inclined to consider sustainable alternatives. Han, Seo, and Ko (2017) examined sustainable fashion product consumption and highlighted the continuing gap between consumers' sustainability concerns and actual purchasing behaviour.

Environmental awareness is particularly relevant to Generation Z because digital media has made environmental information widely accessible. Young consumers can easily encounter content relating to textile waste, pollution, ethical production and climate change. However, knowledge must be accompanied by accessible and attractive products if it is to influence purchasing behaviour consistently.

2.3 Generation Z and Sustainable Consumption

Generation Z represents a significant consumer segment because of its digital connectivity and increasing market influence. The generation is frequently associated with authenticity, social responsibility and greater awareness of social issues. At the same time, Generation Z is highly exposed to fashion trends through social media.

This creates an interesting contradiction. Social media can educate young consumers about sustainability while simultaneously encouraging rapid consumption through influencer culture, trend cycles and digital advertising. Consequently, social media can both support and challenge sustainable fashion consumption.

2.4 Product Quality, Price and Style

Sustainability is unlikely to become a strong purchasing factor if the product does not satisfy basic consumer expectations. Quality, durability, comfort and style remain important considerations. McNeill and Moore (2015) demonstrate that consumers may support sustainable fashion but continue to be influenced by fashionability and personal preferences.

Price is another major consideration. Sustainable products may be perceived as expensive because of the cost of responsible materials, ethical production and smaller production volumes. Joergens (2006) observed that consumers may hesitate to purchase ethical clothing when it requires a significant financial sacrifice.

This suggests that sustainable fashion brands need to move beyond simply communicating environmental benefits. Consumers must also perceive the product as offering good value, attractive design and durability.

2.5 Social Media and Influencer Influence

Social media has become an important part of the fashion decision-making process. Platforms such as Instagram, YouTube and other digital channels provide consumers with product information, reviews, styling ideas and recommendations. Influencers can also shape perceptions of fashion brands and products.

Vrontis et al. (2021) describe social media influencer marketing as an increasingly important part of the consumer decision journey. For sustainable fashion, influencers can potentially make environmental concepts easier to understand by connecting sustainability with everyday fashion choices. However, the influence of social media is not always positive. Rapidly changing trends can encourage frequent purchases and short product life cycles. Sustainable fashion communication must therefore compete with the culture of constant novelty promoted by digital platforms.

2.6 Theoretical Perspective

The Theory of Planned Behaviour provides a useful framework for understanding sustainable purchasing decisions. Ajzen (1991) explains that behavioural intentions are influenced by attitudes, subjective norms and perceived behavioural control. In the context of sustainable fashion, a consumer may have a positive attitude towards environmentally responsible clothing, may feel that friends and social groups support sustainable consumption, and may intend to purchase sustainable products. However, the actual purchase can still be prevented if the product is too expensive, unavailable or difficult to access. This theoretical perspective helps explain why a positive attitude does not automatically result in regular purchasing behaviour. The consumer may support sustainability but lack the practical ability or opportunity to act on that belief.

3. Research Gap

Existing literature indicates that environmental awareness, ethical values, product quality, social influence and sustainability perceptions influence consumer attitudes towards sustainable fashion.

However, an attitude-behaviour gap remains evident. Consumers may express strong support for sustainability while continuing to purchase conventional fashion because of price, availability, style, convenience or established shopping habits. A further gap exists in understanding this issue within the Indian Generation Z market. Much of the established literature has examined sustainable consumption in international contexts, while consumer behaviour in India is shaped by its own economic, cultural and digital environment. The present study therefore focuses specifically on Generation Z consumers in India and examines how awareness and positive attitudes relate to actual sustainable fashion purchasing behaviour.

4. Research Objectives

The study aims to examine the consumer behaviour of Generation Z towards sustainable fashion products.

The specific objectives are:

1. To assess the awareness of Generation Z consumers regarding sustainable fashion.
2. To examine their attitudes towards sustainable fashion products.
3. To identify factors influencing sustainable fashion purchase decisions.
4. To understand the influence of environmental concern on purchasing behaviour.
5. To examine the influence of price, quality, style and availability.
6. To assess the role of social media in sustainable fashion decisions.
7. To examine the relationship between sustainability awareness and actual purchasing behaviour.
8. To identify barriers to regular sustainable fashion consumption.
9. To provide recommendations for fashion brands and marketers.

5. Research Methodology

The study adopted a quantitative, descriptive and cross-sectional research design. The research was designed to understand the existing perceptions and purchasing behaviour of Generation Z rather than manipulate any variables. A quantitative approach was considered appropriate because the study required measurable responses that could be presented through frequencies and percentages.

The target population consisted of Generation Z consumers in India. For the purpose of the study,

respondents between 18 and 27 years of age were considered. A convenience sampling technique was used because respondents were selected based on accessibility and willingness to participate.

Primary data were collected from 25 respondents using a structured questionnaire. The questionnaire included questions related to demographic characteristics, awareness of sustainable fashion, environmental concern, attitudes, purchasing behaviour, product quality, price, social media and willingness to pay more for sustainable products.

The questionnaire was distributed online through Google Forms. The survey link was shared through platforms including WhatsApp, Instagram, Facebook, LinkedIn, email and student networks. Participation was voluntary, and respondents were informed about the academic purpose of the study.

The collected responses were analysed using Microsoft Excel. Frequency and percentage analysis were used to identify general patterns in the responses. Since the sample consisted of only 25 respondents and was selected through convenience sampling, the findings should be viewed as exploratory rather than representative of the entire Indian Generation Z population.

6. Findings and Discussion

6.1 Awareness of Sustainable Fashion

The findings show that most respondents were familiar with the concept of sustainable fashion. Only a small number reported limited awareness. This suggests that sustainability has become a relatively familiar concept among the young consumers represented in the sample.

The finding is consistent with the broader literature, which identifies younger consumers as increasingly exposed to environmental information and sustainability-related communication. However, awareness should not be interpreted as proof of regular sustainable purchasing.

6.2 Actual Purchasing Behaviour

Most respondents reported purchasing sustainable fashion products at least once. However, regular purchasing was considerably lower. This is one of the most important findings of the study because it clearly demonstrates the difference between awareness and behaviour.

The finding supports the attitude-behaviour gap discussed by Joergens (2006), Niinimäki (2010) and McNeill and Moore (2015). Consumers may genuinely care about sustainability but still make conventional purchases when faced with practical considerations. In everyday shopping, consumers are not necessarily choosing between sustainability and non-sustainability in isolation. They are also considering whether a product looks good, fits their budget, is easily available and offers sufficient quality. Sustainable consumption therefore requires products that fit naturally into consumers' everyday lives.

6.3 Environmental Concern

Respondents generally demonstrated a high level of awareness regarding the environmental impact of the fashion industry. Most respondents selected positive responses when asked whether they were aware of fashion's environmental effects.

This indicates that environmental concern is present among the respondents and may contribute to positive attitudes towards sustainable fashion. The finding is consistent with the Theory of Planned Behaviour, where positive attitudes can contribute to purchase intention (Ajzen, 1991). However, environmental concern alone does not guarantee behaviour. The lower frequency of regular sustainable purchases suggests that other factors continue to intervene between concern and action.

6.4 Importance of Sustainable Fashion

A majority of respondents considered sustainable fashion important. They also expressed positive attitudes towards brands that adopt ethical and environmentally responsible practices. This finding suggests that sustainability has moved beyond being a niche concern. For young consumers, the values associated with a fashion brand can contribute to its attractiveness. This supports the argument that sustainable fashion can become part of a consumer's identity and value system (Niinimäki, 2010).

6.5 Product Quality and Durability

Quality and durability received consistently positive responses. Respondents appeared to value products that last longer and provide good value for money. This finding is particularly significant because it provides a practical direction for fashion brands. Sustainable fashion should not be

promoted only as an environmental choice. Durability can be communicated as a direct consumer benefit. A garment that lasts longer can be presented as both economically sensible and environmentally responsible. Lundblad and Davies (2016) similarly emphasised the importance of purchasing criteria and personal values in sustainable fashion consumption. Consumers are more likely to engage with sustainable fashion when the product provides meaningful value in addition to its ethical characteristics.

6.6 Social Media Influence

The findings indicate that social media has a moderate to strong influence on sustainable fashion awareness and purchasing decisions. Respondents reported using digital platforms as sources of information about fashion and sustainability.

This is important because Generation Z interacts with fashion through a highly visual digital environment. Influencers, brand campaigns, styling content and consumer reviews can affect how sustainable products are perceived.

At the same time, social media creates a challenge. The same platforms that promote sustainable fashion can also encourage frequent consumption. Sustainable fashion communication should therefore focus on quality, long-term use, styling versatility, repair, reuse and responsible consumption rather than simply encouraging another purchase.

6.7 Willingness to Pay More

Most respondents indicated some willingness to pay a slightly higher price for sustainable fashion. However, price remained a barrier for some participants.

This finding demonstrates that consumers may accept a moderate price premium when they understand the value of the product, but affordability remains important. Joergens (2006) and McNeill and Moore (2015) similarly demonstrate that ethical preferences can be weakened when consumers perceive sustainable products as financially demanding.

Brands therefore need to make sustainability economically accessible rather than treating it only as a premium market proposition.

7. Implications for the Fashion Industry

The findings provide several practical implications for fashion brands targeting Generation Z in India.First, brands should focus on **value rather than sustainability claims alone**. Quality, durability, comfort and design should be integrated into sustainability communication.

Second, brands should work towards **greater affordability and accessibility**. Sustainable fashion will have greater potential for wider adoption when products are available at price points that young consumers can realistically afford.

Third, brands should use **social media for education as well as promotion**. Content can explain materials, production processes, garment care, repair, recycling and responsible consumption in a simple and engaging way.

Fourth, brands should maintain **transparency in sustainability communication**. Consumers are increasingly exposed to sustainability claims, and unclear or exaggerated claims can weaken trust. Transparent information about materials, production and sourcing can strengthen credibility.

Finally, fashion businesses should explore **circular consumption models**, including repair, resale, recycling, upcycling and clothing take-back programmes. Such approaches can provide consumers with practical ways to participate in sustainability beyond purchasing new products.

8. Limitations of the Study

The study has several limitations. The most important limitation is the small sample size of 25 respondents. Since convenience sampling was used, the respondents cannot be considered statistically representative of the entire Generation Z population in India.

The research also relies on self-reported responses. Respondents may express stronger environmental concern than they demonstrate through actual purchasing behaviour. This possibility is particularly relevant to sustainability research because social expectations can influence responses.

The study is also limited to Generation Z and does not compare their behaviour with Millennials or other age groups. In addition, the research provides a snapshot of consumer behaviour at one point in

time. Consumer attitudes may change as sustainable products, technologies, social media trends and economic conditions develop.

Future research could use a larger and more diverse sample and compare respondents across different Indian cities, income groups and educational backgrounds. Qualitative interviews could also provide deeper insight into why consumers who support sustainable fashion do not always purchase it regularly.

9. Conclusion

The study demonstrates that Generation Z consumers represented in the sample have a generally positive attitude towards sustainable fashion and considerable awareness of environmental and ethical issues associated with the fashion industry. Many respondents had purchased sustainable fashion products at least once and expressed a preference for brands that demonstrate responsible practices.

However, awareness has not yet translated into consistent purchasing behaviour. The lower level of regular purchasing indicates the continued presence of an attitude-behaviour gap. Price, accessibility and availability remain important barriers, while product quality and durability strongly influence consumer decisions.Social media has emerged as an important source of awareness and influence. This creates an opportunity for fashion brands to engage Generation Z through credible, transparent and educational digital communication. However, digital marketing should avoid promoting sustainability while simultaneously encouraging excessive consumption.

The findings suggest that the future of sustainable fashion among Generation Z will depend on making responsible choices practical and desirable. Consumers need products that are sustainable, but they also expect quality, attractive design, affordability and convenience. For this reason, the transition towards sustainable fashion should not be presented as a choice between fashion and environmental responsibility. Instead, sustainability should become part of the overall value offered by a fashion product.Generation Z has the potential to become an important driver of sustainable fashion consumption in India. Realising this potential will require cooperation among brands, consumers, educators and policymakers. If sustainable products become more accessible, affordable, durable and trustworthy, positive consumer attitudes may increasingly translate into regular

purchasing behaviour.

References

- Ajzen, Icek. 1991. “The Theory of Planned Behavior.” *Organizational Behavior and Human Decision Processes* 50 (2): 179–211.
- Connell, Kyra Y. H. 2010. “Internal and External Barriers to Eco-Conscious Apparel Acquisition.” *International Journal of Consumer Studies* 34 (3): 279–286.
- Han, Jinghe, Yuri Seo, and Eunju Ko. 2017. “Staging Luxury Experiences for Understanding Sustainable Fashion Consumption: A Balance Theory Application.” *Journal of Business Research* 74: 162–167.
- Joergens, Catrin. 2006. “Ethical Fashion: Myth or Future Trend?” *Journal of Fashion Marketing and Management* 10 (3): 360–371.
- Lundblad, Louise, and Iain A. Davies. 2016. “The Values and Motivations behind Sustainable Fashion Consumption.” *Journal of Consumer Behaviour* 15 (2): 149–162.
- McNeill, Lisa, and Rebecca Moore. 2015. “Sustainable Fashion Consumption and the Fast Fashion Conundrum: Fashionable Consumers and Attitudes to Sustainability in Clothing Choice.” *International Journal of Consumer Studies* 39 (3): 212–222.
- Niinimäki, Kirsi. 2010. “Eco-Clothing, Consumer Identity and Ideology.” *Sustainable Development* 18 (3): 150–162.

Chapter 5

Deep fakes and Misinformation: Challenges for Digital Communication

Saburi Puri, Student,

Jagannath Community College

Abstract

The rapid advancement of artificial intelligence (AI) has transformed digital communication by enabling the creation of highly realistic synthetic media, commonly known as deepfakes. While this technology offers innovative applications in entertainment, education, and content creation, it also presents significant challenges by facilitating the spread of misinformation, fake news, and manipulated digital content. This chapter examines the growing impact of deepfake technology on digital communication, with a particular focus on its role in the dissemination of misinformation across social media platforms and online news environments. It explores the technological foundations of deepfakes, their implications for journalism, public opinion, and democratic processes, and the challenges faced by governments, media organizations, and technology companies in detecting and mitigating AI-generated deceptive content. The chapter also discusses existing regulatory frameworks, AI-based detection techniques, media literacy initiatives, and ethical considerations necessary to combat the misuse of deepfake technology.
By reviewing recent literature and real-world examples, this chapter aims to provide a comprehensive understanding of the risks associated with deepfakes while highlighting strategies for promoting responsible AI use, digital literacy, and trustworthy communication in the evolving digital landscape.

Keywords: Deepfakes, Artificial Intelligence, Misinformation, Digital Communication, Fake News, Social Media

1. Introduction

Digital communication has changed considerably over the past two decades. The movement from traditional mass media toward social media, video-sharing platforms, messaging applications, and algorithmically driven news environments has made communication faster, more participatory, and more accessible. Individuals are no longer only consumers of information. They can record, edit,

publish, distribute, and comment on content within seconds. This transformation has created new opportunities for public participation and creative expression, but it has also made the information environment increasingly vulnerable to manipulation.

Among the most significant developments in this changing environment is the emergence of deepfake technology. Deepfakes are synthetic or manipulated forms of audio, video, and imagery generated or modified through artificial intelligence. They can make an individual appear to say something they never said, perform an action they never performed, or appear in a situation in which they were never present. Advances in generative artificial intelligence have made the production of realistic synthetic media increasingly accessible to ordinary users rather than limiting it to highly specialized technical communities (Mirsky & Lee, 2021).

The concern surrounding deepfakes is not simply that false content can be produced. Manipulated media existed long before artificial intelligence. Photographs could be edited, videos could be selectively cut, and audio recordings could be taken out of context. Paris and Donovan (2019) therefore distinguish between technologically sophisticated deepfakes and what they describe as "cheap fakes," where simple editing, cropping, selective sequencing, or misleading contextualization can create deceptive impressions. The larger problem is the changing relationship between audiences and evidence. When people can no longer confidently determine whether an apparently authentic video is genuine, the credibility of visual communication itself becomes uncertain.

The challenge is particularly serious because contemporary communication is strongly visual. Videos and images often create an immediate impression of authenticity. Audiences may assume that seeing is equivalent to knowing. A realistic video of a public figure therefore has persuasive power even when the content is completely fabricated. Research by Vaccari and Chadwick (2020) suggests that political deepfakes can increase uncertainty and influence perceptions of political communication, particularly when audiences are already uncertain about the reliability of information.

Deepfakes also operate within an information ecosystem characterized by rapid sharing, personalization, political polarization, and declining attention spans. A fabricated video can move from one platform to another before journalists, fact-checkers, or authorities have an opportunity to verify it. In such circumstances, the speed of communication becomes an advantage for

misinformation. As Wardle and Derakhshan (2017) explain through the broader concept of information disorder, contemporary misinformation problems involve not only false information but also deliberately created disinformation and harmful information designed to manipulate audiences.

The deepfake problem, therefore, cannot be understood only as a technological problem. It is simultaneously a communication, journalistic, social, political, ethical, legal, and educational problem. Addressing it requires cooperation between technology developers, journalists, policymakers, educators, researchers, civil society organizations, and ordinary users.

2. Understanding Deepfake Technology

The term "deepfake" is associated with deep learning, a branch of artificial intelligence that uses neural networks to learn patterns from large amounts of data. Early deepfake systems relied heavily on generative adversarial networks, commonly known as GANs. GANs involve two competing neural networks. One generates synthetic content while another attempts to distinguish generated content from authentic material. Through repeated training, the generator becomes increasingly capable of producing realistic outputs.

More recent developments have expanded beyond traditional GAN-based systems. Diffusion models and other generative architectures can now create highly detailed images, modify existing photographs, synthesize voices, and produce increasingly sophisticated video content. Contemporary research identifies image, video, audio, and multimodal deepfakes as important categories of synthetic media (Croitoru et al., 2024). The increasing quality of these systems has created a moving target for researchers working on detection.

Deepfakes can generally be divided into several forms. Face-swapping involves replacing one person's face with another. Facial reenactment attempts to reproduce facial expressions or movements. Lip-syncing changes mouth movements to match new speech. Voice cloning can reproduce the characteristics of an individual's voice. Fully synthetic media can create people or scenes that never existed. These techniques can also be combined, allowing creators to produce highly convincing audio-visual fabrications.

There are legitimate uses for these technologies. Film production can use synthetic media to recreate historical environments or support visual effects. Education can employ synthetic characters for

simulations and interactive learning. Artists and content creators can experiment with new forms of storytelling. Accessibility applications can use synthetic voices to assist people who have lost their ability to speak. Therefore, deepfake technology should not automatically be treated as harmful.

The problem arises when the technology is used deceptively, without appropriate disclosure or consent. The same tools that can support creative communication can also be used to impersonate public figures, fabricate statements, manipulate political messages, conduct fraud, or damage reputations. Chesney and Citron (2019) argue that deepfakes create distinctive challenges because they can affect privacy, personal reputation, democratic processes, and national security simultaneously.

This dual nature makes policy difficult. A blanket prohibition on synthetic media could restrict legitimate creative and artistic applications. At the same time, a completely unrestricted environment could allow malicious actors to exploit the technology. The central challenge is therefore to distinguish responsible synthetic media from deceptive manipulation.

3. Deepfakes and the Spread of Misinformation

Misinformation refers broadly to false or inaccurate information that is shared without necessarily having the intention to cause harm. Disinformation, in contrast, involves deliberately creating or distributing false information to deceive or manipulate audiences. Deepfakes can contribute to both forms of information disorder.

An individual may unknowingly share an AI-generated video because they believe it is authentic. In this situation, the person becomes a distributor of misinformation without intending to deceive anyone. In another situation, a political organization or coordinated influence campaign may deliberately create a fabricated video to damage an opponent. Here, the same technology becomes an instrument of disinformation.

The distinction matters because solutions must address both the content and the communication environment. Removing a fake video may be necessary, but it does not address the audience's reasons for believing it. Similarly, educating users about deepfakes will not be sufficient if organized networks deliberately create and distribute thousands of misleading materials.

Deepfakes can be particularly effective when they confirm an existing belief. A person who already dislikes a political leader may be more willing to believe a fabricated video showing that leader behaving irresponsibly. This reflects the wider influence of confirmation bias and motivated reasoning in digital communication. Deepfakes therefore do not necessarily create beliefs from nothing. In many cases, they strengthen existing assumptions.

The emotional nature of social media further increases this risk. Content involving anger, fear, outrage, humiliation, or surprise tends to attract attention and engagement. A fabricated video designed around an emotionally powerful event may therefore travel more quickly than a carefully researched correction. The result is an imbalance between the speed of false information and the slower process of verification.

This problem existed before generative AI. However, deepfakes add a new layer because they can create apparently direct evidence. A false written statement can be questioned as a claim. A realistic video appears to provide visual proof. This can make audiences more vulnerable to manipulation.

Paris and Donovan (2019) emphasize that the danger does not come exclusively from sophisticated artificial intelligence. A misleading caption attached to a genuine photograph, an old video presented as a recent event, or a genuine speech edited to remove important context may be equally influential. Consequently, the fight against deepfakes must be part of a broader effort to improve information quality.

4. Deepfakes, Social Media, and Digital Communication

Social media platforms have fundamentally changed the way information is distributed. Traditional news organizations operated through editorial systems in which professional journalists, editors, and gatekeepers determined which information reached the public. Social media has introduced a more decentralized system where individuals, influencers, organizations, automated accounts, and anonymous users can all participate in information distribution.

This environment creates both opportunities and risks. On the positive side, social media can democratize communication and allow citizens to document events that might otherwise receive limited media attention. On the negative side, the absence of consistent editorial gatekeeping makes it easier for fabricated content to circulate.

The architecture of social platforms can amplify this problem. Content is often recommended on the basis of engagement, relevance, previous user behavior, and other algorithmic signals. A controversial deepfake may therefore receive greater visibility because users react strongly to it. Once the content becomes popular, its popularity itself can be interpreted as evidence of importance or credibility.

Another difficulty is cross-platform circulation. A fabricated video may initially appear on a small account, then be copied to messaging applications, reposted on short-video platforms, embedded in blogs, and eventually discussed by mainstream media. At each stage, the original source can become less visible. Audiences may encounter the content without knowing where it originated.

India provides an important example of this challenge because of its enormous and linguistically diverse digital population. During the 2024 Indian general election, AI-generated videos involving Bollywood actors were circulated as if the actors had expressed political opinions. The fabricated videos involving Aamir Khan and Ranveer Singh demonstrated how synthetic media could be used within a highly active political information environment (Associated Press, 2024).

The Indian context also demonstrates why deepfake detection cannot depend only on technological tools. Content may circulate across multiple languages, regional networks, political communities, and messaging applications. A detection system developed primarily for one linguistic or cultural environment may not perform equally well in another.

5. Impact on Journalism and News Credibility

Journalism has a particularly complicated relationship with deepfakes. On one hand, journalists are expected to investigate and expose manipulated media. On the other hand, journalists themselves may become victims of synthetic content or may unintentionally amplify a false claim by reporting on it before verification.

The traditional journalistic principle of verification becomes even more important in an environment of synthetic media. A journalist cannot assume that a video is authentic simply because it appears to come from a recognizable account. The source, metadata, context, date, location, original upload, and other available evidence need to be examined.

Deepfakes can also create what may be called a "verification burden." Journalists may have to spend substantial time authenticating material before publishing it. This is especially difficult in breaking news situations, where audiences expect immediate updates.

At the same time, journalists face another danger: the "liar's dividend." When society becomes aware that realistic synthetic media can be created, genuine evidence can also be dismissed as fake. Chesney and Citron (2019) highlight this broader concern. A public figure accused of wrongdoing may simply claim that an authentic recording is an AI-generated fake. Consequently, deepfakes do not only create false evidence. They can weaken confidence in genuine evidence.

This creates a difficult environment for investigative journalism. Journalists need increasingly sophisticated digital forensic skills, but they must also retain traditional reporting practices such as source verification, corroboration, interviews, documentation, and contextual analysis.

News organizations should therefore develop internal protocols for synthetic media verification. These protocols may include reverse image searches, frame-by-frame analysis, source tracing, metadata examination, audio analysis, geolocation, comparison with independent recordings, and consultation with forensic experts.

6. Deepfakes and Public Opinion

Public opinion is shaped not only by the information people receive but also by how credible they believe that information to be. Deepfakes threaten this process by creating uncertainty about the authenticity of media.

Vaccari and Chadwick (2020) found that exposure to synthetic political videos can affect perceptions of political communication and increase uncertainty. The significance of this finding goes beyond individual deception. Democracy depends on citizens having access to reasonably reliable information when forming political judgments.

A fabricated video of a political candidate making an offensive statement, announcing a policy, or appearing to support an extremist position could influence voter perceptions. Even if the video is later disproved, some viewers may remember the original claim. Corrections do not always reach the same audience as the original content.

The impact can also be psychological. When audiences repeatedly encounter claims that videos may be fake, they may become increasingly skeptical of all political media. Excessive skepticism can be as problematic as excessive trust. A healthy information environment requires critical evaluation rather than complete distrust.

The problem becomes particularly serious during elections. Political campaigns operate under tight deadlines, and false information can influence voters before corrective information becomes available. During the 2024 United States election, fabricated videos and other misleading materials were used to create false narratives about candidates and election processes. U.S. authorities also investigated fabricated election-related videos attributed to Russian influence actors (Associated Press, 2024).

These cases demonstrate that deepfakes can become tools of political influence. Their purpose may not always be to persuade audiences to support one particular candidate. Sometimes the objective may simply be to create confusion, reduce confidence in institutions, or increase political polarization.

7. Threats to Democratic Processes

Democratic communication depends on informed citizens, credible institutions, free expression, and access to reliable information. Deepfakes can affect all four.

First, they can manipulate political discourse. Second, they can impersonate candidates or public officials. Third, they can create false evidence of political misconduct. Fourth, they can increase distrust in election institutions. Finally, they can contribute to political polarization by creating emotionally charged narratives.

However, it is important not to exaggerate the threat. Deepfakes are not automatically successful. Their influence depends on context, distribution, audience characteristics, political circumstances, and the credibility of the source. Paris and Donovan (2019) remind us that relatively simple forms of manipulation can sometimes be more effective than technically sophisticated deepfakes.

The most important democratic risk may therefore be the cumulative effect of information uncertainty. If citizens begin to believe that any inconvenient evidence could be artificially

generated, public debate becomes more difficult. This creates an environment in which facts become increasingly contested. The solution cannot be to create an authority that simply declares which information is true. Such a system could itself become a tool for censorship or political control. Instead, democratic societies need transparent verification mechanisms, independent journalism, digital literacy, platform accountability, and clear legal safeguards.

8. Deepfake Detection: Opportunities and Limitations

Technological detection is one of the major responses to deepfakes. Researchers have developed systems that examine visual artifacts, facial movements, eye patterns, lighting inconsistencies, compression characteristics, audio irregularities, and other signals. Machine learning models can be trained on known authentic and synthetic examples to identify patterns associated with manipulation.

However, detection is an ongoing technological race. As generative systems improve, the artifacts that once made deepfakes relatively easy to identify may disappear. A detector trained on one generation of synthetic media may perform poorly against content created using a newer model. Recent research has highlighted the generalization problem. State-of-the-art detectors can perform well on known datasets but may fail when presented with deepfakes produced by previously unseen generators (Croitoru et al., 2024). This is a major concern because real-world content is rarely identical to controlled laboratory datasets. Human detection is also imperfect. Research involving thousands of participants found that ordinary viewers and machine-based detectors can make different types of errors, while combining human and machine judgments can sometimes improve performance (Groh et al., 2021). This suggests that the most effective approach may involve human-machine collaboration rather than complete dependence on automated detection. NIST's work on synthetic content similarly emphasizes that no single technical solution is sufficient. Content provenance, watermarking, labeling, detection, testing, and auditing are all parts of a broader synthetic media transparency ecosystem (Chandra et al., 2024).

Therefore, detection should be viewed as one layer of defense rather than a complete solution.

9. Content Provenance and Authentication

One promising approach is content provenance. Instead of asking only whether a piece of content

looks fake, provenance systems attempt to record information about how the content was created, edited, and distributed. The Coalition for Content Provenance and Authenticity has developed the C2PA standard and Content Credentials framework to establish information about the origin and history of digital media. Such systems can provide users with information about whether content was created or modified by particular tools or organizations.

The strength of provenance is that it shifts part of the problem from detection to authentication. Rather than waiting for a detector to identify suspicious pixels, a platform can potentially provide information about the content's history. However, provenance is not a perfect answer. C2PA itself notes that valid provenance information does not automatically mean that the content is truthful. A person can create authentic provenance for deliberately misleading content. Similarly, content without Content Credentials should not automatically be assumed to be fake or untrustworthy (C2PA, 2026). This distinction is essential. Authenticity of origin and truthfulness of meaning are different concepts. A genuine person may intentionally create misleading content. Provenance can tell us who created or modified something, but journalists and audiences still need to evaluate its context and claims.

10. Regulatory and Legal Responses

Governments around the world are increasingly developing rules related to synthetic media. Regulatory approaches differ according to national legal traditions, free speech protections, election systems, and technological capabilities.

The European Union has adopted one of the most significant regulatory approaches through the AI Act. Article 50 establishes transparency requirements concerning certain AI-generated and manipulated content. As of August 2, 2026, relevant transparency obligations require providers to enable machine-readable marking of synthetic content and require certain deployers to clearly disclose deepfakes and specific AI-generated content concerning matters of public interest (European Commission, 2026).

The regulatory principle is important because it does not necessarily prohibit synthetic media. Instead, it seeks to make the use of AI more transparent. Such transparency can help users distinguish between authentic recordings and synthetic creations. Legal regulation, however, must

balance competing interests. Excessively broad laws may restrict satire, artistic expression, political criticism, parody, or legitimate experimentation. At the same time, weak regulation may leave victims without effective remedies.

Legal frameworks also need to address consent and identity misuse. If a person's face or voice is used without permission, particularly in defamatory, fraudulent, or sexually exploitative material, existing legal protections may need to be supplemented by technology-specific provisions. International cooperation is also important because digital content does not respect national borders. A deepfake created in one country can be distributed globally within minutes. Fragmented regulation can therefore create enforcement gaps.

11. Ethical Challenges

The ethical questions surrounding deepfakes extend beyond whether content is technically legal. Consent is one of the central concerns. Individuals should have meaningful control over the use of their face, voice, identity, and personal characteristics.

The issue is especially serious when deepfakes are used to humiliate or sexually exploit individuals. Women and public figures have been disproportionately targeted by certain forms of non-consensual synthetic media. Such practices can cause serious reputational and psychological harm and raise questions about privacy, dignity, and digital rights.

Another ethical concern is transparency. Audiences deserve to know when they are viewing synthetic content, especially when the material could influence important decisions. Disclosure should be clear rather than hidden in technical terms that ordinary users cannot understand.

There is also an ethical responsibility on journalists and media organizations. A newsroom may publish a deepfake while attempting to expose it, but repeatedly displaying manipulated material can unintentionally increase its visibility. Journalists therefore need to consider whether showing the fake is necessary for public understanding and how it should be presented.

Technology companies also carry responsibility. Platform policies need to address the creation and distribution of deceptive synthetic media while avoiding arbitrary censorship. Companies should provide clear rules, accessible reporting mechanisms, timely responses, and transparency about

enforcement.

12. Media Literacy as a Long-Term Solution

Technology alone cannot solve the deepfake problem. Even the most sophisticated detection system cannot protect users who automatically share emotionally appealing content without checking its source.

Media literacy should therefore become an essential component of digital citizenship. Users need to understand how digital content is produced, edited, distributed, and monetized. They should be encouraged to pause before sharing suspicious material and ask basic questions: Who created this content? Where did it first appear? Is there an original source? Has a reliable news organization independently verified it? Does the date and context match the claim? Media literacy should also teach audiences that "looking real" is not sufficient evidence of authenticity. Educational institutions have an important role in this process. Schools, colleges, and universities can integrate AI literacy and digital verification into communication, journalism, computer science, and social science curricula. Journalism education should include practical exercises involving manipulated photographs, synthetic audio, AI-generated video, reverse searches, source verification, and fact-checking.

Media literacy should not become a lesson in simply distrusting social media. The goal should be informed skepticism. Students and citizens should learn how to evaluate evidence rather than reject everything that appears online.

13. Role of Media Organizations and Technology Companies

Media organizations need to develop institutional strategies rather than treating deepfakes as occasional anomalies. Every newsroom should have a basic synthetic media verification protocol.

First, suspicious material should be traced to its earliest available source. Second, journalists should seek independent confirmation. Third, technical verification should be used when necessary. Fourth, the content should be compared with authentic recordings or statements. Finally, journalists should clearly communicate uncertainty when verification is incomplete. Technology companies can complement these efforts through labeling, provenance tools, detection systems, user reporting, and

rapid-response mechanisms. Platforms should also invest in multilingual detection because misinformation frequently crosses linguistic boundaries. Another important issue is transparency. Platforms should publish regular reports explaining how much synthetic content they detect, how much they remove or label, how quickly they respond, and what limitations their systems have. The industry should also support researchers by providing appropriate datasets and evaluation frameworks while protecting privacy. Research into deepfake detection requires diverse examples representing different languages, cultures, devices, compression levels, and generation techniques.

14. Future Directions

The future of deepfake communication will likely involve increasingly sophisticated multimodal systems. Instead of producing only a manipulated face, future systems may simultaneously generate voice, facial expressions, body movement, background environments, and contextual details. This may make traditional visual inspection increasingly unreliable.

Research will therefore need to move toward multimodal and context-aware verification. Detection systems should examine not only individual frames but also relationships between speech, facial movement, background sound, lighting, geography, timestamps, and external evidence.Another important direction is real-time verification. As live synthetic video becomes more realistic, communication platforms may require systems capable of identifying manipulation while content is being streamed. There is also a need for stronger international research collaboration. Deepfake technologies evolve quickly, and isolated research efforts can become outdated. Shared benchmarks, diverse datasets, common standards, and transparent testing can help improve detection systems.At the same time, researchers should continue studying the human side of the problem. Understanding why people believe particular deepfakes may be just as important as improving algorithms. Factors such as political identity, emotional response, trust in institutions, social networks, and digital literacy all influence how people respond to synthetic media.

15. Conclusion

Deepfakes represent one of the most complex challenges facing contemporary digital communication. They demonstrate both the creative potential and the destructive possibilities of artificial intelligence. The technology itself is neither inherently good nor inherently harmful. Its

social consequences depend largely on how it is designed, used, distributed, regulated, and understood.The central challenge is not simply that people can create fake videos. Manipulated media has existed for generations. What makes the current moment different is the combination of realism, accessibility, speed, scale, and platform-driven distribution. A fabricated video can now be created with relatively limited technical expertise and distributed to a global audience almost instantly.For journalism, deepfakes reinforce the importance of verification and professional responsibility. For social media platforms, they create pressure to improve detection, labeling, provenance, and accountability. For governments, they raise difficult questions about regulation, privacy, political communication, and free expression. For educational institutions, they create an urgent need for AI and media literacy. For ordinary users, they make critical thinking an essential part of digital citizenship. Ultimately, the goal should not be to create a digital world in which every image, video, or audio recording is treated with suspicion. Such an environment would undermine the value of communication itself. The goal should be to create an information ecosystem in which people have better tools, stronger skills, clearer evidence, and greater institutional support for determining what they can reasonably trust. In an age where seeing is no longer sufficient to establish truth, responsible verification, transparency, and digital literacy will become central foundations of trustworthy communication.

References

- Associated Press. (2024, May 2). *As India votes, misinformation surges on social media: "The whole country is paying the price."* Associated Press. (AP News)
- Chandra, B., Dunietz, J., Roberts, K., Lee, Y., Fontana, P., & Awad, G. (2024). *Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency* (NIST AI 100-4). National Institute of Standards and Technology. (NIST)
- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*, 1753-1819. (BU Law Scholarly Commons)
- Coalition for Content Provenance and Authenticity. (2026). *C2PA specifications and Content Credentials*. (C2PA Specification)
- Croitoru, F. A., Hiji, A. I., Hondru, V., Ristea, N. C., Irofti, P., Popescu, M., Rusu, C., Ionescu, R. T., Khan, F. S., & Shah, M. (2024). *Deepfake media generation and detection in*

the generative AI era: A survey and outlook. arXiv. (arXiv)

- European Commission. (2026). *Code of practice on transparency of AI-generated content*. European Commission. (Digital Strategy)
- European Commission. (2026). *Quick facts: Transparency rules for AI systems*. European Commission. (Digital Strategy)
- Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2021). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*. (arXiv)
- Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys, 54*(1). (arXiv)
- Paris, B., & Donovan, J. (2019). *Deepfakes and cheap fakes: The manipulation of audio and visual evidence*. Data & Society Research Institute. (Data Society)

Chapter 6

Digital Literacy and AI Readiness among School Teachers

Mr. Abhishek Kumar

Abstract

Digital literacy and artificial intelligence readiness have become vital skills for school teachers in the rapidly developing educational landscape. As AI-powered tools increasingly impact teaching, learning, assessment, and classroom management, teachers are expected to hold not only basic digital skills but also the self-assurance and understanding required to integrate these technologies efficiently and sensibly. This study explores the implication of digital literacy and AI readiness among school teachers and highlights their role in refining the worth of education. The paper debates how digital literacy empowers teachers to access, evaluate, produce, and share educational resources while nurturing meaningful learning skills. It also observes AI readiness as a blend of technological knowledge, positive attitude, principled awareness, and willingness to implement innovative teaching practices. Despite the growing availability of digital tools, many educationalists continue to face problems such as inadequate professional training, limited technological infrastructure, lack of formal support, and concerns about data confidentiality and ethical use of AI. The study highlights the importance of constant professional development, policy support, and school-level initiatives to support teachers' digital proficiencies and prepare them for AI-assisted education. Additionally, the paper advocates for the accountable and comprehensive use of AI to guarantee that technological progression complements human teaching rather than substituting it. The findings emphasise that developing digital literacy and AI readiness among school teachers is fundamental to building future-ready classrooms. Authorizing teachers with the required knowledge, skills, and ethical understanding will allow them to create appealing, inclusive, and technology-enhanced learning surroundings that encounter the evolving requirements of twenty-first-century education.

Keywords: Digital Literacy, AI Readiness, School Teachers, Technology Integration, Digital Competence

1. Introduction and Context

1.1 Changing Landscape of School Education

School education has undergone a significant transformation due to the rapid development and integration of digital technologies. The incorporation of information and communication technology, online learning resources, digital materials, and interactive applications has changed the pedagogical processes and organizational structure of teaching and learning. Classrooms are gradually shifting from the traditional teacher-centred paradigm towards a more technologically integrated, interactive, collaborative, communicative, creative, and innovative model of instruction and learning. Consequently, teachers must not only be well-versed in their subjects but also proficient in utilizing digital technologies and resources (UNESCO, 2023).

The COVID-19 pandemic led to a faster worldwide acceptance of digital education and showed that technology is not merely a supplementary resource any more but an essential part of modern schooling. It became necessary for teachers to give their lessons online, produce digital learning materials, communicate with both students and parents via virtual platforms, and carry out remote assessments of the learners. While a number of teachers managed to make the transition successfully, the experience also highlighted big differences in the digital skills of teachers, the infrastructure available to them, and the support provided by their institutions (OECD, 2023). Because of this, there is now an even greater demand for teachers who are able to use technology confidently, critically, and responsibly.

In this situation, digital literacy has become a basic professional skill. It goes beyond simply being able to use devices or software since it also involves accessing, assessing, producing, communicating and managing information via digital technologies, as well as incorporating critical thinking, responsible digital citizenship, collaboration and the ethical use of online resources (Redecker, 2017). Teachers who have a high level of digital literacy are in a better position to design interesting learning experiences, to support a wide range of learners and to take an active part in professional learning communities.

1.2 Emergence of AI in Teaching and Learning

As part of the wider digital transformation, Artificial Intelligence (AI) has become one of the most influential technologies in the field of education. Tools powered by AI are becoming more and more common for providing personalised learning, offering adaptive teaching, carrying out automated assessments, generating feedback, using learning analytics, employing educational chatbots, and managing classrooms. There is the potential for these technologies to enhance teaching efficiency and to support students in accordance with their individual learning needs (Holmes et al., 2022).

The development of generative AI has now broadened the range of educational opportunities even more. With the help of AI programmes, teachers are able to prepare lesson plans, produce instructional materials, design differentiated activities, and give out feedback in a timely manner. AI can likewise promote accessibility by means of translation, text-to-speech, and other features that support inclusive learning. Yet the educational value of AI is mostly determined by teachers' capacity to assess and properly incorporate these tools. Since AI-generated content might include inaccuracies, bias, or information that is out of context, it has to be subject to professional judgement and critical examination before being used in the classroom.

AI should therefore be seen as a tool which assists teachers rather than one that replaces them; since effective teaching requires empathy, an understanding of context, ethical judgement, and pedagogical decision-making—qualities which are essentially human—the increasing role of AI in education underscores the need to prepare teachers to use such technologies responsibly and in a meaningful way (UNESCO, 2021).

1.3 Need for Teacher Digital Literacy and AI Readiness

Because of the growing use of AI in schools, the range of skills expected of teachers has gone beyond the usual digital skills. Although digital literacy forms the basis for incorporating technology, being ready for AI calls for extra knowledge, confidence, an awareness of ethical issues, and a readiness to embrace innovative approaches. Teachers need to know the capabilities and limitations of AI, assess the reliability of the outputs produced by AI, guard student data, and make sure that activities supported by AI are in line with educational objectives.

AI readiness may be regarded as a multi-dimensional ability which involves technological knowledge, a positive attitude towards innovation, confidence in the use of AI tools, an understanding of teaching methods, a sense of ethical responsibility, and a commitment to ongoing professional development. A teacher might have sufficient digital skills but could still be unprepared

to use AI effectively since the integration of AI requires critical evaluation, reflective practice, and making informed decisions. Research shows that a great many teachers still encounter difficulties including poor professional training, a lack of technological facilities, inadequate institutional support, and worries concerning the privacy, security, and ethical use of AI (OECD, 2023; UNESCO, 2023). The fact that such problems exist means that simply giving teachers access to technology is not enough. Educational systems should therefore make ongoing professional development and provide schools with supportive environments that enhance both their digital skills and their readiness for AI. Even though digital literacy and educational technology have been widely studied, a relatively smaller number of studies look at digital literacy and AI readiness as linked abilities that affect teachers' professional practice. Most of the existing research concentrates on digital competence or AI adoption on their own, which results in only a limited understanding of how these competencies together support the responsible integration of AI into school education. The present chapter removes this gap by offering an integrated view that ties together technological competence, pedagogical preparedness, ethical awareness, and institutional support.

1.4 Purpose, Scope, and Significance of the Chapter

The chapter explores the relationship between digital literacy and AI-readiness among school teachers in the context of the need to adapt to the rapid development of educational technologies. It examines how the level of teachers' digital competencies, confidence in using technology tools, awareness of ethical aspects, and readiness for innovation affects their ability to adopt AI in teaching, learning, assessment, and classroom management.

Particular attention is given to both the opportunities offered by AI-supported education and the challenges related to professional preparation, infrastructure, organizational readiness, and ethical concerns. Drawing on existing research and conceptual frameworks the discussion adopts a human-centred perspective, recognising that meaningful AI integration depends on strengthening teachers' professional capabilities rather than replacing human teaching. By analysing the current state-of-the-art and identifying implications for practice, the author supports the need to empower teachers to create inclusive and technology-driven learning environments in order to meet the demands of modern education and prepare students for the requirements of the 21st century.

The article offers an in-depth perspective on the concepts of digital literacy and AI readiness, providing insight into how these competencies prepare teachers to engage in AI-driven pedagogical

practices. It employs a human-centred approach, arguing that responsible integration of AI in education is possible only through developing not only technical competencies but also ethical awareness and organizational readiness. By combining different theoretical frameworks, empirical research, practical implications, and policy considerations, the article contributes valuable knowledge to the educational literature. The resulting perspective offers numerous implications for teachers, school leaders, teacher educators, policymakers, and researchers seeking to develop responsible, inclusive, and future-ready educational systems.

2. Conceptual and Theoretical Foundations

2.1 Digital Literacy and Digital Competence

Digital literacy is a concept that makes it possible for school teachers to interact with digital resources, access and evaluate information, create and share content, and communicate successfully using digital technologies. This concept includes such components as media literacy, information literacy, digital communication literacy, and content creation (Redecker, 2017). Digital competence, in its turn, is the ability to use digital technologies confidently and critically to support teaching and learning processes (European Commission, 2022). To be competent in digital technologies, school teachers need to acquire digital literacy skills which enable them to engage effectively in classroom activities and address learners' needs.

2.2 Concept of AI Readiness among Teachers

AI readiness refers to "teachers' competence, skills, attitudes, and confidence to understand, critically assess, and use Artificial Intelligence (AI) in an educational setting" and it "encompasses technological proficiency and competence, confidence and attitudes towards innovation, pedagogical competence, ethics, and continuous learning" (UNESCO, 2024). Teachers that are AI ready are able to critically apply and assess AI outputs while understanding their limitations and ensuring privacy and protection of their students. Therefore, the concept of AI readiness goes beyond digital literacy to include more diverse aspects of learning about artificial intelligence.

2.3 Relationship Between Digital Literacy and AI Readiness

Both digital literacy and AI readiness are related but different concepts where the second one is more than simply the use of technology and includes ethical concerns as well. The term digital literacy is associated with the ability to use digital technologies and resources appropriately while AI readiness entails being able to operate and interpret various AI tools critically and ethically to improve learning outcomes while understanding their pedagogical and ethical implications. Therefore, even though being digitally literate could be sufficient for some educational settings, it may not be enough to effectively utilise AI because it involves additional components beyond simply knowing how to use technology. Consequently, it can be argued that AI readiness builds upon digital competence while adding more layers of complexity because of the need to critically utilise and assess new AI solutions (Holmes et al., 2022).

2.4 Relevant Theories and Models of Technology Adoption

Several theoretical frameworks have been developed to understand teacher readiness to adopt technology and artificial intelligence in their practice. For instance, **the Technology Acceptance Model** (**TAM**) explains how teachers accept new technologies based on how useful and easy they find them to use (Davis, 1989). While this is a popular model that has been widely used to predict technology adoption, it does not consider additional pedagogical and ethical factors. Another framework is the **Technological Pedagogical Content Knowledge (TPACK)** which focuses on how teachers can combine their knowledge of pedagogy, technology, and content to effectively utilise technology in teaching and learning (Mishra & Koehler, 2006). However, it does not explicitly address some of the ethical issues involved in using AI in the classroom. Finally, another conceptual framework is the **European Framework for the Digital Competence of Educators (DigCompEdu)**, which is a generic competence framework for developing educators' digital literacy (Redecker, 2017). It offers a comprehensive guide on how to become digitally competent but says little about AI-specific competencies. Hence, these frameworks indicate that preparing teachers to utilise artificial intelligence requires combining several different components beyond simply enhancing their digital literacy.

3. Digital Literacy and AI Readiness in Teaching Practice

3.1 Digital Literacy as a Foundation for Technology Integration

Digital literacy is the foundation upon which technology-informed school education rests. It empowers teachers to locate, assess, produce, and share digital educational resources while also assisting them in choosing the most appropriate technologies for accomplishing their pedagogical goals (European Commission, 2022). It is noteworthy that digital literacy goes beyond mere technical proficiency to encompass knowledge of how to use technology in ways that complement instruction and foster deeper learning. As an extension, digital literacy creates the groundwork for teachers to embrace even more sophisticated tools, including artificial intelligence (AI).

3.2 AI Applications in Teaching and Learning

AI presents tremendous potential to be leveraged by teachers in various ways. First, it can be used to plan, design, execute, differentiate, and assess lessons and classes. For instance, an adaptive learning system can determine the best resources to recommend to individual learners based on their learning progress. On the same note, AI-driven tools can translate and make learning contents accessible to all learners (Holmes et al., 2022). Nevertheless, teachers should always cross-check information retrieved from AI tools since it can sometimes be erroneous or inclined to biases. Therefore, teachers should always ensure that they utilise resources produced by AI tools appropriately in accordance with the intended pedagogical goals (UNESCO, 2023).

3.3 AI-Assisted Assessment and Classroom Management

AI can also be used to evaluate tests and assignments automatically, especially if they contain objective response elements that can be scored with ease. In the same way, automatic attendance systems can be used to track learners' attendance records while also managing documentation and administrative duties. This would free teachers from some of their mundane but time-consuming duties while enabling them to focus on more crucial areas, such as teaching and learning (OECD, 2023). Similarly, teachers should always consider such soft skills as learners' creativity and socio-emotional behaviours when assessing their classes.

4. Opportunities and Educational Benefits

4.1 Personalised and Inclusive Learning

Digital literacy and AI-readiness enable teachers to create inclusive and learner-centred learning

environments that are sensitive to the needs of individual learners and respond to their diverse needs. Adaptive learning systems rely on AI algorithms to recommend the best resources to individual learners based on their unique characteristics and learning needs. For example, an AI-driven adaptive learning system can recommend resources to individual learners based on their learning styles, pace, and proficiencies. Ideally, such systems should be informative but not dictate the instructor's or learner's behavior, especially if the information is unreliable or lacks empirical backing. As an integral part of the educational process, teachers need to ensure that they always scrutinise information generated by AI tools and apply their professional skills and knowledge to utilise it appropriately. Specifically, teachers need to ensure that they utilise such data in a manner that is conducive to achieving the intended learning objectives. Likewise, educational materials generated by AI tools should be used to supplement classroom instruction while also considering individual learners' educational needs.

It is worth noting that AI-driven technologies, including translation tools, speech-to-text programmes, and accessibility features, can also be used to cater to the needs of language diverse and special needs learners, thus fostering inclusive education (Holmes et al., 2022). Nevertheless, teachers need to utilise their analytical skills to critically evaluate the usefulness of recommended resources in individual learners' contexts before using them as a basis for instruction.

4.2 Educational Resource Development and Teacher Efficiency

AI-assisted technologies can be leveraged by teachers to design study plans, prepare worksheets, and plan lessons. Generative AI-powered tools can be used in learning environments to produce various educational resources, including lesson plans and differentiated worksheets, which can save teachers' time and effort while also enhancing their productivity (UNESCO, 2023). However, teachers need to ensure that they do not utilise such tools indiscriminately but instead evaluate the veracity and appropriateness of the generated resources in accordance with the intended instructional objectives.

4.3 Assessment, Feedback, and Learning Support

Digital literacy and AI-readiness can be used by teachers to perform assessments and offer feedback to their learners. For instance, teachers can utilise intelligent marking systems to score objective

response questions automatically. In addition, they can use learning analytics tools to determine learners' progress and offer the feedback needed to promote their learning. In a similar vein, such tools can be used to identify individual learners' needs and recommend the interventions to address them (OECD, 2023). Ideally, such information should be used to supplement the teacher's professional skills in ascertaining the best way of promoting the learners' academic growth and addressing the gaps revealed by the analytics. It is paramount that educators remember that AI tools are not equipped to score higher-order cognitive skills or provide useful insights regarding learners' socio-emotional behaviours. Therefore, such tools should only be used to support the teachers' professional skills in teaching and learning.

5. Challenges, Ethical Issues, and Barriers

5.1 Inadequate Professional Training and Digital Competence

Despite a growing number of teachers embracing digital technologies, teachers are not well prepared in classroom AI. Teacher professional development programs are inadequate and emphasize more on foundational digital skills than the pedagogical implications of classroom AI. Additionally, teachers may not be motivated to learn about technology and are faced with a range of challenges such as technophobia (OECD, 2023). Therefore, professional development for teachers must include practice-centric experiences to improve digital proficiency and AI literacy.

5.2 Infrastructure and Digital Inequality

A lack of digital infrastructure and unequal access to educational technology constrains teachers' capacity to participate fully in an AI-powered education system. The availability of educational technology, internet connectivity, and electricity is often limited in remote and disadvantaged communities thus limiting opportunities for teachers to acquire digital skills and apply AI in classrooms (UNESCO, 2023). Lack of access to digital infrastructure and educational technology limits digital fluency, which undermines educational equity and hinders efforts to embrace new technologies in teaching and learning.

5.3 Data Privacy, Confidentiality, and Security

Teachers are concerned about the ethical implications of using AI in education such as

confidentiality, data security, and cyber safety. AI-powered education systems involve collecting and sharing students' and teachers' data. Therefore, teachers should ensure that they collect and store data securely and ethically to avoid risks such as data breaches, cybercrimes, and loss of privacy. Additionally, the use of cloud-based technologies raises questions about data ownership and privacy thus raising concerns regarding informed consent (UNESCO, 2021). Therefore, education professionals need to have a solid understanding of how to safeguard data against cyber threats and use digital technologies ethically and safely.

5.4 Ethical Issues and Responsible Use of AI

Most AI algorithms are susceptible to built-in biases that can generate prejudiced results against marginalized groups of society. Such biases often stem from the training data used to develop AI systems and can have negative implications such as limiting learning opportunities for some students based on gender, language, disability, and socioeconomic status among other factors. Therefore, the ethical use of AI in education requires auditing educational AI systems to identify and eliminate biases while ensuring inclusive and equitable education for all students.

AIs in education can create challenges concerning intellectual property, including copyrights, trademarks, and plagiarism among students. For instance, generative AI systems can encourage students to submit plagiarized content as their original work by providing answers to assignments, essays, and research papers thus promoting academic dishonesty and undermining integrity in education. Therefore, teachers should advise students on how to avoid academic misconduct when using generative AI in their studies while ensuring originality and authenticity of students' work. Thus, ethical considerations should guide the use of AI in education to promote academic integrity while avoiding intellectual property infringements (UNESCO, 2024; International Centre for Academic Integrity, 2021).

Teachers must critically evaluate information generated by AI systems before using it for instructional purposes. Generative algorithms can often generate information that is incorrect, misleading, and contains biased data. Additionally, generative AI systems can encourage inappropriate and unprofessional classroom behavior such as cheating and dishonesty among students by providing answers to assignments and examinations. Therefore, teachers must critically examine information generated by AI to avoid such unethical practices in classrooms. Furthermore,

issues surrounding academic integrity, copyright, intellectual property, responsibility, and accountability have become more complex with the increased use of generative AI in education (UNESCO, 2024). Therefore, ethical issues such as plagiarism, and inappropriate use of AI-generated content should be considered in developing policy and practices for using AI in education. Professional skepticism is critical in ensuring that AI-generated information is not used in education without being authenticated, verified, and evaluated for its accuracy, reliability, and validity. Finally, teachers should avoid relying on AI technology to replace critical thinking and decision-making in learning settings.

These challenges cannot be ignored if one wants to ensure that teachers can fully benefit from the opportunities presented by artificial intelligence. The issues range from the necessity for appropriate institutional policies and ethical frameworks to proper training, professional development, and governance. After all, educators need to know how they can integrate AI into their practices in a manner that promotes quality education while also being beneficial to their professional skills and competencies (Holmes et al., 2022).

6. Professional Development, School Support, and Policy

6.1 Continuous Professional Development for AI Readiness

Both continuous professional development to strengthen teachers' digital literacy and readiness to adopt AI, and the design of professional learning communities centred on classroom experimentation are critical areas for further research. Since AI is set to develop rapidly, it is critical for teachers to gain the knowledge and confidence to use such technology in classes. Ongoing professional development, including practice-focused training sessions, classroom experiments, reflections, peer-to-peer collaboration, and mentoring, should be prioritized in order to strengthen teachers' proficiency in using AI (OECD, 2023).

In addition, professional learning communities can support educators in sharing their practices and struggles with colleagues and developing effective approaches to teaching with AI (OECD, 2023).

6.2 Role of School Leadership and Institutional Support

Schools require competent leadership and administration to foster favourable conditions, structures,

and procedures to promote innovation through the adoption of AI technology. Leaders play the role of promoting the necessary digital infrastructure, internet access, technical support structures, free time for professional development, and ethical norms and organizational policies to assure data confidentiality and student privacy. Additionally, organizational cultures that encourage professionalism, collaboration, and mutual learning raise the likelihood of teachers experimenting with novel approaches and effectively incorporating AI into pedagogical practices (UNESCO, 2023). Such support from leadership is vital in motivating teachers to acquire further competencies and skills in the application of the technology in education.

6.3 Policy Support and Teacher-Education Programmes

National education policy and teacher-education programs play a critical role in empowering teachers with digital and AI capabilities. As per the **National Education Policy (NEP) 2020**, technology-enabled pedagogy promotes teacher professional learning, digital infrastructure, and new technologies to enhance learning outcomes (Ministry of Education, 2020). In addition, **National Digital Education Architecture (NDEAR)** focuses on strengthening digital education ecosystems and building interoperable digital platforms and ecosystems to enable accessible and reliable digital resources (Ministry of Education, 2021). Various teacher-education institutions must include technological pedagogy, AI, literacy, and data privacy policies in their agenda to empower teachers to adopt technologies in classrooms.

Other national initiatives such as the **DIKSHA platform, PM eVIDYA, and National Curriculum Framework** focus on providing digital resources and professional skills while promoting digital education through multiple modalities and encouraging the need for digital literacy, critical thinking, and related competencies in school education (Government of India, 2020; NCERT, 2023).

7. Future Prospects and Recommendations

7.1 Emerging Trends in Teacher AI Readiness

The future of school education will be increasingly shaped by such trends as the development of generative AI, intelligent tutoring systems, learning analytics, and personalized learning. These technologies can be used to plan and create lessons, teach, train, and evaluate students, and make professional decisions. The teacher's professional learning through such training can offer them

individualization and continuously update their knowledge and skills (UNESCO, 2024). Nevertheless, for the proper use of these technologies, it is necessary to develop teachers' critical thinking about the use of such innovations.

7.2 Future Competencies for School Teachers

As educational technologies evolve, teachers need to acquire more than just digital skills. Future teachers should be AI literate, digitally competent, critically minded, data-aware, ethically aware, and able to engage with AI to maximize learning potential. Educators should also be able to critically assess information generated by AI, protect data privacy, recognize bias in algorithms, and teach students how to use AI responsibly. Moreover, professional development and reflective practices will be critical in preparing educators to embrace emerging technologies and address their challenges (European Commission, 2022).

The future directions of the study should include the impact of AI on teachers' practices and students' learning in different contexts and settings. Moreover, more research should be done on how AI can be applied in different disciplines since educational settings and assessment practices may vary. Also, of great interest are the validity and reliability of assessments produced with the help of AI and their ethical implications (Holmes et al., 2022; UNESCO, 2024).

7.3 Recommendations for Teachers, Schools, and Policymakers

Strengthening AI readiness requires that teachers take part in professional development programmes, collaborate with peers, and critically assess technologies before adopting them in teaching practice. Schools need to put in place reliable digital infrastructure, provide appropriate training and support, and formulate ethical guidelines when adopting AI technologies. Teacher-education institutions need to incorporate AI literacy, digital pedagogy, ethics, and related issues in initial teacher education and professional development programmes for teachers. Additionally, policymakers need to develop strategies to promote access to digital technologies, professional development, and ethical and responsible AI while ensuring privacy and transparency in operations (Ministry of Education, 2020; UNESCO, 2021).

Moreover, the successful implementation of AI requires that governments, schools, teachers, technology developers, and researchers work together to create policies, strengthen digital

infrastructure, design professional development programmes, and ensure that the technology meets educational and ethical standards. In addition, partnerships will promote research, co-creation of knowledge, and continuous improvement of professional practices in AI-driven pedagogy (UNESCO, 2024; Holmes et al., 2022).

7.4 Human–AI Collaboration in Future Education

The future of education should be based on cooperation between teachers and artificial intelligence instead of replacing them. AI can perform many monotonous tasks, create teaching materials, and analyse data, but it cannot take over such skills as teaching experience, empathy, social intelligence, and ethics. The future of education cannot be effective without professional educators. A human-centred idea about the future of education implies the use of AI for better teaching. This concept is aimed at improving professional capabilities, providing quality education, and developing inclusive knowledge environments through continuous teacher training and education (Holmes et al., 2022; UNESCO, 2024).

8. Conclusion

8.1 Synthesis of Major Arguments

Digital literacy and AI readiness have become essential skills for school teachers in the age of information technology. The chapter discussed how digital literacy enables individuals to access, assess, create, and share information using digital tools. At the same time, the concept of AI readiness emphasizes additional components of technology knowledge, pedagogical competence, ethical understanding, confidence, and willingness to learn continuously. Thus, it allows teachers to shape teaching, learning, and evaluation processes and adequately address classroom challenges using artificial intelligence (AI) technologies while remaining committed to professional obligations and responsibilities (Holmes et al., 2022; UNESCO, 2024). According to the chapter, the capacity of AI to offer personalization, support planning, facilitate assessment, promote inclusion, and reduce teachers' administrative tasks will be possible through critical evaluation and selection of information and careful consideration of ethical aspects. Nevertheless, the chapter also outlined the key obstacles to implementing these opportunities, including the lack of teachers' preparation, technology, and resources, digital divide, ethical issues, data privacy and security, algorithmic

biases, and the absence of adequate institutional support (OECD, 2023; UNESCO, 2021).

8.2 Implications for Future-Ready School Education

In order to create schools fit for the future, it is essential to invest in continuous professional development and support teachers in acquiring confidence in the use of technologies, including AI, through leadership, infrastructure, and policy. Education faculties and policymakers should cooperate to promote the development of AI literacy, raise awareness, and ensure access to digital tools both inside and outside the classroom. Such measures might support teachers in addressing new developments in education and better responding to the needs of their students (Ministry of Education, 2020; European Commission, 2022).

References:

- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008
- European Commission. (2022). Digital Education Action Plan (2021–2027): 2022 progress report. European Commission. https://education.ec.europa.eu/focus-topics/digital-education/action-plan
- Holmes, W., Bialik, M., & Fadel, C. (2022). Artificial intelligence in education: Promises and implications for teaching and learning (2nd ed.). Centre for Curriculum Redesign.
- Ministry of Education. (2020). National Education Policy 2020. Government of India. https://www.education.gov.in
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
- OECD. (2023). Digital education outlook 2023: Towards an effective digital education ecosystem. OECD Publishing. https://doi.org/10.1787/b14bef59-en
- Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu. Publications Office of the European Union. https://doi.org/10.2760/159770
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO. https://unesdoc.unesco.org/
- UNESCO. (2023). Global education monitoring report 2023: Technology in education—A tool on whose terms? UNESCO. https://www.unesco.org/gem-report
- UNESCO. (2024). AI competency framework for teachers. UNESCO. https://www.unesco.org/
- Government of India. (2021). National Digital Education Architecture (NDEAR): A framework for digital education. Ministry of Education. https://ndear.gov.in/
- National Council of Educational Research and Training. (2023). National Curriculum .

Chapter 7

From Camera to Algorithm: The Transformation of Photography in the Age of AI

Jyanti Kumari, Student,

Jagannath Community College

Abstract

Photography has traditionally been understood as a way of recording reality through the camera. From the earliest film cameras to today's smartphone cameras, photography has continuously evolved alongside technological developments. The emergence of artificial intelligence, however, represents a more significant transformation because AI is no longer limited to helping photographers edit or organize photographs. It can now enhance images, reconstruct visual elements, generate entirely new images, and influence how photographs are produced, distributed, and interpreted. This development has created new possibilities for creativity and visual communication while simultaneously raising important questions about authenticity, ethics, authorship, and public trust.

This chapter examines the transformation of photography in the age of artificial intelligence from the perspective of mass communication and visual storytelling. It explores the transition from traditional photography to digital photography, computational photography, AI-assisted editing, and generative image creation. Particular attention is given to the changing role of the photographer, the impact of AI-generated visuals on photojournalism, and the increasing difficulty of distinguishing between photographs documenting real events and synthetic images that only appear photographic.

The chapter argues that artificial intelligence should not be understood simply as a replacement for the camera or the photographer. Instead, it represents a new stage in the development of visual communication. AI can assist photographers with technical and creative processes, but its use becomes particularly sensitive when photographs are expected to function as evidence of real events. Professional journalism organizations continue to emphasize accuracy, verification, transparency, and human editorial responsibility in the use of AI. The Associated Press, for example, updated its

newsroom standards in July 2026 and continues to prohibit generative AI from creating, altering, or enhancing news photography. It also requires AI-generated or manipulated material used in journalism to be clearly identified and presented in context (Associated Press, 2026).

Keywords: Artificial Intelligence, Photography, Photojournalism, Visual Communication, Generative AI, Digital Photography

1. Introduction

Photography has occupied a distinctive position within human communication because it combines technology, observation, aesthetics, memory, and evidence. Unlike purely verbal forms of communication, photographs appear to offer a direct relationship with the visible world. A photograph is generally understood as the result of a moment in which light from a scene reaches a recording medium through an optical system. This relationship between the physical world and the photographic image has historically contributed to photography's authority as a form of documentation.

The development of photography changed journalism, advertising, politics, science, art, and everyday communication. Photographs enabled distant audiences to witness wars, political events, disasters, social movements, cultural practices, and personal experiences. With the emergence of illustrated newspapers and magazines, photography became an important component of mass communication. Later, television, websites, smartphones, and social media transformed the speed and scale at which photographs could circulate.

However, the relationship between photography and reality has never been completely straightforward. Photographers select subjects, angles, moments, lenses, lighting conditions, and compositions. Editors crop and sequence images. Publishers select captions and headlines. Audiences interpret images through cultural and political contexts. Consequently, photographs have never been completely neutral representations of reality (Sontag, 1977).

The transition from analogue photography to digital photography intensified debates about photographic truth. Digital files could be copied, altered, compressed, transmitted, and manipulated without the physical limitations associated with film. William J. Mitchell (1992) described this transformation as a challenge to traditional assumptions about photographic truth. The digital image

made manipulation easier and weakened the assumption that every visible element in a photograph necessarily originated from the photographed scene.

Artificial intelligence introduces another major transformation. Contemporary AI systems can perform functions that were previously associated with skilled human photographers, retouchers, designers, and image editors. AI can identify subjects, remove unwanted objects, improve resolution, reconstruct missing areas, generate backgrounds, simulate lighting, create portraits of nonexistent people, and generate complete scenes from textual instructions. In this environment, the camera is no longer the only or even necessarily the primary source of the final visual image.

The transformation can therefore be understood as a movement from camera-based photography toward algorithmically mediated image production. The photograph is increasingly becoming the product of a technological ecosystem consisting of lenses, sensors, computational photography systems, machine learning models, editing applications, cloud platforms, recommendation algorithms, and social media interfaces.

2. From Chemical Photography to Digital Photography

The history of photography is a history of technological transformation. Early photographic processes required considerable technical knowledge and physical preparation. The photographer worked with chemical processes, glass plates, film, controlled exposure, and mechanical cameras. The limitations of film meant that photography involved significant decisions before and during exposure.

The development of smaller cameras, faster film, automatic exposure, autofocus, and digital sensors gradually simplified photographic production. Digital photography represented a major turning point because the photograph became a collection of numerical values rather than a chemical image fixed on film. The digital camera separated image capture from physical film processing. Photographers could immediately review photographs, store thousands of images, transmit them electronically, and manipulate them using software. This transformed professional photography and made photography accessible to a much wider population. The smartphone produced an even more significant transformation. Photography became embedded within a multifunctional communication device. The smartphone camera is not simply a miniature conventional camera. It is part of a computational

imaging system in which software contributes substantially to the final image.

Delbracio et al. (2021) explain that modern mobile photography has been transformed by computational methods including burst photography, noise reduction, super-resolution, and other forms of algorithmic processing. The resulting photograph is therefore increasingly constructed through a combination of optical capture and computational interpretation. This development changes the traditional understanding of the photographic moment. In conventional photography, the shutter captured a particular exposure. In computational photography, several frames may be captured and combined. Software can analyse those frames and determine which pixels should contribute to the final image. The smartphone therefore demonstrates that photography had already become computational before the arrival of generative AI. AI is not the beginning of algorithmic photography. Rather, generative AI extends an existing transformation.

3. The Emergence of Computational Photography

Computational photography represents a conceptual shift from photography as optical recording to photography as computational image formation. Traditional photography depends heavily on the physical characteristics of the camera, lens, sensor, exposure, and lighting. Computational photography supplements these characteristics with software algorithms.

High Dynamic Range photography is an important example. Instead of relying on a single exposure, a camera may capture multiple images at different exposure levels and combine them to create a photograph containing details across a wider range of light and shadow. Similarly, portrait mode on smartphones uses computational techniques to estimate depth and separate foreground subjects from backgrounds. Night photography can combine multiple frames to produce brighter and cleaner images in low-light environments. Super-resolution algorithms can reconstruct image details from multiple frames or computational models. These technologies have changed the meaning of photographic skill. The photographer still needs to understand composition, lighting, perspective, subject positioning, and timing. However, some technical limitations that previously required specialised equipment can now be addressed computationally.

The significance of this change is substantial for visual communication. If software determines exposure, colour, sharpness, depth, noise reduction, facial appearance, and other characteristics, then

the final image is already partly algorithmic before the photographer opens an editing application.

4. AI-Assisted Photography

AI-assisted photography refers to the use of artificial intelligence to improve, modify, organize, or interpret photographs. It exists on a spectrum ranging from relatively minor enhancement to substantial reconstruction.At the simplest level, AI can automate technical corrections. It can reduce noise, improve sharpness, correct exposure, recognize faces, optimise colour, and enhance low-resolution photographs. Such applications are often understood as extensions of traditional editing.

More advanced systems can remove unwanted objects, replace backgrounds, reconstruct damaged areas, enlarge photographs, modify facial expressions, or change aspects of lighting. These functions complicate the distinction between editing and creation.For photographers, AI can reduce repetitive work and allow more time for creative decision-making. A professional photographer dealing with hundreds or thousands of photographs can use AI-based systems to classify, sort, select, and organise images. AI can identify technically stronger photographs, detect faces, and assist with keywording.

AI can also support accessibility. Automated captioning and image description can help visually impaired audiences engage with photographic content. Image recognition can make large photographic archives easier to search.The positive potential of AI should therefore not be ignored. The technology can increase efficiency, democratise sophisticated editing techniques, and enable new forms of artistic experimentation.

Nevertheless, the ethical implications depend on the purpose of the photograph. In advertising or conceptual art, substantial AI modification may be acceptable when the audience understands the nature of the image. In documentary photography, news photography, scientific documentation, and evidence-based communication, the same modification may be unacceptable because it can change what the image claims to represent.

5. Generative AI and the Creation of Synthetic Photographs

Generative AI introduces a fundamentally different capability. Earlier digital editing systems generally started with an existing image. Generative AI can create visual content without requiring a

conventional photographic source.Text-to-image systems can generate scenes from natural language prompts. A user can describe a person, location, historical event, fashion concept, architectural environment, or imaginary situation, and the system can produce an image resembling a photograph.This changes the ontology of the photographic image. A conventional photograph normally begins with a physical scene. A generative image may begin with language, statistical patterns, or a latent representation learned from large datasets.The resulting image can possess many visual characteristics traditionally associated with photography. It may contain realistic lighting, depth of field, shadows, skin texture, lens effects, and photographic composition. Yet no camera may have captured the depicted event.

This distinction is critical for journalism. A synthetic image may look like a photograph while making a false claim about reality.Recent research demonstrates the growing sophistication of AI-generated images. Papa and Theodosiou (2026), for example, examined real and AI-generated images associated with news coverage and found that generative systems can reproduce visual characteristics associated with established news values. Their findings highlight how AI can imitate the visual language of photojournalism even when no corresponding event was photographed.

The problem is therefore not simply that AI images are becoming more realistic. The deeper issue is that the visual conventions through which audiences recognise "news photographs" can themselves be simulated.

6. Photography and the Crisis of Visual Truth

Photography has historically possessed a special evidentiary status. A photograph is often treated as proof that something existed, happened, or was present before a camera.This assumption has always required qualification. Photographs are selective. A photographer can choose one angle rather than another, capture one moment rather than another, and frame a scene in a way that affects interpretation. Captions can also change the meaning of photographs.

Sontag (1977) argued that photography shapes how people encounter and understand reality. Photographs do not simply record the world; they influence how the world is seen, remembered, and interpreted.Digital manipulation challenged the assumption that photographs were inherently trustworthy. AI-generated imagery takes the problem further because the image may contain no

photographic capture at all.

The consequence can be described as a **crisis of visual verification**. Audiences can no longer depend exclusively on visual appearance to determine whether an image represents an actual event.This has consequences beyond journalism. AI-generated photographs can affect political communication, advertising, historical memory, celebrity culture, education, social media, and interpersonal relationships.

For example, a synthetic photograph showing a political leader at an imaginary event may be circulated as evidence of political behaviour. A fabricated historical photograph may be presented as an archival document. An AI-generated portrait may be used to create a fictional identity. In each case, the image's persuasive power comes partly from its photographic appearance.Research on social media also indicates that AI-generated images can contribute to distorted perceptions of reality. A large-scale study by Ibrahim et al. (2025) examined AI-generated imagery and user reactions on Facebook and identified mechanisms through which fabricated visual content can encourage trust in false material.

7. The Changing Role of the Photographer

The rise of AI does not necessarily eliminate the photographer. Instead, it changes the photographer's responsibilities.Traditionally, the photographer's primary tasks included observing, composing, capturing, selecting, and editing. In AI-assisted photography, the photographer may increasingly become a curator, director, editor, prompt designer, verifier, and ethical decision-maker.The photographer's creative role may move further away from manually controlling every technical parameter and toward defining the desired visual outcome.

For example, an advertising photographer may photograph a product and use AI to create different environments around it. A portrait photographer may use AI to make controlled changes to lighting or background. A documentary photographer may use AI for cataloguing while maintaining strict limits on image alteration.This suggests that the future photographer may require a combination of traditional photographic literacy and AI literacy.Traditional photographic knowledge remains essential because AI cannot replace understanding of composition, visual storytelling, light, perspective, timing, culture, and human behaviour. At the same time, photographers need to

understand what AI tools are doing to their images.

The photographer must therefore be able to ask:

1. What was captured by the camera?
2. What was generated computationally?
3. What was changed after capture?
4. What information might have been lost?
5. Does the modification change the meaning of the photograph?
6. Does the audience need to be informed?

8. AI and Photojournalism

Photojournalism represents one of the most sensitive areas of AI-based image transformation because photographs often function as evidence.A news photograph does more than illustrate an article. It can establish presence, document events, provide emotional context, and influence public understanding. The credibility of the photograph is therefore connected to the credibility of journalism itself.The Associated Press has adopted a particularly cautious approach. Its updated July 2026 standards allow AI to assist with certain newsroom functions, including research, summarisation, transcription, translation, headlines, and other limited tasks, but AP continues to prohibit generative AI from creating, altering, or enhancing news photography. The standards also emphasise human verification and accountability (Associated Press, 2026). AI may help a journalist search an archive or organise photographs. However, generating a photograph of an event that did not occur would undermine the evidentiary function of photojournalism.

AP's earlier guidance similarly stated that generative AI should not be used to add or subtract elements from news photographs and advised journalists to verify suspicious material through source checking, reverse image searches, and comparison with trusted reporting (Associated Press, 2023). World Press Photo also emphasises accurate representation and prohibits intentional misleading through recreated or staged events within its ethical framework. These approaches suggest that the central principle of photojournalistic ethics remains representation rather than technological purity. The objective is not to prevent all technological assistance. The objective is to prevent technology from creating a false representation of reality.

9. Authenticity, Provenance, and Content Credentials

As AI-generated imagery becomes increasingly difficult to identify through visual inspection, provenance becomes more important. Provenance refers to information about where a piece of media came from and what happened to it during its lifecycle. In photography, provenance can include information about the camera, creator, editing process, software, and subsequent transformations. The Coalition for Content Provenance and Authenticity has developed Content Credentials as a technical approach to recording the origin and history of digital content. The C2PA framework uses cryptographically verifiable information to help establish the provenance of digital assets and document changes made during their lifecycle.

This approach does not simply ask audiences to identify whether an image "looks real." Instead, it attempts to provide information about how the image was created and modified. Such systems can be particularly valuable for journalism. A photograph captured by a professional camera could potentially carry information indicating its origin. Subsequent editing could be recorded as part of the asset's provenance. However, provenance systems should not be treated as a perfect solution. Metadata can be removed or lost during distribution. Not every platform necessarily preserves provenance information. Users may also misunderstand what a credential proves.Therefore, technological verification must be combined with journalistic verification and media literacy.

10. AI, Visual Storytelling, and Creativity

The discussion of AI photography should not be limited to risk. Generative AI also offers substantial creative possibilities.

Visual storytellers can use AI to develop concepts that would otherwise require expensive sets, locations, actors, or production equipment. Photographers and designers can experiment with surreal environments, speculative futures, historical reconstructions, conceptual narratives, and visual metaphors. For education, AI can provide students with opportunities to explore composition, visual narrative, storytelling, and design. Students can compare conventional photographs with synthetic images and discuss differences in realism, authorship, and meaning. For advertising, AI can enable rapid experimentation with different visual concepts. A brand may generate multiple visual possibilities before producing a final campaign. For artists, AI can become a medium rather than

simply a tool. The creative process may involve photography, prompting, selection, generation, compositing, and post-processing.

This development requires a broader understanding of photographic authorship. If a person photographs an image, edits it using AI, and then combines it with generated elements, authorship becomes distributed across human intention and machine processes.The important issue is transparency. An AI-generated artwork does not necessarily need to imitate documentary photography. If audiences understand that the image is synthetic or artistically constructed, the technology can support creativity without necessarily creating deception.

11. The Ethics of AI-Generated Human Images

Human portraits present particularly sensitive ethical questions. Generative AI can create realistic people who never existed. It can also modify real people's faces, bodies, clothing, expressions, or environments. These capabilities have implications for privacy, consent, dignity, reputation, and identity. The ability to generate realistic images of individuals creates risks of impersonation and non-consensual representation. AI can also reinforce unrealistic beauty standards and stereotypes. Recent research has found that disclosure does not necessarily eliminate the psychological effects of AI-generated images. Bennett et al. (2026), for example, reported that exposure to AI-generated influencer images affected body satisfaction even when participants were informed about the role of AI. The study also found that disclosure reduced trust in technology among informed participants.

This finding is important because it demonstrates that labeling is necessary but may not always be sufficient. Audiences need broader media literacy to understand how synthetic imagery influences perception. Photography education therefore needs to expand beyond camera operation and image composition. Students must also learn about algorithmic bias, synthetic media, consent, provenance, disclosure, and visual verification.

12. AI, Social Media, and the Speed of Visual Misinformation

Social media has transformed the distribution of photographs. A photograph can reach millions of people within minutes without passing through the editorial processes historically associated with newspapers or broadcasters. Generative AI increases the volume of available visual material.

Anyone with access to an image generation system can potentially create realistic scenes and distribute them through social platforms.

Audiences should be encouraged to ask basic questions:

Who created the image?

Where did it first appear?

Is there an original source?

Does independent reporting confirm the event?

Does the image contain unusual visual or contextual features?

Is the image labeled as AI-generated or manipulated?

13. The Problem of "Seeing Is Believing"

One of the most important consequences of generative AI is the weakening of the cultural assumption that seeing something means that it happened.

For centuries, photographs have contributed to a social relationship between images and evidence. A photograph could be used to support a claim because audiences assumed that the camera had physically encountered the scene. This phenomenon can be described as the liar's dividend. When synthetic images become common, individuals accused of wrongdoing can claim that authentic evidence is AI-generated. In such circumstances, the existence of deepfakes can create uncertainty even when genuine photographs are available.

This makes verification more important than visual intuition. Reuters fact-checking has demonstrated this problem in practice. In one case, a genuine photograph from a political rally was falsely accused of being AI-generated, while independent visual evidence and forensic analysis supported its authenticity (Reuters, 2024). Therefore, the AI era produces a paradox. Society must become more sceptical of images, but it must not become so sceptical that all visual evidence is dismissed.

14. The Transformation of Visual Journalism

AI is likely to transform visual journalism in several ways.First, routine production processes may become increasingly automated. Image classification, archiving, caption assistance, transcription, translation, and search can be supported by AI.Second, journalists may increasingly need to verify the provenance of images received through social media and unofficial sources.Third, news organisations may need dedicated synthetic media verification teams.Fourth, visual journalists may increasingly use AI for data visualisation, reconstruction, explanatory graphics, and clearly labeled illustrations.Fifth, editorial policies will become more important.The newsroom of the future may therefore contain photographers, video journalists, data journalists, AI specialists, fact-checkers, visual editors, and provenance specialists working together.This does not necessarily reduce the importance of the photographer. Instead, it increases the value of professional judgment.A camera can capture a moment, but a journalist determines whether that moment is relevant, ethical, contextualized, and accurately represented.

15. The Future of Photography Education

Photography education must respond to the changing technological environment.Traditional photography curricula generally focus on camera operation, exposure, composition, lighting, lenses, genres, editing, and visual storytelling. These remain important. However, students now also require knowledge of computational photography and generative AI.

16. Copyright, Authorship, and Ownership

AI-generated photography also raises difficult questions about copyright and ownership.Traditional photography generally has a clear human creator. The photographer operates the camera, makes creative decisions, and produces the image.AI complicates this model because the final visual may result from a combination of training data, algorithms, prompts, human selection, editing, and machine generation. Another major issue concerns the images used to train AI systems. Photographers and visual creators have questioned whether their copyrighted work should be used to train generative models without permission or compensation. These debates demonstrate that AI photography is not only a technological issue. It is also an economic and legal issue. Photographers depend on licensing, commissions, editorial assignments, stock photography, and creative rights. If

organisations can generate inexpensive synthetic images for commercial purposes, traditional markets may be disrupted. At the same time, new professional roles may emerge. Photographers may become AI visual directors, image consultants, dataset curators, authenticity specialists, or hybrid visual storytellers.

17. Authenticity as a Social Value

Authenticity has become one of the most important concepts in the AI photography debate. An authentic image does not necessarily mean an image that has never been edited. Professional photography has always involved editing, cropping, colour correction, exposure adjustment, and other interventions. A conceptual artwork can be authentic even when every element is digitally created because it openly communicates its constructed nature.A commercial advertisement can use synthetic imagery if it does not mislead audiences about a product or person.

A news photograph, however, has a different expectation because it claims to document an event. The same technological intervention can therefore have different ethical meanings in different contexts. This is why universal rules such as "AI is bad" or "AI is good" are inadequate. Ethical evaluation requires consideration of purpose, context, audience, degree of alteration, and potential harm.

18. Conclusion

The transformation of photography from camera to algorithm represents one of the most significant developments in contemporary visual communication. Photography has never been technologically static. It has moved from chemical processes to film, from film to digital sensors, from digital cameras to smartphones, and from conventional image processing to computational photography.

Artificial intelligence represents the next stage of this transformation.AI can enhance photographs, automate technical processes, assist with image organisation, create new visual possibilities, and generate images that resemble photographs with remarkable realism. These developments can expand creativity, improve accessibility, reduce repetitive work, and democratise sophisticated visual production.

However, the same technologies can also challenge the evidentiary value of photographs. When

synthetic images can imitate photojournalism, audiences can no longer depend only on visual appearance to determine authenticity. The problem becomes particularly serious when images are used to document political events, conflicts, disasters, public figures, historical events, or other matters of public interest.

The future of photography will therefore not simply be a competition between humans and machines. It will increasingly be a collaborative relationship between photographers, algorithms, cameras, software, platforms, editors, and audiences.

The photographer's role may change from simply operating a camera to directing an increasingly complex visual system. Yet the fundamental importance of human judgment will remain. Photography has value because people use it to communicate experiences, document events, preserve memories, construct narratives, and make claims about the world. Artificial intelligence can transform the technical process through which images are created, but it cannot remove the ethical responsibility attached to those claims.

References

- Associated Press. (2023). *Standards around generative AI*. The Associated Press. (The Associated Press)
- Associated Press. (2026, July 23). *AP updates newsroom standards for artificial intelligence*. The Associated Press. (The Associated Press)
- Barthes, R. (1981). *Camera lucida: Reflections on photography* (R. Howard, Trans.). Hill and Wang. (Original work published 1980).
- Bennett, C., et al. (2026). Filtering trust: Disclosing the role of artificial intelligence decreases trust in technology, but does not prevent harm to body image after viewing AI-generated content. *International Journal of Eating Disorders*. (DOI)
- Coalition for Content Provenance and Authenticity. (2026). *Content credentials: C2PA technical specifications*. C2PA. (C2PA)
- Delbracio, M., Kelly, D., Brown, M. S., & Milanfar, P. (2021). Mobile computational photography: A tour. *Annual Review of Vision Science, 7*, 571–604. (arXiv)
- Ibrahim, M., et al. (2025). The illusion of reality: How AI-generated images are fooling social media users. *Computers in Human Behavior, 176*, 108876. (DOI)

- Mitchell, W. J. (1992). *The reconfigured eye: Visual truth in the post-photographic era*. MIT Press. (MIT Press)
- Papa, V., & Theodosiou, Z. (2026). Visual news values in the age of AI: Exploring newsworthiness in generative image creation. *Journalism*. (DOI)
- Reuters. (2024, August 12). *Photo from Harris-Walz rally in Michigan was not AI-generated*. Reuters Fact Check. (Reuters)
- Sontag, S. (1977). *On photography*. Farrar, Straus and Giroux. (Google Books)
- Suo, J., Zhang, W., Gong, J., Yuan, X., Brady, D. J., & Dai, Q. (2021). Computational imaging and artificial intelligence: The next revolution of mobile vision. *IEEE Signal Processing Magazine, 38*(6), 34–43. (arXiv)
- Thorshaug, P. S., Schjelderup, H. G., & Papachristos, E. (2026). How AI-generated images affect human preference and trust in websites. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. (DOI)
- UNESCO. (2025). *AI and media ethics: Press councils from South-East Europe and Türkiye adopt landmark declaration*. UNESCO. (UNESCO)
- World Press Photo. (2026). *2026 contest code of ethics*. World Press Photo. (World Press Photo)

Chapter 8

From Manuscript to Machine: AI-Assisted Research, Writing, and the Reimagining of Scholarly Communication in the Digital Ecosystem

Sridevi. B,
Ph.D Full Time Research Scholar,
Department of Library and Information Science,
Periyar University, Salem, Tamil Nadu, India-636011.
E-Mail ID-sridevіphd@periyaruniversity.ac.in ,
Orcid ID- https://orcid.org/0009-0008-4872-7168

Jayaprakash. M,
University Librarian,
Periyar University, Salem, Tamil Nadu, India-636011.
E-Mail ID-lisjayaprakash@gmail.com ,
Orcid ID-https://orcid.org/0000-0003-4154-3055

Abstract

The rapid diffusion of artificial intelligence (AI) tools into academic life is reshaping how scholars discover literature, design studies, analyze data, draft manuscripts, and disseminate findings. This chapter provides a comprehensive examination of the evolving role of AI-assisted research and writing within the broader transformation of scholarly communication. Drawing on an interdisciplinary body of literature, it traces the shift from AI as a peripheral support tool to AI as an embedded collaborator in knowledge production, while critically engaging with concerns around authorship, academic integrity, factual reliability, bias, and the epistemic authority of generative systems. The chapter reviews institutional, publisher, and professional-body responses to AI adoption; considers questions of global equity in access to AI tools; and offers a structured framework of recommendations for responsible integration of AI across the research lifecycle. It concludes that the changing digital ecosystem requires scholars, publishers, and institutions to jointly develop new norms of transparency, accountability, and disclosure so that AI's benefits efficiency,

accessibility, and the democratization of research can be realized without compromising scholarly rigor or public trust in science.

Keywords: artificial intelligence, scholarly communication, academic writing, generative AI, research ethics, digital ecosystem, authorship, academic integrity

1. Introduction

Scholarly communication has always evolved in tandem with the technologies available to researchers from the printing press, to the typewriter, to word processors, to digital repositories and open-access platforms. Each transition altered not merely the speed of knowledge production but also its social organization: who could participate, what counted as a legitimate contribution, and how trust was established and maintained. The emergence of generative artificial intelligence particularly large language models (LLMs) such as GPT-based systems represents the latest, and arguably most disruptive, shift in this long trajectory (Bommasani et al., 2021).

AI tools are now used across nearly every stage of the research lifecycle: literature searching and synthesis, hypothesis generation, statistical analysis, manuscript drafting, language editing, visualization, and even peer review support (Dwivedi et al., 2023). Unlike earlier digital tools, which primarily automated mechanical tasks (formatting, reference management, spell-checking), generative AI systems can produce fluent, contextually appropriate prose, summarize complex arguments, and simulate expert-level reasoning across disciplines (Floridi & Chiriatti, 2020). This qualitative shift from tool-as-instrument to tool-as-collaborator has generated both considerable enthusiasm and considerable unease within the academic community.

Proponents highlight AI's potential to reduce the burden of routine tasks, improve accessibility for non-native English speakers, accelerate literature reviews, and democratize participation in global scholarship (Lund & Wang, 2023). Critics, meanwhile, warn of risks to originality, academic integrity, and the reliability of AI-generated content, given the well-documented tendency of generative models to produce plausible-sounding but factually incorrect outputs—commonly termed "hallucinations" (Alkaissi & McFarlane, 2023; van Dis et al., 2023).

This chapter situates AI-assisted research and writing within the broader theme of the changing digital ecosystem, asking four guiding questions:

1. **How is AI altering the practices and workflows of scholarly communication?**
2. **What ethical, epistemological, and institutional questions does this raise?**
3. **How have publishers, universities, and professional bodies responded so far?**

4. **What pioneering frameworks might guide responsible adoption going forward?**

The chapter proceeds as follows. Section 2 outlines the changing digital ecosystem of scholarship as background context. Section 3 reviews the literature on AI applications across the research lifecycle. Section 4 offers a conceptual framework distinguishing AI-as-collaborator from AI-as-author. Section 5 examines authorship, integrity, and epistemic trust in depth. Section 6 surveys institutional and publisher responses. Section 7 considers global equity concerns. Section 8 presents illustrative scenarios drawn from documented cases in the literature and press. Sections 9 and 10 discuss opportunities and risks, respectively. Section 11 proposes a recommendations framework, and Section 12 outlines directions for future research before the chapter concludes in Section 13.

2. Background: The Changing Digital Ecosystem of Scholarship

Before examining AI specifically, it is useful to situate it within the wider digital transformation of scholarly communication over the past two decades. This transformation includes the rise of open-access publishing, the proliferation of preprint servers, the growth of altmetrics and alternative impact measures, and the increasing role of social media and digital platforms in research dissemination (Zawacki-Richter et al., 2019). Each of these developments has incrementally reshaped who can access knowledge, how quickly findings circulate, and how scholarly credibility is signaled and verified.

Generative AI enters this already-shifting ecosystem rather than a static one. It intersects with open science movements by potentially lowering barriers to participation (e.g., through AI-assisted translation and writing support for non-native English speakers), while simultaneously raising new questions about verification and trust that echo and in some ways intensify earlier debates about predatory publishing, reproducibility, and peer review quality (Hosseini et al., 2023). Understanding AI-assisted research and writing therefore requires attention not only to the technology itself but to the institutional and infrastructural context into which it is being absorbed.

3. AI Across the Research Lifecycle: A Literature Review

3.1 Literature Discovery and Synthesis

AI-powered tools now assist researchers in navigating an increasingly overwhelming volume of published literature. Machine learning–based systems can cluster related studies, summarize findings, and identify research gaps at a scale unattainable through manual review (Zawacki-Richter

et al., 2019). Such tools lower the barrier to conducting systematic reviews and meta-analyses, allowing researchers to process thousands of abstracts in a fraction of the time required by traditional methods. However, scholars caution that AI-generated syntheses still require human verification for accuracy, contextual nuance, and correct attribution, since even sophisticated summarization tools can misrepresent nuanced findings or omit important contextual caveats (Aydın & Karaarslan, 2022).

3.2 Research Design and Data Analysis

Beyond literature review, AI is increasingly embedded in research design and analysis. Machine learning techniques support pattern detection in large datasets, assist in experimental design optimization, and can flag anomalies that might otherwise go unnoticed by human analysts. In fields such as biomedicine, AI-assisted analysis has become integral to processing high-dimensional data (King, 2023). This integration, however, raises questions about the interpretability of AI-derived findings and the risk that researchers may over-rely on "black box" outputs without fully understanding the underlying analytic logic.

3.3 Writing and Language Editing

Generative AI has become particularly prominent in manuscript drafting and language refinement. For researchers writing in a second language, AI tools can substantially improve clarity, grammar, and overall readability, reducing the linguistic disadvantage historically faced by non-native English speakers in international publishing (Kitamura, 2023). Salvagno, Taccone, and Gerli (2023) argue that AI writing assistance can meaningfully support scientific communication when used transparently and combined with expert human review.

At the same time, this capability has intensified debate over what constitutes acceptable use versus academic misconduct, particularly when AI contributions are undisclosed (Cotton et al., 2023). A widely discussed illustration of the risks of insufficient human oversight involves published papers in which residual AI-generated phrases such as boilerplate disclaimers characteristic of chatbot outputs—were inadvertently left in the final text, prompting corrections and public scrutiny (Else, 2023). Such incidents underscore that even when AI assistance is not inherently problematic, careless integration can visibly compromise a manuscript's credibility.

3.4 Peer Review and Editorial Processes

AI is also entering the editorial pipeline, with journals experimenting with AI-assisted plagiarism detection, reviewer-manuscript matching, and preliminary manuscript screening for methodological red flags. While these applications promise efficiency gains for overburdened editorial systems, they raise questions about transparency and the potential for algorithmic bias to influence which research is deemed publishable or which reviewers are selected (Hosseini et al., 2023). There is also growing debate about whether reviewers themselves should be permitted to use AI tools to help draft review reports, given confidentiality concerns around uploading unpublished manuscripts to third-party AI systems (COPE, 2023).

3.5 Dissemination, Metrics, and Public Engagement

Finally, AI is reshaping how research is disseminated and evaluated after publication. AI-driven recommendation systems influence which papers gain visibility on academic search platforms and social media, while natural language generation tools are increasingly used to produce lay summaries, press releases, and multilingual translations of research findings, potentially broadening public engagement with science (Lund & Wang, 2023). This raises new questions about the fidelity of AI-generated science communication and the risk of oversimplification or distortion when complex findings are translated into accessible language without adequate expert oversight.

4. Conceptual Framework: AI as Collaborator, Not Author

A recurring theme across the literature is the need to distinguish between AI as a *supportive collaborator* in the research process and AI as an *author* in the formal scholarly sense. This chapter adopts a three-tier conceptual framework to organize this distinction:

Tier 1 — Mechanical Assistance: AI is used for narrowly defined, verifiable tasks such as grammar correction, formatting, or translation. Risk to integrity is low, and disclosure expectations are correspondingly minimal in most current guidelines.

Tier 2 — Cognitive Augmentation: AI is used to assist with literature synthesis, brainstorming, drafting sections of text, or generating first-pass data interpretations. Here, human verification, disclosure, and critical evaluation become essential, since the AI's contribution shapes substantive content rather than surface form (Dwivedi et al., 2023).

Tier 3 — Autonomous Generation: AI is used to generate substantial portions of a manuscript's

intellectual content—such as full literature reviews, discussion sections, or even hypotheses—with minimal human input. Current consensus across major publishers holds that this tier raises the most serious integrity concerns and, in its unsupervised form, is broadly discouraged or prohibited (Thorp, 2023; Stokel-Walker, 2023).

This tiered framework helps clarify why blanket bans on "AI use" are both impractical and conceptually imprecise: the ethical and epistemic stakes differ substantially depending on *where* in this spectrum a given use falls.

5. Authorship, Integrity, and Epistemic Trust

5.1 The Authorship Debate

A central controversy in this space concerns authorship. Major publishers and editorial bodies, including *Nature* and *Science*, have explicitly stated that AI tools cannot be listed as authors, since authorship implies accountability that a non-sentient system cannot bear (Thorp, 2023; Stokel-Walker, 2023). This position reflects a broader consensus that while AI can assist in the *mechanics* of writing, the *intellectual responsibility* for a work's claims must remain with human researchers. Authorship, in this view, is inseparable from accountability: an author must be able to answer for the accuracy, originality, and ethical soundness of a work, a capacity that current AI systems do not possess.

This consensus has led journals and institutions to develop disclosure requirements, mandating that authors explicitly state where and how AI tools were used in manuscript preparation (Dwivedi et al., 2023). Such policies aim to preserve trust in the scholarly record while acknowledging that outright prohibition is neither enforceable nor necessarily desirable (Anders, 2023).

5.2 Hallucination and Factual Reliability

Because generative models predict statistically plausible text rather than verify truth, unsupervised use in research writing risks introducing fabricated citations, misattributed findings, or subtly distorted claims into the literature. Alkaissi and McFarlane (2023) document cases in which AI systems generated entirely fabricated but stylistically convincing references and clinical claims in medical writing contexts a phenomenon they term "artificial hallucination." Similarly, Gao et al. (2023) found that AI-generated scientific abstracts could convincingly mimic genuine research abstracts, sometimes evading detection even by experienced reviewers, raising concerns about the

potential for AI-generated pseudo-research to enter the scholarly record undetected.

These findings underscore the continued necessity of human oversight, critical evaluation, and domain expertise positioning AI as an augmentation of scholarly judgment rather than a replacement for it. Van Dis et al. (2023) argue that addressing this challenge requires coordinated action across five priority areas: verifying AI outputs, developing rules for accountability, investing in genuinely open-source AI tools for research, embedding AI literacy in academic training, and fostering ongoing debate about the appropriate boundaries of AI use in science.

5.3 Plagiarism, Originality, and Detection Tools

The rise of generative AI has also complicated existing frameworks for detecting plagiarism and ensuring originality. Traditional plagiarism-detection software identifies textual overlap with existing sources, but AI-generated text is, by construction, often original at the sentence level even when it derives its substance from patterns learned across vast training corpora. This has spurred the development of AI-text detection tools, though their reliability remains contested, with reported concerns about both false positives (incorrectly flagging human-written text) and false negatives (failing to detect AI-assisted content) (Cotton et al., 2023). The unreliability of detection tools reinforces the argument that disclosure-based transparency, rather than detection-based policing, may be the more sustainable long-term approach.

5.4 Bias and Representational Concerns

A further concern involves the propagation of bias through AI-assisted research and writing. Because large language models are trained on existing bodies of text, they can reproduce and even amplify biases present in their training data including linguistic biases favoring dominant academic traditions, disciplinary framings, and geographic or cultural perspectives (Bommasani et al., 2021). This raises the risk that widespread reliance on AI tools could subtly homogenize academic writing styles and argumentative structures, potentially marginalizing alternative epistemic traditions and non-Western scholarly voices unless actively counteracted.

6. Institutional and Publisher Responses

6.1 Journal and Publisher Policies

In response to these concerns, major publishers have moved relatively quickly to establish formal policies. As noted, both *Nature* and *Science* have stated that AI systems cannot be credited as authors (Thorp, 2023; Stokel-Walker, 2023), while requiring authors to disclose AI use in the methods or acknowledgments sections of a manuscript. Other major publishers have adopted broadly similar frameworks, generally converging on a model in which AI use is permitted but must be transparently disclosed and remains subject to full human accountability (Dwivedi et al., 2023).

6.2 University and Funder Guidelines

Universities and research funders have likewise begun developing institutional guidance, though these frameworks remain considerably more fragmented than publisher policies, varying widely across institutions and countries. Common elements include requirements for AI-use disclosure in theses and grant applications, updated academic integrity policies explicitly addressing generative AI, and, in some cases, the introduction of AI literacy training for graduate students and early-career researchers (Cotton et al., 2023).

6.3 Professional and Editorial Bodies

Professional bodies overseeing research ethics and publication standards have also issued guidance. The Committee on Publication Ethics (COPE) has articulated a position that AI tools cannot meet the criteria for authorship and that their use should be transparently disclosed, with authors remaining fully responsible for submitted content (COPE, 2023). Similarly, editorial guidance from major medical journal consortia emphasizes that AI-assisted content must be verified for accuracy and that confidential, unpublished material such as manuscripts under peer review—should not be uploaded to public AI tools due to data privacy and confidentiality concerns.

Taken together, these institutional responses reveal an emerging, if still evolving, consensus: AI is broadly accepted as a legitimate research and writing aid, provided its use is transparent, verified, and does not substitute for human accountability.

7. Global Equity and the Digital Divide in AI-Assisted Scholarship

The integration of AI into scholarly communication does not occur on a level playing field. Access to advanced AI tools particularly premium, more capable models is unevenly distributed across institutions, disciplines, and world regions, raising concerns that AI could exacerbate rather than

alleviate existing inequities in global scholarly participation (Zawacki-Richter et al., 2019). Researchers at well-funded institutions in high-income countries may have privileged access to cutting-edge AI tools, institutional training, and technical support, while scholars elsewhere may face cost barriers, infrastructural limitations, or restrictions on data transmission to AI providers located in other jurisdictions.

At the same time, AI tools hold genuine potential to *reduce* certain longstanding inequities—particularly the disadvantage faced by non-native English speakers in a scholarly publishing system still dominated by English-language journals (Kitamura, 2023). Realizing this equity-enhancing potential, rather than allowing AI to become a new axis of stratification, will require deliberate policy attention: subsidized or open-access AI tools for under-resourced institutions, capacity-building initiatives, and publisher policies that do not inadvertently penalize legitimate, disclosed AI-assisted writing support.

8. Illustrative Scenarios from the Literature and Press

Several documented cases illustrate both the promise and the pitfalls of AI-assisted scholarly work. Reported instances in which peer-reviewed papers retained visible AI-generated boilerplate phrases—text characteristic of chatbot disclaimers drew widespread public and editorial attention, prompting corrections and renewed scrutiny of editorial safeguards (Else, 2023). Separately, controlled studies have shown that AI-generated research abstracts can be difficult even for trained scientists to distinguish from genuine human-written abstracts, raising concerns about the potential for AI to be used to generate convincing but unsubstantiated research summaries (Gao et al., 2023).

These cases do not suggest that AI assistance is inherently harmful to scholarly communication; rather, they illustrate the consequences of insufficient human oversight and inadequate editorial verification processes. They reinforce the chapter's central argument that responsible integration—not prohibition—is the more realistic and productive institutional response.

9. Opportunities in the Changing Digital Ecosystem

Despite the risks outlined above, the literature identifies substantial opportunities associated with AI-assisted research and writing:

- **Efficiency gains**: Automating routine tasks such as reference formatting, initial drafting, and literature summarization frees researcher time for higher-order intellectual work (Dwivedi et al., 2023).
- **Accessibility and inclusion**: AI-assisted language editing can reduce linguistic barriers for non-native English-speaking researchers, potentially broadening global participation in international scholarship (Kitamura, 2023).
- **Accelerated discovery**: AI-assisted literature synthesis and pattern detection can accelerate the identification of research gaps and novel hypotheses (Zawacki-Richter et al., 2019).
- **Enhanced public engagement**: AI-assisted science communication tools can help translate complex findings into accessible formats for broader public audiences, provided fidelity to the original findings is maintained (Lund & Wang, 2023).

9. Risks and Challenges

- **Factual unreliability and hallucination**, with the potential to introduce fabricated or distorted claims into the scholarly record (Alkaissi & McFarlane, 2023; van Dis et al., 2023).
- **Erosion of authorship accountability**, if AI contributions are undisclosed or if human oversight is insufficient (Thorp, 2023).
- **Detection and enforcement limitations**, given the unreliability of current AI-text detection tools (Cotton et al., 2023).
- **Bias propagation and epistemic homogenization**, particularly affecting non-dominant academic traditions and languages (Bommasani et al., 2021).
- **Unequal access**, risking a new digital divide in scholarly participation (Zawacki-Richter et al., 2019).

10. A Recommendations Framework for Responsible Integration

Synthesizing the literature reviewed above, this chapter proposes the following principles for the responsible use of AI in research, writing, and scholarly communication, organized by stakeholder group:

For individual researchers:

5. Disclose the nature and extent of AI assistance used in any given work.

6. Independently verify all AI-generated content citations, summaries, data interpretations before inclusion in scholarly outputs.
7. Retain full intellectual and ethical responsibility for all submitted work, regardless of AI involvement.

For journals and publishers:

8. Maintain clear, consistently enforced policies distinguishing acceptable AI assistance from prohibited autonomous generation.
9. Require explicit AI-use disclosure statements as a standard manuscript submission element.
10. Invest in editorial training to help reviewers and editors recognize signs of unsupervised or undisclosed AI use.

For universities and funders:

11. Develop clear, updated academic integrity policies that explicitly address generative AI.
12. Provide AI literacy training for researchers at all career stages, particularly graduate students and early-career scholars.
13. Ensure equitable institutional access to AI tools to prevent AI from becoming a new axis of academic stratification.

For the broader scholarly ecosystem:

14. Support the development of transparent, ideally open-source, AI tools tailored to research use cases, reducing dependency on opaque commercial systems (van Dis et al., 2023).
15. Foster ongoing, cross-disciplinary dialogue on the evolving boundaries of acceptable AI use, recognizing that static policies will likely require regular revision as the technology continues to develop (Dwivedi et al., 2023; Hosseini et al., 2023).

11. Directions for Future Research

Several avenues merit further scholarly attention. First, longitudinal studies are needed to assess how AI-assisted writing practices affect the long-term quality, originality, and diversity of scholarly output across disciplines. Second, further empirical work should examine the actual (rather than perceived) prevalence of undisclosed AI use in published literature, and the effectiveness of various disclosure and detection mechanisms. Third, comparative research across world regions and institutional contexts could illuminate how unequal access to AI tools is shaping global patterns of scholarly participation. Finally, research is needed on the pedagogical dimensions of AI

integration—specifically, how AI-assisted writing tools affect the development of critical thinking, analytical writing, and research skills among early-career scholars.

12. Conclusion

AI-assisted research and writing mark a pivotal development in the ongoing transformation of scholarly communication. Rather than displacing human scholarship, current evidence suggests AI is best positioned as a collaborative tool that enhances efficiency and accessibility while demanding renewed attention to integrity, transparency, and equity. The tiered framework proposed in this chapter distinguishing mechanical assistance, cognitive augmentation, and autonomous generation offers one way of clarifying where the most significant ethical and epistemic risks lie, and where existing safeguards are most urgently needed.

As the digital ecosystem continues to evolve, pioneering academic engagement with these tools critical, informed, and ethically grounded will be essential to ensuring that AI strengthens, rather than undermines, the credibility of the scholarly record. The task ahead is not to resist this transformation but to shape it deliberately: building norms, policies, and infrastructures that allow AI's genuine benefits to be realized without compromising the foundational values of accountability, originality, and trust that underpin scholarly communication

References

- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*(2), e35179.
- Anders, B. A. (2023). Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking? *Patterns, 4*(3), 100694.
- Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT generated literature review: Digital twin in healthcare. In Ö. Aydın (Ed.), *Emerging Computer Technologies 2* (pp. 22–31). İzmir Akademi Derneği.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). *On the opportunities and risks of foundation models*. Stanford Center for Research on Foundation Models (CRFM), Stanford University.

- Committee on Publication Ethics (COPE). (2023). *Authorship and AI tools: COPE position statement*. Committee on Publication Ethics.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642.
- Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature, 613*, 423.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681–694.
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine, 6*, 75.
- Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. *Accountability in Research*, 1–9.
- King, M. R. (2023). The future of AI in medicine: A perspective from a Chatbot. *Annals of Biomedical Engineering, 51*(2), 291–295.
- Kitamura, F. C. (2023). ChatGPT is shaping the future of medical writing but still requires human judgment. *Radiology, 307*(2), e230171.
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News, 40*(3), 26–29.
- Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care, 27*, 75.
- Stokel-Walker, C. (2023). ChatGPT listed as author on research papers. *Nature, 613*, 620–621.
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313.
- van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023).

Chapter 9

Technology, Society, and Digital Culture

Shaivi Singh , Student,

Jagannath Community College

Abstract

Technology is no longer something that exists separately from everyday life. It shapes how people communicate, learn, work, shop, create, entertain themselves, and understand the world around them. From smartphones and social media to artificial intelligence, cloud computing, digital payments, and immersive technologies, digital tools have gradually become part of the social fabric of modern life. This transformation has created new opportunities for education, economic growth, creativity, participation, and global communication. At the same time, it has introduced difficult questions about privacy, inequality, misinformation, identity, employment, mental well-being, cultural change, and the ethical use of technology.

This chapter explores the relationship between technology, society, and digital culture, with particular attention to the ways in which technology influences human behaviour and, in turn, is shaped by human values and social practices. It examines the emergence of digital culture, the changing nature of communication and relationships, the digital divide, the role of artificial intelligence, transformations in education and employment, online identity, privacy, misinformation, and the growing need for digital responsibility. Rather than viewing technology as either entirely beneficial or entirely harmful, the chapter presents it as a powerful social force whose impact depends largely on how people, institutions, businesses, and governments choose to design and use it.

Keywords

Technology, Society, Digital Culture, Artificial Intelligence, Social Media, Digital Divide, Digital Literacy, Privacy, Cybersecurity, Digital Identity, Ethics

1. Introduction

Imagine an ordinary morning in the life of a young person today. The alarm rings on a smartphone. The first few minutes may involve checking messages, scrolling through social media, reading the

news, or watching a short video. A digital payment might be used to buy breakfast. A map application may guide the journey to college or work. Classes may involve online resources, and assignments may be completed through cloud-based platforms. In the evening, entertainment might come through a streaming service, while communication with friends and family continues through messaging applications.

None of these activities necessarily feels extraordinary anymore.

That is perhaps the most significant feature of modern technology: it has become ordinary.

Technology is no longer simply a collection of machines, software, or devices. It has become embedded in everyday routines and social relationships. The smartphone in a person's hand can simultaneously function as a telephone, camera, newspaper, bank, classroom, entertainment system, workplace, diary, and social space. As a result, understanding technology today requires more than understanding how a device works. It requires understanding what technology does to people, what people do with technology, and how both influence society.

The growth of digital technologies has also changed the meaning of culture. Traditionally, culture developed through language, traditions, family, religion, art, literature, music, institutions, and shared social experiences. These influences continue to matter, but digital platforms have added a new layer to cultural life. Memes, viral videos, hashtags, influencers, podcasts, gaming communities, digital art, and user-generated content have become part of how people express themselves and participate in culture. This new environment can be described as digital culture.

Digital culture is not simply culture that happens on the internet. It is the broader transformation of human behaviour, communication, values, creativity, and social interaction brought about by digital technologies. The relationship between technology and society is therefore two-way. Technology changes society, but society also determines which technologies are created, adopted, regulated, and rejected. A technology reflects the priorities of the people and institutions that design it. The same technology can produce very different outcomes depending on the social environment in which it is used. The challenge of the digital age, therefore, is not to decide whether technology is “good” or “bad.” The more important question is: How can technology be developed and used in ways that strengthen human life rather than weaken it?

2. From Digital Revolution to Digital Culture

The digital revolution did not happen overnight. It developed through several stages, beginning with the transition from analogue systems to digital forms of storing and processing information. The

growth of personal computers, the internet, mobile phones, and eventually smartphones transformed digital technology from a specialised tool into a mass phenomenon. The arrival of the internet was particularly significant because it changed communication from a largely one-directional process into an interactive one. Earlier forms of mass communication such as newspapers, radio, and television generally involved a relatively small number of organisations producing content for large audiences. The internet gradually blurred this distinction. Today, an individual with a smartphone can be a consumer, producer, commentator, photographer, broadcaster, entrepreneur, or journalist.

This shift has democratised communication in important ways. People can share stories that may never have reached traditional media organisations. Small businesses can reach customers without establishing physical stores. Independent artists can find audiences across national borders. Students can access educational resources from institutions around the world.However, greater participation also creates new problems. When everyone can publish information, distinguishing reliable information from misleading information becomes more difficult. The ability to communicate widely does not automatically mean that the information being communicated is accurate. This is one of the central contradictions of digital culture: technology has made communication easier, but not necessarily communication wiser.

3. Technology and the Transformation of Human Communication

Communication is one of the areas most dramatically transformed by digital technology.

A message can now travel across continents within seconds. Video calls allow families separated by distance to see one another. Online communities allow people with similar interests to connect even when they have never met physically. Social media platforms have created spaces where personal experiences, political opinions, entertainment, news, and advertising exist alongside one another.

Yet faster communication does not always mean deeper communication. Digital communication often rewards speed. People may react to a headline before reading the article, respond to a message before understanding its context, or share a post before checking whether it is true. Short-form content can make complex issues appear simple, while algorithms may repeatedly present users with content similar to what they have already viewed.

As a result, communication has become increasingly shaped by attention. In the digital economy, attention has value. Platforms compete for the user's time, and content creators compete for views, likes, comments, shares, and subscribers. This can encourage creativity, but it can also encourage sensationalism. Content designed to provoke an emotional reaction may perform better than content

that requires careful thought. At the same time, digital communication has created new forms of expression. Emojis, GIFs, memes, reaction videos, voice notes, and short videos have developed into forms of digital language. Young people in particular often move naturally between text, image, video, audio, and symbols while communicating. Language itself is evolving. Digital culture has created new slang, abbreviations, expressions, and patterns of humour. A meme can communicate an idea or emotion that would require several sentences in traditional writing. These changes demonstrate that technology does not simply provide new communication tools; it also influences the way people think about communication.

4. Social Media and the Construction of Digital Identity

One of the most personal aspects of digital culture is the creation of an online identity.

People have always presented different versions of themselves in different social situations. However, social media has made this process more visible and continuous. A person's profile can contain photographs, opinions, achievements, friendships, interests, professional experiences, and personal moments. Over time, these elements form a digital representation of the individual.

This creates both freedom and pressure. On one hand, social media allows people to express aspects of themselves that may not find space in their immediate surroundings. Someone living in a small town can find a global community interested in their art, music, language, identity, or ideas.

On the other hand, the pressure to appear successful, attractive, happy, or interesting can become exhausting. Social media often shows carefully selected moments rather than ordinary reality. Comparing one's everyday life with another person's carefully edited online presence can create unrealistic expectations. Digital identity also has a lasting quality. A photograph, comment, or video posted casually may remain accessible long after the person has forgotten about it. This makes digital literacy particularly important. People need to understand that online actions can have consequences beyond the moment in which they occur. The question of identity has become even more complicated with artificial intelligence. AI-generated images, synthetic voices, virtual influencers, and realistic digital avatars challenge traditional ideas about authenticity. In the future, determining whether something is genuinely created or experienced by a human may become increasingly difficult.

5. The Digital Divide: Access Is Not Equal

The digital revolution is often described as universal, but access to technology remains unequal.

The International Telecommunication Union reported that 5.5 billion people were online in 2024, representing 68 percent of the world's population, while approximately 2.6 billion people remained offline. These numbers reveal an important reality: technological progress does not automatically create equal opportunity. The digital divide refers to differences in access to digital technologies and meaningful digital participation. These differences may be caused by income, geography, education, gender, age, disability, language, or affordability. For example, having access to a smartphone is not necessarily the same as having meaningful digital access. A student may own a phone but lack reliable high-speed internet, a suitable device for academic work, a quiet environment for online classes, or the digital skills required to use educational platforms effectively.

The divide therefore has several dimensions: access, affordability, skills, quality, and meaningful participation. Bridging the digital divide is therefore not simply about providing internet connections. It is about ensuring that people have the knowledge, resources, and opportunities necessary to participate meaningfully in digital society.

6. Artificial Intelligence and the Changing Relationship Between Humans and Technology

Artificial intelligence has introduced a new stage in the relationship between technology and society. Earlier digital technologies primarily helped humans store, retrieve, process, and communicate information. AI systems can now perform tasks associated with activities such as language generation, image creation, pattern recognition, prediction, translation, recommendation, and decision support. This creates enormous possibilities. In education, AI can support personalised learning and help students explore difficult concepts. In healthcare, AI can assist with analysis and research. In business, it can automate repetitive tasks and support decision-making. In creative industries, it can become a tool for brainstorming, editing, design, and experimentation.

However, AI also raises serious questions. What happens when an automated system makes a mistake? Who is responsible for an unfair decision made by an algorithm? How should personal data be protected? What happens to jobs when certain tasks become automated? How can society distinguish AI-generated content from human-created content?These are not purely technical questions. They are social and ethical questions.

UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted in 2021, places human rights and human dignity at the centre of AI governance and emphasises principles including transparency, fairness, accountability, and human oversight.

The most important lesson is that AI should not be viewed simply as a replacement for human intelligence. Its most constructive role may be as a tool that expands human capabilities while leaving important decisions subject to appropriate human judgment. The future will therefore require AI literacy, not merely technical AI skills. People need to understand what AI can do, what it cannot do, how its outputs can be wrong, and how to use it responsibly.

7. Technology and Education

Education has changed significantly because of digital technology.

A student no longer depends entirely on a physical classroom or textbook for knowledge. Online lectures, digital libraries, educational applications, virtual classrooms, open courses, simulations, and AI-based learning tools have expanded access to information.

Technology can make learning more flexible and interactive. Students can pause a lecture, revisit difficult concepts, use visual explanations, collaborate with classmates online, or explore subjects beyond their formal curriculum.However, access to information is not the same as understanding. The internet contains an enormous amount of information, but information without evaluation can create confusion. Students therefore need critical thinking alongside digital access. They must learn how to identify reliable sources, compare perspectives, recognise misinformation, understand algorithms, protect personal data, and use AI tools ethically. The role of teachers is changing as well. Technology does not eliminate the need for teachers; instead, it can change what teachers focus on. Rather than simply delivering information, educators increasingly have an opportunity to guide students in interpreting, questioning, applying, and creating knowledge. The classroom of the future may therefore be less about memorising information and more about learning how to navigate an environment where information is constantly available.

8. Technology, Work, and the Future of Employment

The workplace has also been transformed by digital technology. Remote work, digital collaboration platforms, cloud computing, online recruitment, automation, data analytics, and AI-assisted tools have changed how organisations operate. A person can work with colleagues located in different cities or countries without sharing the same physical office.

This flexibility can create opportunities, particularly for people who might otherwise face geographical or physical barriers to employment. At the same time, automation has created anxiety about job displacement. History shows that technological change does not simply eliminate work; it

often changes the nature of work. Certain tasks disappear, while new occupations emerge. However, the transition can be difficult for individuals whose existing skills no longer match labour-market demands. This makes lifelong learning increasingly important. A qualification obtained at the beginning of a career may not remain sufficient for an entire working life. Workers may need to continuously develop digital, analytical, creative, interpersonal, and problem-solving skills. The future of employment should therefore not be understood only as a competition between humans and machines. It is more accurately a question of how humans and machines will work together.

Skills that involve empathy, judgement, creativity, leadership, communication, ethical reasoning, and complex problem-solving are likely to remain particularly valuable because they involve dimensions of human experience that cannot be reduced easily to automated processes.

9. Privacy, Data, and the Invisible Digital Footprint

Everyday digital activity generates data. Searches, purchases, locations, photographs, browsing behaviour, interactions, and preferences can all contribute to a person's digital footprint. Data can make services more convenient and personalised, but it can also create risks when collected, stored, shared, or used without sufficient understanding or consent. Privacy is therefore no longer simply about keeping secrets. It is about maintaining a degree of control over personal information. Many people accept digital permissions without reading them because the process is complicated or because refusing access may make a service difficult to use. This creates a gap between formal consent and meaningful understanding.n Digital privacy therefore requires responsibility from individuals, companies, and governments. Individuals need stronger digital awareness. Companies need responsible data practices. Governments need effective frameworks that protect rights without unnecessarily restricting innovation. Privacy should not be treated as an obstacle to technological progress. It should be understood as part of responsible technological progress.

10. Misinformation, Deepfakes, and the Crisis of Trust

One of the greatest challenges of digital culture is the speed at which false or misleading information can spread. A traditional rumour might once have travelled through a neighbourhood. Today, a misleading post can reach thousands or millions of people within hours.

The problem has become more complex with AI-generated content. Deepfake videos, synthetic voices, fabricated images, and automatically generated text can create highly convincing representations of events or people that never actually occurred. This creates a crisis of trust.

If people cannot determine whether an image, video, or voice recording is authentic, they may become sceptical of everything. This can be damaging because a healthy society depends on a shared ability to establish basic facts. The solution is not to distrust every digital source. Instead, society needs stronger media and information literacy. Users should develop simple habits: check the original source; look for independent confirmation; examine the date and context; avoid sharing emotionally provocative content immediately; distinguish opinion from verified information; be cautious with images and videos that seem unusually dramatic; and use credible sources when the information has serious consequences. In the digital age, responsible communication is not only about having freedom of expression. It is also about understanding the consequences of what we choose to amplify.

11. Technology and Human Well-Being

Technology has made life more convenient, but convenience does not always equal well-being.

Constant connectivity can make people feel that they are never completely offline. Notifications, messages, work emails, social media updates, and breaking news can create an environment of continuous stimulation. There is also a psychological dimension to digital platforms. Likes, comments, views, and notifications can become measures of social approval. For some users, this may create pressure to remain constantly visible and active. The answer, however, is not necessarily to reject technology. Digital tools can also support well-being. People use technology to maintain relationships, find communities, access educational resources, express creativity, and obtain important information. The key issue is the quality and purpose of use.

A healthier digital culture requires intentional use. This may involve setting boundaries around screen time, turning off unnecessary notifications, taking periods away from social media, protecting sleep, and creating spaces where people can interact without constantly looking at screens. Technology should serve human life. Human life should not be organised entirely around the demands of technology.

12. Digital Culture and the Changing Meaning of Community

Digital platforms have transformed the concept of community. A community was once strongly associated with physical location: a neighbourhood, village, school, workplace, or religious institution. Today, communities can also be formed around interests and shared experiences rather than geography.

A person interested in a particular form of music, sport, literature, gaming, photography, filmmaking, or social cause can find people with similar interests anywhere in the world.

This can be deeply empowering. However, digital communities can also become isolated from opposing viewpoints. When users repeatedly encounter people who share their beliefs, algorithmic recommendation systems may reinforce existing preferences. Over time, this can contribute to polarisation. Digital communities therefore have two possibilities: they can broaden people's worlds or make those worlds narrower. The outcome depends partly on how consciously people engage with them. A healthy digital culture should encourage curiosity rather than constant agreement. Exposure to different perspectives can be uncomfortable, but it is essential for developing a mature understanding of society.

13. Technology and Cultural Diversity

Digital technology has created new opportunities for cultural preservation and expression.

Languages, folk music, traditional crafts, regional stories, historical photographs, and local traditions can now be documented and shared with global audiences. A young creator can introduce an international audience to a local tradition without waiting for traditional cultural institutions to provide that opportunity. At the same time, global digital platforms can create cultural pressure. Popular trends can spread rapidly across countries, sometimes overshadowing local languages, customs, or forms of expression. Digital culture is therefore both global and local.

It creates a shared global environment while allowing people to express highly specific local identities. The challenge is to ensure that global connectivity does not become cultural uniformity.

Technology should ideally make the cultural world larger, not smaller.

14. Ethics and Responsibility in the Digital Age

Technological development often moves faster than social institutions can respond.

A new application can be launched in months. A new AI capability can spread globally within weeks. Laws, educational systems, and social norms may take much longer to adapt. This creates an ethical gap. Responsible technology requires asking questions before problems become widespread: Is the technology safe? Who benefits from it? Who might be harmed? Whose data is being collected? Can the system discriminate against certain groups? Can users understand how important decisions are made? Who is accountable when something goes wrong? Is the technology environmentally sustainable?

These questions should not be left exclusively to technology companies. Governments, educators, researchers, journalists, designers, civil society, and ordinary users all have roles to play. Ethics should not be treated as something added after a technology is developed. Ethical thinking should be part of the design process itself.

15. Building a More Responsible Digital Society

A responsible digital society requires more than advanced technology. It requires capable and informed citizens. Digital literacy should therefore become a fundamental life skill.

A digitally literate person does not simply know how to operate an application. They understand how digital systems influence information, privacy, communication, opportunities, and decision-making. Schools and universities can play an important role by teaching students not only how to use technology but also how to question it.

Businesses can contribute by designing products that respect privacy, accessibility, safety, and human well-being. Governments can establish clear and balanced regulatory frameworks that protect citizens while allowing responsible innovation. Technology companies can adopt greater transparency and accountability. Individuals also have responsibilities. Every digital user contributes to digital culture through what they post, share, support, ignore, and create. A responsible digital society is therefore a shared project.

16. Conclusion

Technology has become one of the defining forces of contemporary society. It has changed how people communicate, learn, work, create, entertain themselves, participate in public life, and understand their identities. Digital culture has emerged from this transformation, creating new forms of expression and community while also producing new challenges.

The digital age has demonstrated that technological progress is never purely technological. Every innovation carries social consequences.

The internet can connect people and spread misinformation. Social media can create communities and intensify comparison. Artificial intelligence can expand human capabilities and introduce new ethical risks. Digital platforms can increase economic opportunity while also deepening existing inequalities. There is therefore no simple answer to whether technology is changing society for better or worse.

It is doing both. What matters is what society chooses to do with the power technology provides.

The future of digital culture should not be built around the idea that everything that can be automated should be automated, or that everything that can be connected should be connected. Instead, technology should be evaluated through a human question: Does this make human life better, fairer, safer, and more meaningful? The most important technology of the future may not be a machine at all. It may be the ability of people to use technology with judgement, empathy, responsibility, and awareness. Technology will continue to evolve. Society will continue to adapt. Digital culture will continue to change. But the human values guiding that change must remain at the centre

References

- International Telecommunication Union. (2024). Measuring Digital Development: Facts and Figures 2024.
- International Telecommunication Union. (2024). Facts and Figures 2024: Internet Use.
- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.
- World Bank. (2016). World Development Report 2016: Digital Dividends.
- World Bank Group. Digital and AI.
- UNICEF. (2025). Childhood in a Digital World.
- UNICEF. Keeping Children Safe Online.

Chapter 10

AI-Driven Consumer Behaviour: The Role of Data Analytics and Personalisation in Digital Business

Dr Sandhya Godara,

Assistant Professor , GJU

Abstract

Artificial intelligence is changing the way digital businesses comprehend, forecast, and influence consumer behaviour. With the growth of e-commerce, social media, apps, streaming, and digital payments, an unprecedented volume of consumer data has been generated. Artificial intelligence, machine learning, predictive analytics, natural language processing, and generative AI facilitate the ability of organisations to convert these data sources into actionable insights and increasingly personalised consumer experiences. Through the analysis of various data, businesses can make informed decisions. As a result, brands are increasingly turning away from mass marketing and are utilising more data-driven and personalised methods.

The chapter explores the role of AI, data analytics, personalisation, and consumer behaviour in digital business. The conceptual bases of AI-enabled consumer behaviour, data analytics for consumer understanding, and AI-enabled personalisation across digital marketing and customer experience management are covered in this course. The chapter also studies how personalisation affects consumer awareness, evaluation, purchase decisions, satisfaction, loyalty, and trust. Alongside this, the growing dependence on consumer data creates challenges related to privacy, algorithmic bias, transparency, surveillance, exploitation, and consumer autonomy. Newer developments such as generative AI, hyper-personalisation, AI shopping agents, predictive analytics, and emotion awareness technologies are expected to further enhance digital consumer experiences. The chapter argues that success in AI-driven consumer engagement will not just depend on technology but also on how these initiatives are governed, that is, their transparency, fairness, privacy protection, and human-centred implementation. The chapter describes AI-driven consumer behaviour as a continuous feedback loop, wherein consumer data helps to inform prediction and personalisation, while consumer responses generate new behaviour data.

Keywords:
Artificial Intelligence, consumer 132ersonal, data analytics, 132ersonalization, digital business, predictive analytics, consumer engagement, digital marketing, customer experience, generative AI, AI ethics.

1. Introduction

Consumer expectations have evolved due to the digital transformation of business in the cloud economy. A website consumers interact with today includes mobile apps, social media, digital marketplaces, search engines, video streaming, online banking, and connected devices. This information generates a huge amount of data on their likes and dislikes, buying actions and behaviours. Digital businesses can observe consumers constantly and analyse them at scale. This is different from a typical business environment in which organisations have access to limited customer information collected intermittently.

Artificial intelligence has become one of the main technologies that allow organisations to convert this growing amount of data into consumer intelligence. Artificial Intelligence (AI) refers to computer systems that mimic human intelligence to perform specific tasks. Machine learning involves the use of data to create a system that can learn, while natural language processing enables organisations to analyse and respond to human language. Generative AI has also automated the production of text, images, audio, video, and conversational responses.

First and foremost, the rise of data-driven personalisation led to the growing role of AI in consumer markets. In traditional marketing, the company segments customers in broad categories based on demographic or geographic characteristics. Digital technologies allow for far more granular approaches. Businesses can observe searches performed, products viewed, time spent browsing, purchases made, and replies to marketing messages. AI can then predict what the consumer is likely interested in based on these signals.

You can see this in recommendation systems. E-commerce sites suggest products you may like based on what you have browsed or bought. Similarly, streaming services recommend content that suits your viewing habits. Social media tailor the feed based on previous interactions, while

digital advertising tools use bulk assessment of data points to figure out what advertisement would lure the user (Davenport et al., 2020; Rajpoot et al., 2026).

The rising use of personalisation presents great opportunities for businesses and consumers alike. Enhancing engagement, conversion, retention, and customer value for businesses. Consumers can receive relevant information, save search costs, get a tailored service, and gain convenience.

Nonetheless, these technologies also pose significant ethical and social threats. Gathering and evaluating personal data may jeopardise privacy. Algorithms could replicate social biases. Customers may confuse why they are being shown specific recommendations or ads. Persuasive systems that are highly sophisticated could limit consumer freedom in various ways. So, the issue is not whether AI can predict consumer behaviour at all, but how it ought to be used.

This chapter explores AI-driven consumer behaviour through the connected lenses of data analytics, personalisation, digital business strategy, consumer psychology, and AI ethics. AI is changing consumer behaviour from a controllable output of marketing activity into a more active relationship where algorithms simultaneously observe, predict, and influence consumer choice.

Consequently, this chapter conceptualises AI-driven consumer behaviour not as linear but as a feedback relationship. The consumer interacts, data is generated, and AI drives consumer behaviour by predicting and personalising. As a result, consumers respond.

2. Artificial Intelligence and the Transformation of Consumer Behaviour

The process of consumer behaviour begins with need recognition and ends with post-purchase evaluation. According to classical models, consumer behaviour is influenced by psychology, culture, society, and economics. Digitalisation has played an important technological role in this.

In the online space, consumers are participating in algorithmically induced choices. Search engines rank information, e-commerce sites recommend products, streaming sites highlight

content, and social media tailor feeds. Consequently, AI does not just analyse consumer behaviour after it happens; it is increasingly reshaping the environment in which it happens. According to Davenport et al. (2020), AI is very significant in marketing, as it improves customer insights, personalisation, and marketing decision-making. Huang and Rust (2021) likewise identify AI as a strategic technology able to modify marketing functions by connecting together automation, intelligence, and interaction.

The linear sequence forms the basis of the more dynamic feedback loop presented in Figure 1. In the first stage, digital businesses will collect or receive the information generated by consumers. During the second phase, AI and analytics spot trends and forecast likely behaviour. In the third stage, organisations use those insights to tailor communication and personalisation of recommendations, offers, services and experiences.



Figure 1 illustrates a continuous cycle in which consumer interactions generate data, data areanalysed to identify patterns, AI models generate predictions, and these predictions inform personalised experiences. Consumer responses subsequently generate new behavioural data, allowing the cycle to continue.

It makes personalisation gradually sharper and more effective. Because of that, the relationship isn't one-directional; it is dynamic: consumers affect algorithms by way of their behaviour, while algorithms affect consumers on account of their experiences.

1. Data Analytics as the Foundation of AI-Driven Consumer Insights

AI-driven consumer applications are very effective, depending solely on the availability of data. The information used by digital businesses can come from a wide variety of sources, including transaction systems, websites, mobile applications, loyalty programmes, social media, customer service, reviews and ads.

1.1 Big Data and Consumer Intelligence

Big data is when we have high volume, velocity, and variety. Consumer data may comprise structured information like the value of transactions, product categories, dates of purchase, and demographic characteristics. It could even consist of customer reviews, social media comments, photographs, voice interactions, and video.

AI comes in handy in such settings, as traditional analytical methods may not effectively pick up the sign of complex relationships in large, varied datasets. To identify links and behavioural patterns, large datasets can be processed by machine-learning algorithms.

The goal of digital businesses is not just data but also consumer intelligence. Consumer intelligence is intelligence that an informed business can apply to consumer knowledge.

For instance, an e-retailer can examine historical purchases and browsing behaviour to identify which consumers are likely to purchase a new item. A streaming service can analyse how long users watch and what genres they're searching for and filling out.

1.2 Descriptive, Predictive, and Prescriptive Analytics

There are mainly three categories of data analytics: descriptive, predictive, and prescriptive.
Descriptive analytics studies what happened in the past. A business may look at its monthly sales, website visits, customer complaints, etc.

Predictive analytics examines what may happen. AI models can help predict customer behaviour, such as whether a customer will purchase a product, cancel a subscription, or come to a promotion.

Prescriptive analytics provides recommendations for actions that a company should take. A system is able to recommend which customers to target with an offer, which products to promote, or which channel to use.

Businesses are now proactively transitioning from historical to predictive and prescriptive analytics at all levels. Organisations are increasingly able to anticipate consumer behaviour rather than respond to it. As such, analytics is the process of moving from knowing past behaviour to predicting future behaviour and suggesting the appropriate action.

1.3 Customer Segmentation

Traditionally, demographic characteristics have been used for segmentation. AI allows for more active behavioural segmentation.

Based on various behavioural variables, consumers may be grouped in terms of their purchase frequency, spend, browsing activity, brand choice, engagement, discount sensitivity, etc.

For example, an e-commerce business might identify:

- high-value customers;
- price-sensitive customers;
- customers at risk of leaving; and
- customers likely to adopt a new product.

Such segmentation allows businesses to develop more relevant marketing strategies.

Nevertheless, using sensitive characteristics can also create problems with segmentation. Companies should make sure that their analytical practices do not discriminate.

1.4 Sentiment Analysis

Companies can use natural language processing to study customer reviews, comments, survey results, and social media discussions. Sentiment analysis can classify opinions and recognise

trends. With this capability, organisations can recognise the emergence of customer concerns arising from new products or campaigns.

However, emotional determination is not flawless. Misinterpretation occurs due to sarcasm, cultural differences, irony, multilingualism, and contextual meaning. Review by people is still important for impactful decisions.

2. AI-Powered Personalisation in Digital Business

Personalisation has emerged as one of the most prominent uses of AI. It is the modification of a consumer's experience as per personal preferences, behavioural traits, context, or predicted need. Personalisation now encompasses sophisticated systems that can adaptively change content, recommendations, offers, and interactions rather than relatively simple mechanisms like using the customer's name in an email.

2.1 Recommendation Systems

Recommendation systems are critical for many online businesses. They assist consumers in finding products or content that might interest them. There are many techniques for recommendation, which are collaborative filtering, content filtering, and hybrid filtering. Collaborative methods establish the relationship between user and products. A content-based approach uses similarities between items and consumer preferences. Part of the business value of recommendations is reducing information overload. In settings with millions of products or pieces of content, consumers cannot evaluate all alternatives. We can limit our selection through AI. However, recommendations can also affect consumer preferences by determining which alternatives are in view. This creates a form of algorithmic gatekeeping.

2.2 Personalised Advertising

Advertising is data-driven in today's world. AI technologies help in selecting adverts according to the predicted interests, behaviour, context and engagement possibility. The personalisation of advertising could make it more relevant to consumers' interests and past

behaviour. But very targeted marketing can also be unsettling. Advertising may frighten

consumers when the ads seem to reflect information they did not provide. It is the challenge of businesses to achieve relevance but not become intrusive in their surveillance

2.3 Personalised Offers and Dynamic Pricing

Businesses can use AI to know which kinds of consumers respond to which offers. Companies can offer loyalty rewards to customers who visit often or discounts to price-sensitive customers. Dynamic pricing systems can shift prices based on demand, inventory, timing, and other variables (Priester et al., 2020). These methods could increase efficiency, yet inconsistent treatment brings a question of fairness. If consumers receive substantially different prices or other benefits, businesses may need to determine whether the system may cause discrimination or undermine trust (Heidary & Pluut, 2025; Duani et al., 2024).

2.4 Generative AI and Personalised Content

The use of generative AI has enhanced personalisation. Businesses can now leverage the power of AI to generate mass customised content. Marketing messages can be tailored according to customer interests, past interactions, preferred language, and communication channels. It is possible to auto-generate or modify product descriptions, emails, ads, social media posts, and customer-service responses (Hermann & Puntoni, 2025; Kshetri et al., 2024). The benefits here are enormous. Nevertheless, the risks posed by large-scale automated content production include factual errors, inconsistencies, and a lack of authenticity. It is necessary to have human oversight to ensure that the content generated is aligned with the organisational values and provides accurate information (Hermann & Puntoni, 2025)

Figure 2. AI-Powered Personalisation in Digital Business

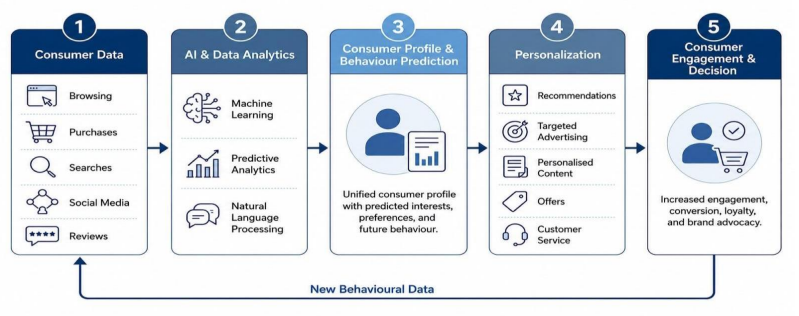


Figure 2: *The figure illustrates how consumer data are transformed through AI analytics and behavioural prediction into personalised recommendations, content, offers, and services.*

3. Influence of AI on Consumer Decision-Making

Section 4 looks at how AI helps with personalisation; here we examine how such personalised interventions may shape different stages of consumer decision-making. AI personalisation can impact almost every stage of the consumer decision-making process.

3.1 Need Recognition

AI can bring consumer needs to consciousness when products or services are presented at the right time. Personalised advertisements can help recognition of products you ordinarily wouldn't have considered.

This may help if the recommendation is relevant. Nonetheless, it also shows the convincing power of algorithmic systems.

3.2 Information Search

Digital consumers can often find themselves overwhelmed with information. This complexity can be reduced by using recommendation systems and AI assistants.

Instead of going through hundreds of alternatives, consumers may receive a small AI-chosen set. While this may enhance efficiency, consumers may become over-reliant on their choices.

3.3 Evaluation of Alternatives

Three AI processes help online shoppers: comparing merchandise, summarising reviews, and identifying features. Also, they can recommend alternatives. Soon, generative AI assistants will perform these tasks conversationally.

The consumer may thus search no longer for information but interact with the intelligent intermediary.

3.4 Purchase Decision

Personalised recommendations and offers, reminders, and other customised promotions influence the final purchase decision. AI can streamline checkout, answer queries, and forecast preferred payment or delivery options, thereby minimising friction in the process.

3.5 Post-Purchase Behaviour

After purchase, AI can assist with customer service, suggest complementary products, solicit feedback, and provide personalised loyalty rewards. When they are useful and not excessive, these interactions can lead to customer satisfaction and retention.

4 Consumer Engagement, Satisfaction, and Loyalty

Personalisation strengthens engagement and makes digital interactions more relevant, thereby reducing consumer search effort. Experiencing this may create satisfaction and a stronger relationship with the brands. However, just because something is personal doesn't mean I am a loyal consumer; I may switch if it is a competitor who provides me with greater and better value, privacy protection, or experience overall. Trust is thus an essential condition for lasting consumer relationships.

There should be a distinction made between value-based and surveillance-based personalisation. Consumers receive something useful in exchange for their data through value-based personalisation. Constant monitoring and expectation prediction but it does not consider consumers' expectations. Digital firms may need to prove that personalisation has real consumer benefits to stay competitive in the future. The distinction points out how personalisation's value

depends not only on predictive accuracy, but also whether consumers think the tradeoff of data and benefits is justified and proportional.

5 Ethical, Privacy and Social Challenges

The rise of consumer analytics through AI raises various ethical issues.

5.1 Privacy and Data Protection

Consumer data can reveal highly detailed information about individuals. Even apparently ordinary behavioural information can become sensitive when combined with other datasets. Organisations therefore need strong data governance, secure storage, appropriate access controls, clear retention policies, and responsible data-sharing practices (Cloarec, 2024; Yao et al., 2024). Privacy should be treated as an ongoing relationship between the organisation and consumer rather than simply as a checkbox during registration (Cloarec, 2024).

5.2 Lack of Transparency

Many consumers do not understand how recommendation and advertising systems operate. A consumer may see a recommendation without knowing whether it was based on previous purchases, browsing behaviour, demographic information, or inferred interests.

Greater transparency can improve consumer understanding and trust. Organisations should explain AI use in language that ordinary consumers can understand rather than relying entirely on complex legal documents.

5.3 Algorithmic Bias

AI often reflects the biases that exist in its data. The algorithms can reproduce inequalities if the data are historical. When making purchase recommendations, advertising exposure, credit assessments, pricing, or service provision, bias may creep in. Regular assessment, audit, surveillance, and remedy are necessary for responsibility in AI. As AI systems train on behavioural data, they can also reproduce bias or incompleteness if it occurs in follow-up

feedback loops.

5.4 Consumer Manipulation

When personalisation is designed to exploit psychological vulnerabilities, it can become an ethical problem. An AI system may distinguish what type of message would most likely influence a particular consumer. When used to maximise sales regardless of consumer welfare, personalisation can verge on manipulation. Businesses must therefore set boundaries with regard to persuasive AI. (Valenzuela et al. 2024, Spais et al. 2024).

5.5 Digital Inequality

Personalisation options being offered to consumers through AI can also help create digital inequality. Consumers with fewer digital skills may have less understanding of how algorithms steer their choices. Consumers need greater AI and digital literacy to enable them to make informed decisions in an increasingly algorithmic marketplace.

6 Responsible AI and Consumer-Centred Digital Business

Responsible AI requires organisations to consider the consequences of AI systems throughout their lifecycle.

Several principles are particularly important.

1. **Transparency:** Consumers should receive understandable information about significant AI-driven processes.
2. **Fairness:** AI systems should be tested for discriminatory outcomes.
3. **Privacy:** Organisations should collect and use data proportionately and responsibly.
4. **Accountability:** Businesses should identify who is responsible for AI-driven decisions.

5. **Human oversight:** High-impact decisions should not be left entirely to automated systems.
6. **Security:** Consumer information must be protected against unauthorised access and misuse.
7. **Consumer control:** Where possible, consumers should have meaningful choices regarding personalisation and data use.

These principles align with more general international developments in responsible AI governance. According to the OECD AI Principles (OECD, 2019), the focus is on values including inclusive growth, human-centred values, transparency, robustness, security, and accountability. UNESCO's Recommendation on the Ethics of Artificial Intelligence stresses the importance of human rights, transparency, fairness, and accountability (UNESCO, 2021).The EU's AI Act is another important initiative in risk-based AI governance and is particularly important in terms of transparency as well as the regulation of high-risk AI (European Union, 2024). As far as businesses are concerned, responsible AI should be built into strategy, not just a technical or legal fix.

7 Future Trends in AI-Driven Consumer Behaviour

7.1 Hyper-Personalisation

The upcoming phase of personalisation will probably feature experiences that are more real-time and contextual. Instead of relying solely on historical profiles, AI systems can include current behaviour and context. Experiences may adapt according to context, including location, device, time, prior experiences, and current needs.

7.2 Generative AI

We expect generative AI to proliferate further within marketing, customer service, commerce, and content production. Consumers can engage with businesses using conversational systems instead of standard web pages. Consumers need not search the product categories; they just describe their need and receive personalised suggestions.

7.3 AI Shopping Agents

AI agents may become important intermediaries between consumers and businesses. An agent could potentially search for products, compare prices, evaluate reviews, monitor availability, and assist with purchasing. This development may shift the consumer journey from direct platform interaction toward AI-mediated search, comparison, recommendation, and transaction. This could shift competitive advantage. Businesses may no longer compete only for consumer attention on websites and social media platforms; they may also need to ensure that their products are appropriately represented and discoverable by AI agents.

7.4 Predictive Customer Experience

Consumers will predict needs before consumers will demand them more and more from businesses. As an example, a digital service might identify that a customer is having difficulty and offer help. Systems like these can improve the quality of services but not be intrusive.

7.5 Emotion-Aware AI

Technologies that seek to determine emotions from voice and facial expressions, or text and behaviour, open up new avenues for customer service and marketing. Nonetheless, recognising emotion is a technical and ethically sensitive topic. An emotional state is complex and contextual, and it isn't always easy to infer.

7.6 Human–AI Collaboration

In the future, humans will likely not be replaced by technology like AI. Many digital business functions, instead, would be a collaboration of humans and intelligent systems. Through the rapid analysis and prediction of data by AI, humans can provide contextual judgement, empathy, creativity, and ethical reasoning. The growing role of AI assistants and agents indicates a possible shift from direct consumer–business interaction to multi-service AI-mediated interaction, as illustrated in Figure 3.

Figure 3. Future AI-Consumer Relationship in Digital Business

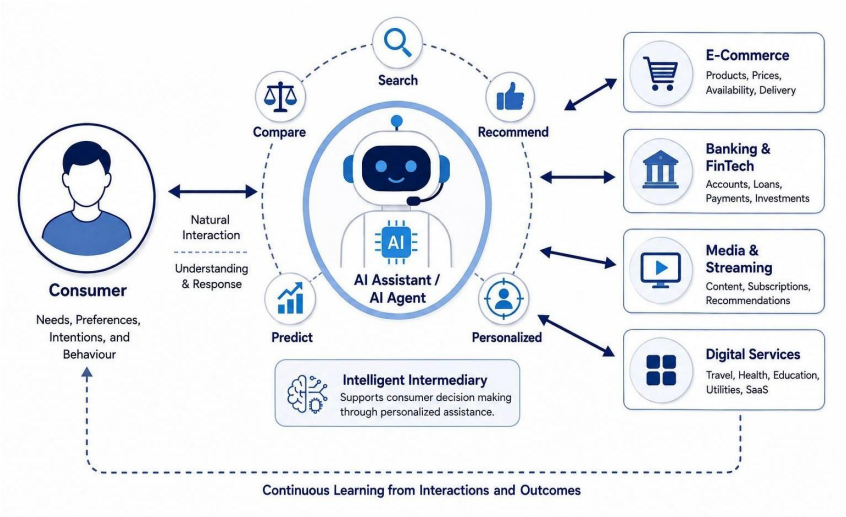


Figure 3: Future AI–Consumer Relationship—showing the emerging role of AI assistants/agents between consumers and multiple digital businesses.

8 Strategic Implications for Digital Businesses

First, organisations must treat data as a strategic asset while ensuring its quality and governance. Second, consumers should judge personalisation by value rather than conversion.Governing AI requires expertise from company departments like marketing, technology, legal, data science, and more.Investing in AI adoption means investing in training your employees. It refers to instilling AI literacy to understand its capabilities and limitations.

It’s important to continuously assess AI systems whenever consumer preferences and market conditions change.

9. Conclusion

Artificial intelligence reshapes consumer behaviour and digital business by transforming the way organisations collect, analyse, predict and respond to consumer information. Data analytics allows the identification of patterns from complex behavioural datasets, while AI helps in transforming them into predictions and personalised experiences.

The joint use of AI and personalisation has opened up vast opportunities. The consumer journey can be made more relevant and efficient while enhancing engagement and decision support. For businesses, these functions can help with customer engagement, marketing efficiency, retention and strategic decision-making.

Nonetheless, these advantages come along with substantial issues related to privacy, fairness, transparency, manipulation, and consumer autonomy. Even if an AI system is technically successful, it can still be socially or ethically problematic if it doesn't respect consumer rights or expectations.

The future will most probably be more personal with the help of generative AI, AI agents, predictive analytics, and context-aware systems. As time progresses, consumers will likely deal with AI instead of humans when conducting business. This transformation will generate new opportunities but will also require more robust governance and a more aware consumer.

There are three key developments. First, artificial intelligence is shifting consumer analyticsfrom retrospective observation towards predicting behaviour & influencing it. Furthermore, personalisation establishes a feedback loop whereby consumer reactions yield data that can be used for subsequent predictions. AI assistants and agents may put AI in the middle of consumers and a wide range of digital services.

References

- Cloarec, J. (2024). Transformative privacy calculus: Conceptualising the personalisation–privacy paradox on social media. *Psychology & Marketing*. https://doi.org/10.1002/mar.21998
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*, 24–42. https://doi.org/10.1007/s11747-019-00696-0
- Duani, N., Barasch, A., & Morwitz, V. (2024). Demographic pricing in the digital age: Assessing fairness perceptions in algorithmic versus human-based price

discrimination. *Journal of the Association for Consumer Research*, *9*(3), 257–268. https://doi.org/10.1086/729440

- European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Official Journal of the European Union.
- Heidary, K., & Pluut, H. (2025). All is (not) fair in personalised pricing: Antecedents and outcomes of consumer fairness perceptions. *Journal of Revenue and Pricing Management*, *24*, 204–221. https://doi.org/10.1057/s41272-024-00509-2
- Hermann, E., & Puntoni, S. (2025). Generative AI in marketing and principles for ethical design and deployment. *Journal of Public Policy & Marketing*, *44*(3), 332–349. https://doi.org/10.1177/07439156241309874
- Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*, 30–50. https://doi.org/10.1007/s11747-020-00749-9
- Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research*, *58*(3), 263–267. https://doi.org/10.2501/JAR-2018-035
- Kshetri, N., Dwivedi, Y. K., Davenport, T. H., & Panteli, N. (2024). Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda.
- *International Journal of Information Management, 75*, 102716.https://doi.org/10.1016/j.ijinfomgt.2023.102716
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420
- OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD Publishing. https://oecd.ai/en/ai-principles
- Priester, A., Robbert, T., & Roth, S. (2020). A special price just for you: Effects of personalised dynamic pricing on consumer fairness perceptions. *Journal of Revenue and Pricing Management*, *19*, 99–112. https://doi.org/10.1057/s41272-019-00224-3
- Rajpoot, C. S., Tiwari, V., & Vishwakarma, S. K. (2026). *Emerging trends of recommender systems for e-commerce: A comprehensive review*. Discover Computing, 29, 63.
- Spais, G., Phau, I., & Jain, V. (2024). AI marketing and AI-based promotions impact consumer behaviour and the avoidance of consumer autonomy threats. *Journal of Consumer Behaviour*. https://doi.org/10.1002/cb.2248
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
- Valenzuela, A., Puntoni, S., Hoffman, D., Castelo, N., De Freitas, J., Dietvorst, B., Hildebrand, C., et al. (2024). How artificial intelligence constrains the human experience. *Journal of the Association for Consumer Research*, *9*(3), 241–256.

Chapter 11

AI-Driven Citizen Journalism: Transforming Political Mobilization in India

Kashish Sahni, Research Scholar,
Department of Mass Communication, Jagannath University, Delhi NCR
kashishsahni9@gmail.com

Abstract

The convergence of AI and citizen journalism today is creating and influencing news and information in many different ways and on all levels. This chapter introduces readers to the field of AI-driven citizen journalism and explains how, by means of various digital media, i.e., YouTube, Facebook, Instagram, WhatsApp, X, etc., citizens and users today can create, publish, and distribute news of all types. This, in turn, has been made possible by AI technologies and tools such as Natural Language Processing (NLP), Machine Learning (ML), Computer Vision (CV), automated translation, and fact-checking. These technologies enhance the reach and speed of information flow to citizens in real-time and enable multilingual communication.

This chapter also highlights how AI technology is enabling various forms of political mobilization and digital activism, thereby amplifying participatory democracy. The chapter also covers a host of current problems that have arisen from the increased use of AI, including the spread of misinformation, issues caused by disinformation and deepfakes, algorithmic bias, online privacy risks, and ethical concerns affecting the credibility of online information. With suitable governance, AI technologies can enhance democratic communication. It also highlights the need to build media literacy, adopt ethical journalism, and have a suitable regulatory framework.

The chapter concludes by stating that the 'Age of AI-driven citizen journalism' presents endless opportunities for increasing transparency and holding people accountable in order to boost citizen participation in democracy. This can, however, only be possible if we manage to strike the right balance between technological growth and ethics and, in the end, remain committed to upholding the values of a democratic society.

Keywords:

AI-driven Citizen Journalism, Political Mobilization, Democracy, Artificial Intelligence, India.

1. Introduction

The rapid progress of artificial intelligence (AI) has dramatically changed the global media environment, enabling new models of production, distribution, and consumption of news. Especially, AI-based news applications enable citizen journalism in a completely new dimension by empowering every single citizen to generate, distribute, and consume news within seconds, anywhere in the world. As the Internet and smartphones have become the dominant medium for information consumption in India over the last decade, news applications powered by AI have started to change the paradigm of news consumption. Therefore, it is high time to study the impact of AI-driven news applications and their role in political mobilization and democratization.

The term ‘citizen journalism’ refers to the active participation by citizens in the gathering, processing, and publishing of news and information through digital media. In contrast to ‘traditional journalism,' which is published by media organizations run by professionals, in ‘citizen journalism,’ non-professionals gather information and publish news and information through digital platforms. The integration of AI in the process of citizen journalism has, therefore, the potential to increase the speed of information, its reach, and its accessibility. It enables citizens to produce high-quality information and also translate information published in one language into languages known to people located in different parts of the country.

This chapter examines the role of AI-driven citizen journalism in transforming the ways of political mobilization and in the process of democratic communication in India. This chapter shall explore the impact of AI on news production and dissemination and how it is affecting the political activity of citizens and promoting civic engagement. This chapter shall further bring to notice the various challenges faced by AI-enabled journalism: ethical, legal, and technological. Challenges of misinformation, of algorithms biased towards certain viewpoints, of violation of the right to privacy, and of the regulatory challenges faced by media and information platforms in the digital era.

2. Conceptual Framework

2.1 Understanding AI-Driven Citizen Journalism

News created by individuals rather than by established media channels is referred to as ‘citizen journalism.' By utilizing the low-cost Internet and smartphones, citizens are today able to report

the news as it happens in real time and disseminate it to a local as well as a global audience.
This tends to give voice to the citizens and conveys alternative perspectives and local issues and information that get very limited attention from the mainstream media.
The integration of AI in citizen journalism has revolutionized the medium by enabling the use of a wide array of technologies. AI can aid in the creation of news, editing of news, verification of news, translation of news, distribution of news, and even in recommending news to users. It thus increases the capacity of the citizen journalist to create high-quality news and to disseminate it to a wider audience. Many applications currently exist that utilize AI to aid in the production of news by automatically producing captions for videos, transcribing audio to text, enhancing images and videos to improve their quality, detecting whether images or videos have been tampered with, and even translating news into languages that the original author is not fluent in. These enable citizens to produce high-quality content that is produced with limited technical expertise.

2.2 Evolution of Citizen Journalism in India

Over the years, Indian citizen journalism has seen rapid digital growth. Initially, there were only blogs, online forums, and individual websites that served as the principal source for news and information. Gradually, as mobile phones became affordable and internet reach expanded, millions of Indians could now record, upload, share videos and images, and publish news as and when they happen.
Social media platforms, with little or no editorial intervention, have greatly facilitated the news-gathering and news-dissemination process. At times of elections, during protests, natural disasters, and other crises, live videos, images, and first-person accounts of events covered by ordinary citizens are increasingly becoming the norm. Such instances, which initially gained popularity through social media, later got coverage in mainstream media outlets. Therefore, highlighting the booming influence of citizen-generated content.

2.3 AI Technologies Supporting Citizen Journalism

The broad and diverse field of artificial intelligence is increasingly being harnessed for the purposes of enhancing communication on the internet. There are many technologies that have become especially relevant for citizen journalism.

Natural Language Processing (NLP) is the set of AI algorithms and programs that enable computers to understand, process, and generate human languages and human-language-like texts. It has features like automated transcription, language translation, text summarization, and chatbot-based services. Therefore, NLP can be of tremendous value to many citizen journalists.

Machine learning algorithms enable systems to learn from large datasets and understand user behavior. Social media platforms, for instance, use machine learning algorithms to recommend content and to sort news in users' news feeds. The system also identifies trending topics on social media platforms, which in turn can help important stories reach a wider audience. However, there are also risks of user behavior being reinforced, thus creating information bubbles around users.

Computer vision technologies are a subset of AI that is used to analyze images and videos to understand what they are about. This is used to recognize objects in images, identify individuals in photographs or videos (with their consent), enhance low-quality images, and, most importantly, verify whether images and videos have been tampered with or not. Computer vision is thus being used extensively for visual reporting as well as to identify altered content.

Generative AI offers a wide array of potential applications, from assisting in creating the actual content of a report to completing research and analysis to generating key graphics and infographics. Once a tool has been selected for use in a particular news story, a journalist can save a considerable amount of time by using the generated material while still retaining full responsibility for the content of the final story in terms of accuracy, original reporting, and ethical reporting practices.

2.4 AI, Social Media, and Public Communication

The interplay between social media and AI has transformed public communication dramatically. Digital platforms, where users share information and opinions, are no longer passive mediums. They are now affected by AI-powered recommendation systems. These recommendation systems use various parameters, such as past behavior, interests, and location, to surface specific posts or information. They determine what users can see on their digital platforms. In this sense, social

media is no longer just about consumption of information; it affects how information is consumed.

2.5 Theoretical Perspectives

The study draws on several theories of communication to explore the interplay between technology, media, and democracy.

Public Sphere Theory: Developed by Jürgen Habermas, Public Sphere Theory looks at the role that technology, media, and democracy play. A democratic society is based on a public sphere where citizens hold discussions, exchange information and views on current affairs, and thus influence the decisions of the government and other public authorities. The space of the public sphere has expanded with the advent of digital media supported by AI. Thus, the quality of the public sphere is determined by the quality of deliberation or discussion that takes place within it. In order to have a healthy public sphere, it is necessary to have credible information, equal opportunities, and respect for other views.

Participatory communication theory looks at how people become active participants and producers of information as opposed to simple consumers of information through media. This is exemplified by AI-driven citizen journalism, where individuals are able to document events and contribute to the dissemination of information and sharing of knowledge and experiences with others through public discussion. The use of technology to facilitate participatory communication helps to reduce language, skill, and physical distance barriers.

Network Society Theory, proposed by Manuel Castells, explains how digital networks restructure social, political, and economic relationships. The rapid flow of information across interconnected platforms allows for greater mobilization of support and an effective impact on political processes. Furthermore, AI boosts this network by means of information discovery, information customization, and communication efficacy.

The above theories have been used to understand how AI-driven citizen journalism is changing political participation and the dynamics of communication in democracy. These information and communication technologies have brought about many opportunities as well as challenges like misinformation, algorithmic influence, unequal access, etc.

3. AI-Driven Citizen Journalism in India

3.1 Growth of AI-Driven Citizen Journalism

India has seen a massive digital revolution in the last decade. The proliferation of affordable mobiles, low-cost mobile data, and rising internet penetration has enabled the masses to access information in no time. This has been possible due to Digital India initiatives of the government. Consequently, there has been an exponential rise in the number of citizens who have donned the hats of citizen journalists and are producing and disseminating news and information on the internet.

AI has removed the technical hurdles that had prevented so many from becoming citizen journalists. First was the matter of capturing images or footage at a scene, which could then be uploaded onto social media platforms such as Facebook or Twitter. Enhanced images and better-quality videos, with added features such as subtitles or even translation into local languages and summaries of longer interviews, have all been made possible through the use of AI-powered tools to enhance basic reports and make them of a very high standard.

3.2 Major Digital Platforms Supporting Citizen Journalism

The growth of AI-driven citizen journalism in India is being greatly supported by various digital platforms. Social media is enabling various users to access information and to create and disseminate news and current affairs on their own. Hence, platforms like YouTube, Facebook, Instagram, WhatsApp, and X (formerly Twitter) are some of the popular digital platforms where AI is facilitating the growth of citizen journalism in India.

A very large number of independent journalists in India are producing and uploading content on their own **YouTub**e Channels. They create videos, such as interviews, documentaries, discussion panels, etc. The videos are then promoted through the AI-powered recommendation system on YouTube.

Facebook also turns out to be a powerful news-sharing platform; citizens can surface local civic issues and organize campaigns and public discussions on various public policy issues through local community groups. AI algorithms further aid in personalizing users' news feeds based on past user interactions with similar content.

In India, where most citizens use **WhatsApp** to communicate with each other daily, whether living in a city or village, the messaging app is used to forward local news, share election-related updates, circulate public service messages, and even pass on information about emergencies. While it facilitates quick information dissemination, the app can also spread the worst kind of misinformation in no time, and hence there is an urgent need for digital literacy and fact-checking.

X (formerly Twitter) is a well-known platform for real-time updates across the globe. Politicians, media houses, government organizations, social activists, and citizens use this platform for day-to-day updates on current events, new policy updates, and ongoing news around the globe. Topics tend to trend on this platform with the help of AI-powered algorithms that study user behavior and increase visibility on trends that users are talking about on a real-time basis.

3.3 Opportunities Created by AI

Artificial intelligence has created numerous opportunities for citizen journalism. Firstly, AI increases access to information by enabling the publication of content in numerous languages, including Indian languages. This helps in enhancing democratic participation and in reducing language barriers to information.

Second, AI also greatly increases the efficiency of making news. For example, by automatically transcribing audio and video files and by automatically recognizing speech and making summaries of content.

The role of fact-checking technologies in stopping the spread of misinformation is playing a vital role in society. Nowadays, AI can compare digital content with verified databases to check if there are any manipulated images or inconsistencies and to detect possible misleading information.

3.4 Challenges and Ethical Concerns

However, the use of AI in citizen journalism also presents a number of challenges. The biggest one is the spread of misinformation or even disinformation. So, while ‘fake news’ in the form of images and videos and even audio recordings is becoming more and more believable, it can cause irreparable harm to the media and even affect elections. So, in the end, they can erode people’s faith in the democratic system and how it works.

Deepfake technology is another growing issue and can be used to create fake audio or video content of public figures and spread false information. Media houses must verify facts before publishing, law enforcement agencies must track the source of such information, and social media platforms must remove the content as soon as possible. However, this will only work if citizens are vigilant and report such content immediately.

Algorithmic bias can also have a negative effect on online environments. For instance, AI systems are taught by large amounts of prior data, which can have social, cultural, or even political biases. This, in turn, can affect the way in which online content is recommended. Particular viewpoints could be given more prominence online than others. This in turn could intensify online ideologies.

The issue of privacy and data protection, therefore, is another area of concern in the context of the use of AI by digital platforms to collect and process personal data of citizens and to enable them to engage in citizen journalism. While it is necessary to have robust legal safeguards, it is equally important to ensure transparency and accountability on the part of institutions and bodies that would govern the use of such data to ensure that information is not misused in order to carry out surveillance of citizens.

The digital divide is increasingly becoming a cause of concern as the internet is spreading across the globe; inequalities between urban and rural residents as well as between different socioeconomic groups in terms of their ability to use a variety of different digital technologies are likely to create inequalities in participation by citizens.

3.5 Regulatory and Institutional Responses

India is now confronting the opportunities and risks of AI-driven digital communication. Policy and regulatory measures by government agencies, technology companies, media, civil society, and academic institutions and organizations to promote media literacy and responsible digital communication have been put in place. These include issues concerning digital platforms, the role of intermediaries, data, online safety, and the like. Independent fact-checking organizations have been established.

Innovative technology and advanced AI systems are being used to combat spam, misinformation, manipulated media, and harmful content on the Internet. Above all, there is also a need to strike a

balance between curbing such content and allowing freedom of expression as well. Excessive regulation can restrict public discourse, while lack of oversight can result in havoc caused by the spread of harmful content on the Internet. Therefore, government institutions, technology companies, media, academia, researchers, and citizens must work together for effective governance.

4. AI and Political Mobilization

Political mobilization is about encouraging citizens to participate in political activities, like voting, campaigning, protest, public discussions, and policy advocacy, etc. Traditionally, political mobilization used to take place in the form of public meetings, rallies, newspapers, radio, television, etc. However, with the growing digital technologies, there has been a sea change in the forms of political communication. AI, social media, and citizen journalism have provided a new platform for citizens, political parties, civil society organizations, and government institutions to communicate effectively with the public. In India, where digital media has become a part of the daily life of citizens, AI-driven citizen journalism is playing a significant role in shaping political participation and democratic engagement.

4.1 AI and Digital Election Campaigns

Election campaigns have adopted AI technologies at a large scale. Political parties are using AI-powered data analytics for campaign planning, as this helps in understanding voters, including analysis of demographics, online behavior, and public sentiments.

Another group involved in AI-powered citizen journalism is those who report on and document political campaigns and local political issues. Traditional media outlets focus mostly on national politicians and events. In contrast, citizen journalists focus on local politicians, on issues of local governance, on the state of local infrastructure, education, health, employment, the environment, etc. Thus, they put pressure on politicians to respond to the needs of local communities and to raise awareness of issues of local governance among the public.

Multilingual political communication also got support from AI. With hundreds of languages being spoken in India, political campaigns can sometimes become challenging due to language barriers. While AI-powered translation tools come into play by drafting speeches and campaign materials, they do ensure that these are translated into the relevant regional languages and can

reach the widest possible audience across the country.

4.2 AI and Digital Activism

Digital activism has become important for current political activities and movements in India. Online platforms are used as tools to promote social movements in order to spread awareness and get public support.

Hashtag campaigns have become very important for organizing public debates on social issues. These campaigns are reinforced by AI algorithms that identify trends and recommend them to users. Such mechanisms can foster greater civic participation and collective action. However, this also amplifies emotionally charged or even false information, especially if it has not been verified adequately.

4.3 AI, Public Opinion, and Political Communication

Information flow and the formation of public opinion are increasingly being affected by artificial intelligence (AI). For example, online news, videos, and political discussions are distributed through algorithms that recommend them to social media users. In this way, AI not only distributes information, but it also influences the development of opinions and attitudes.

Citizen journalists play a key role in the formation of public opinions by reporting on local issues in detail, interviewing citizens affected by current events, and more generally reporting issues. These reports can often be seen to complement mainstream journalism, as they offer coverage of diverse perspectives and draw the attention of government towards regional/local issues.

However, the personalization of content poses challenges; AI has the ability to put people in an "echo chamber" where they only view information that supports their interests and already held opinions, limiting their exposure to alternative viewpoints. This contributes to political polarization, thereby reinforcing ideological divisions. The need of the hour is to have access to balanced and credible content.

4.4 Challenges: Misinformation, Disinformation, and Deepfakes

Spread of misinformation and disinformation by AI-driven political communication. Here, misinformation refers to false information, which can be distributed accidentally or without the intention to deceive others. On the other hand, disinformation refers to the distribution of false information with the intention to deceive others and influence public opinion.

Generative AI can create text, videos, images, and recordings in seconds. Though it gives room for legitimate use in industry, it can also be misused to create fabricated content. This can range from deepfake images, videos, and audio recordings, such as a doctored video showing a politician saying something that they have never actually said. In the context of an election, false information can spread fast through social media before being verified by authorized sources. So, the key responsibility falls on citizen journalists to verify information, check the sources of information, and report responsibly. Independent fact-checking and AI-powered verification of information have become important in minimizing the spread of false information, but public awareness and media literacy remain paramount.

4.5 Strengthening Transparency and Accountability

AI-driven citizen journalism significantly contributes to promoting transparency and accountability by enabling and supporting citizens in documenting public events, government projects, election campaigns, and other public activities and thus holding public authorities accountable for their actions.

However, the information gathered by citizen journalists can be difficult to organize, to analyze, and to verify. This is where AI can play a big role. It can help identify patterns in public data and set up automated search systems for government information such as policy reports, official statements, and other documents. AI can also assist with transcription and translation, making it easier for citizens to access information that is published by governments in different languages.

Moreover, public bodies today are increasingly employing AI-powered chatbots in order to better communicate with the citizens and provide them with adequate information on public services and facilities. These systems contain information on public welfare schemes and health care services, and taxation, as well as on various administrative procedures. By means of these systems, citizens are able to obtain the required information in a faster and more efficient

manner. The work of the administrative staff is, in turn, greatly reduced. However, these technologies should not replace but rather complement the existing human services, especially in cases where complex public issues are involved.

5. Ethical Considerations

The use of AI in journalism and political communication more broadly raises a host of new ethical issues. Chief among these is the age-old ethical principle of accuracy. Whether a journalist or a citizen journalist, it is always important to verify content. Thus, AI-generated information must be checked before it is published, especially in the sphere of reporting sensitive political issues and in cases of public emergency.

Another ethical principle of importance to journalism is transparency, which requires the citizen journalist to state the use of any AI tools that they have used for generating, editing, or enhancing their content, thereby allowing the audience to evaluate the information for its accuracy. This also aids in promoting responsible use of emerging technologies.

Next comes fairness and inclusivity; they play a crucial role, as AI must aim to reduce, not to create or to reinforce, existing social inequalities. In designing AI systems as well as in regulating the role of technology, developers, technology companies, and policymakers should seek to regulate algorithmic bias and should ensure equal participation and opportunities on digital platforms in democratic public space.

Media literacy in the AI era is essential when consuming online information. It refers to the skills that people require to access, evaluate, and use information effectively in various forms of media and communications. Educational institutions, civil society, media, governments, and a host of other organizations have a collective role to play in equipping citizens with sufficient levels of media literacy so that they become effective users of information, whether they consume it, create it, or disseminate it.

6. Conclusion

Citizen journalism has been fundamentally transformed by the growing use of artificial intelligence (AI) to generate news and for verification purposes. Furthermore, with the large number of Indian citizens possessing Internet-enabled mobiles, the rapid use of social media has led to a sort of AI convergence, which has enhanced opportunities for participation and thus

strengthened democratic communication and facilitated access to information. This shows that now citizens are active participants in the news ecosystem rather than just passive consumers of news.

While there are several challenges posed by the AI-driven communication of citizen journalism, as of now the situation appears to be grave. The spread of misinformation, which can be as damaging as disinformation, and sensationalism that can be inadvertently amplified by AI-powered recommendation systems can have serious consequences for public opinion and even for democracy. Hence, technology alone is not enough for strengthening democracy. What is required is the application of ethics, accountability, and good governance, and above all, responsible use of technology.

Even greater things can be expected from technology as it advances to newer levels. The array of AI tools and applications that can be harnessed for journalism, for political communication, and for citizen participation is endless. Advanced multilingual platforms, automated fact-checking, real-time translation, and intelligent content moderation are just a few of the many applications that are already available. What is needed to make these abundant resources work is a collaborative effort on the part of all the stakeholders: government, media, technology, academia, and civil society. Together with citizens of the country, they must ensure that the powerful technology is not misused in any way and is harnessed for the greater good of society.

Therefore, AI-driven citizen journalism turns out to be a powerful instrument to enhance transparency and enable more active political participation. At the same time, it also possesses a responsibility, which must be exercised carefully with adequate regulation and ethical conduct in order to safeguard the integrity of democracy. As India is rapidly moving into an era of digital transformation, AI in citizen journalism must be used in a sensible manner to develop into a more informed, participative, and resilient democracy.

References

- Castells, M. (2010). *The rise of the network society* (2nd ed.). Wiley-Blackwell.
- Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT Press. (Original work published 1962)
- Kaplan, A. M., & Haenlein, M. (2019). Siri, Siri, in my hand, who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

- Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2024). *Reuters Institute Digital News Report 2024*. Reuters Institute for the Study of Journalism, University of Oxford.

- Nielsen, R. K., & Ganter, S. A. (2022). *The power of platforms: Shaping media and society*. Oxford University Press.

- Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. Penguin Press.

- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news." *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/

- World Economic Forum. (2024). *The Global Risks Report 2024* (19th ed.). World Economic Forum.

Chapter 12

EdTech Revolution in India: A Digital India Perspective towards Viksit Bharat 2047

Dr. Priya Agarwal,

Associate Professor, Institute of Innovation Technology and Management.

Dr. Manjari Agarwal

Assistant Professor, Samalkha Group of Institutions

Abstract

The EdTech revolution, which is closely related to the goals of the Digital India program, has emerged as a result of the digital technologies that have transformed India's educational system. This chapter examines the development, significance, and prospects of EdTech in India, especially with regard to realizing Viksit Bharat 2047. The study is conceptual in nature and is based on secondary data gathered from pertinent academic literature, government papers, and policy documents like the National Education Policy (NEP) 2020.

The chapter looks at how the rapid growth of online learning platforms like BYJU'S, Unacademy, SWAYAM, and DIKSHA has been made possible by factors like rising internet penetration, reasonably priced data services, and ubiquitous smartphone usage. It also emphasizes how the COVID-19 epidemic has accelerated the use of digital and hybrid learning approaches at all educational levels. Personalized, adaptable, and interactive learning experiences have also been made possible by the incorporation of cutting-edge technology like data analytics and artificial intelligence.

According to the report, EdTech has enormous potential to assist skill development, increase accessibility, and improve learning outcomes—all of which would contribute to a knowledge-driven economy. But issues like the digital divide, low digital literacy, gaps in infrastructure, and worries about data security and quality continue to be major obstacles.

The chapter concludes that although the EdTech revolution is a major force behind India's educational change, inclusive policies, better digital infrastructure, and cooperative efforts among stakeholders are necessary for its success. In order to use EdTech as a catalyst to achieve an egalitarian and internationally competitive education system by 2047, these elements must be strengthened.

Keywords:
EdTech, Digital India, Online Learning, NEP 2020, Digital Transformation, Viksit Bharat 2047

1. Introduction

The chapter concludes that although the EdTech revolution is a major force behind India's educational change, inclusive policies, better digital infrastructure, and stakeholder collaboration are necessary for its success. In order to use EdTech as a catalyst to create an equal and internationally competitive educational system by 2047, these elements must be strengthened.

India's digital journey underwent a sea change in 2015 with the introduction of the Digital India program. Digital India has been instrumental in bolstering digital infrastructure, improving connection, and fostering digital literacy throughout the nation with its objective of building a knowledge economy and society empowered by technology. These advancements have created a solid basis for the expansion of educational technology platforms, making high-quality education accessible to people from all socioeconomic and geographic backgrounds.

Education delivery and consumption have been completely transformed by the rise of EdTech platforms like BYJU'S, Unacademy, Vedantu, SWAYAM, and DIKSHA. These platforms provide individualized, adaptable, and accessible learning experiences that meet the demands of a wide range of learners. This change was further expedited by the COVID-19 epidemic, which led to the adoption of online and hybrid teaching methods by educational institutions nationwide, integrating digital learning into the educational system.

The National Education Policy (NEP) 2020 places a strong emphasis on integrating technology into education to enhance access, equity, and quality in line with this shift. It encourages skill-based education, virtual classrooms, and digital learning—all of which are critical for developing a workforce prepared for the future.

The EdTech revolution has enormous potential to create a knowledge-driven, internationally competitive country in the context of Viksit Bharat 2047. EdTech may make a substantial contribution to India's long-term development objectives by improving educational accessibility, encouraging skill development, and closing learning gaps.

Objectives of the Study

- To assess EdTech's expansion and advancement in India
- To evaluate how the Digital India program advances digital education

- to research how EdTech affects educational quality and accessibility
- To determine the obstacles to EdTech adoption in India
- To assess EdTech's role in realizing Viksit Bharat 2047's vision

2. Literature Review

In recent years, the quick development of educational technology in India has garnered a lot of interest from academics and policymakers. Numerous studies have demonstrated EdTech's revolutionary potential to enhance education's efficiency, quality, and accessibility.

EdTech technologies offer flexibility and individualized learning opportunities, particularly for students in remote places, according to Sharma's (2022) analysis of the impact of digital platforms in improving learning outcomes. In a similar vein, Kumar and Singh (2021) highlighted how the use of technology in education has lowered regional boundaries and made high-quality educational resources more widely accessible.

The availability of reasonably priced smartphones and rising internet penetration have caused the Indian EdTech market to grow exponentially, according to a KPMG report from 2020. The COVID-19 pandemic accelerated the adoption of online education at all levels, according to the research.

In order to increase access, equity, and quality in education, the National Education Policy (NEP) 2020 firmly supports the use of technology. To promote ongoing learning, it suggests creating digital infrastructure, online content, and virtual learning environments. Government programs like DIKSHA and SWAYAM have also been acknowledged as significant advancements in the democratization of education (Government of India, 2020).

Although EdTech has enormous promise, obstacles including the digital divide, a lack of digital literacy, and infrastructure constraints restrict its efficacy, especially in rural regions, according to a study by Joshi (2021). Additionally, academics like Mehta (2023) have expressed worries about data privacy and content quality.

Overall, the body of research indicates that EdTech is a potent catalyst for India's educational revolution. However, resolving important issues and guaranteeing equitable access to digital resources are essential to its success. Building on these observations, this chapter examines EdTech's function within the framework of Digital India and how it advances the goal of Viksit Bharat 2047.

3. Research Methodology

This study is based on secondary data and is conceptual and descriptive in nature.

- Study Type: Conceptual/Descriptive
- Source of Data: Secondary data

Data Sources:

- Government publications (Digital India, NEP 2020)
- Journals and research papers
- Reports from companies such as NASSCOM and KPMG
- Online databases and websites

In order to comprehend the development, effects, and difficulties of EdTech in India as well as its applicability to the country's development objectives, the study examines data and literature already in existence.

Your structure is already sound; all you need to do is turn each goal into concise, understandable academic material. You can either use or slightly alter this well-written main body section for your chapter:

4. EdTech Revolution in India

4.1 Growth of EdTech in India

Over the past ten years, India's EdTech industry has grown rapidly as a result of rising internet penetration, smartphone usage, and demand for high-quality education. Millions of students are switching to online learning platforms, making India one of the world's biggest EdTech markets. Adoption of digital learning has been further accelerated by the emergence of inexpensive data plans and government programs. Significant investments have also been drawn to EdTech businesses, making the industry extremely competitive and dynamic.

5. Role of Digital India Initiative

The Government of India's Digital India Initiative has been instrumental in changing the face of education. Its main goals are to guarantee access to online services, improve digital

infrastructure, and increase digital literacy. In order to give students all around the nation access to free education, platforms such as SWAYAM, DIKSHA, and e-Pathshala were established. Through the widespread availability of educational resources, this program has contributed to closing the gap between urban and rural education systems.

5.1 COVID-19's effects

The EdTech revolution in India was sparked by the COVID-19 outbreak. Online education became the only practical choice when schools and universities were shuttered. As a result, the adoption of EdTech platforms by educators, institutions, and students skyrocketed. Digital tests, webinars, and virtual classrooms have taken the place of traditional classrooms. The shift greatly increased the adoption of online education, despite obstacles like the digital gap and the absence of infrastructure in rural areas.

5.2 Important EdTech Platforms (SWAYAM, BYJU'S, etc.)

The expansion of digital education in India has been aided by a number of EdTech platforms. BYJU'S provides individualized learning opportunities and interactive educational materials. Top universities provide free online courses through SWAYAM. Vedantu provides live online tutoring, while Unacademy concentrates on preparing students for competitive exams. These platforms make learning more interesting and successful by utilizing cutting-edge teaching techniques like videos, tests, and real-time interaction.

5.3 AI's and technology's roles

By offering individualized learning experiences, artificial intelligence (AI) and cutting-edge technology are revolutionizing the EdTech industry. AI assists in evaluating student performance and suggests personalized learning strategies. Virtual reality (VR), data analytics, and machine learning are examples of technologies that improve learning outcomes and increase engagement. Additionally, chatbots and automated technologies let students find immediate answers to their questions, improving education

5.4 The Next Priority for Digital Learning

Digital technology integration in education is a key component of the National Education Policy

(NEP) 2020. It encourages the use of digital resources, virtual labs, and online platforms to improve learning results. NEP promotes the creation of multilingual e-content and digital infrastructure to guarantee accessibility and inclusivity for all students. This is in line with the expansion of EdTech in India.

5.5 Push for Online Learning

Online education is acknowledged by NEP 2020 as a crucial part of contemporary education. It encourages the growth of platforms like DIKSHA and SWAYAM to offer high-quality education to more people. Additionally, the strategy promotes blended learning, which combines online instruction with traditional classroom approaches. This method makes education more learner-centric by giving students greater flexibility and letting them study at their own speed.

5.6 Learning Based on Skills

NEP 2020 places a strong emphasis on employability and skill development. Early on, the policy encourages digital skills, critical thinking, problem-solving, and vocational training. EdTech platforms are essential for providing practical knowledge, certifications, and skill-based courses. In addition to preparing students for real-world issues, this transition from rote learning to skill-based education also connects education with industrial demands.

6. Connection to NEP 2020 (Viksit Bharat 2047)

6.1 Emphasis on Digital Learning

NEP 2020 emphasizes the use of digital technologies as a basis for education that is prepared for the future. To guarantee inclusive education, it supports digital libraries, virtual labs, e-learning platforms, and multilingual content. The goal of Viksit Bharat 2047, which emphasizes the importance of a technologically literate populace for the advancement of the country, is directly supported by this digital drive. India can develop a knowledge-driven economy by increasing digital literacy and access.

For instance, platforms such as DIKSHA offer digital information and QR-coded textbooks to educators and students, particularly helping those in rural areas.

6.2 The Push for Online Education

Online and hybrid learning modalities are encouraged by NEP 2020. It encourages flexibility in education and supports Massive Open Online Courses (MOOCs) via platforms like SWAYAM. In order to achieve equitable growth by 2047, this democratizes access to high-quality education and lowers geographic barriers.

Case Study: Universities implemented hybrid learning approaches during and after COVID-19, fusing online lectures with in-person instruction. This approach keeps increasing access and boosting learning effectiveness.

6.3 Learning Based on Skills

Under NEP 2020, rote learning will give way to skill-based and competency-based education. The policy places a strong emphasis on industry-relevant skills, coding, entrepreneurship, and vocational training. EdTech platforms facilitate this by providing job-oriented programs, practical learning modules, and certification courses.

By producing a competent workforce that can propel innovation, entrepreneurship, and economic growth, this is consistent with Viksit Bharat 2047.
For instance, platforms like Coursera, Unacademy, and Skill India Digital improve young people's employability by providing courses in data analytics, artificial intelligence, communication skills, and entrepreneurship.

7. India's EdTech Challenges

7.1 The Digital Divide

The digital divide between urban and rural areas is one of the largest issues facing the EdTech industry. While many children in rural areas still lack smartphones and basic connectivity, urban kids have easy access to digital gadgets and high-speed internet. Inequality in access to online education results from this.

7.2 Insufficient Infrastructure

In many regions of India, inadequate digital infrastructure—such as poor internet connectivity, a lack of electricity, and restricted device access—hinders the expansion of EdTech. Remote educational institutions frequently find it difficult to successfully implement digital learning.

7.3 Digital Proficiency

A large percentage of educators and students lack the digital skills needed to use internet resources efficiently. This inhibits the uptake of EdTech solutions and diminishes their efficacy, particularly among underrepresented communities and older educators.

7.4 Quality Problems

Not every EdTech platform offers consistent or high-quality content. Learning results may be impacted by problems like excessive commercialization, variable course quality, and a lack of regulation. Maintaining integrity and quality is still a top priority. Maintaining quality and credibility is still a top priority.

8. Suggestions

8.1 Boost Internet Access in Rural Areas

The government ought to concentrate on developing broadband infrastructure in isolated and rural regions. To guarantee that everyone has access to dependable and reasonably priced internet, programs like BharatNet must be strengthened.

8.2 Training for Teachers

Teachers should receive adequate training in online teaching techniques and digital tools. Programs for ongoing professional development can assist teachers in successfully incorporating technology into their lessons.

8.3 Public-Private Collaborations

Government and private EdTech companies working together can enhance quality, accessibility, and innovation. Initiatives for digital education can be expanded nationwide with the aid of public-private partnerships.

8.4 Reasonably Priced EdTech Options

To promote inclusion, EdTech platforms should concentrate on offering free or inexpensive educational resources. Scholarships, discounted access, and freemium models can assist in reaching economically disadvantaged groups.

9. Conclusion

In India, the EdTech revolution has become a disruptive force in the education industry, changing the way that knowledge is acquired and given. Education is now more accessible, adaptable, and inclusive because to the quick development of digital technology, rising internet adoption, and ubiquitous smartphone use. Particularly in underprivileged areas, government programs like Digital India and platforms like SWAYAM, DIKSHA, and e-Pathshala have made a substantial contribution to closing educational disparities.

EdTech has become an essential component of the educational system as a result of the COVID-19 epidemic, which further drove the adoption of online learning. Furthermore, the use of cutting-edge technology like artificial intelligence, data analytics, and virtual learning tools has boosted student engagement and individualized learning experiences.

The emphasis on digital learning, online education, and skill-based development has been reinforced by EdTech's integration with the National Education Policy (NEP) 2020. This change is essential to building a workforce that can satisfy demands around the world in the future. To guarantee equal growth, however, issues like the digital gap, a lack of infrastructure, digital illiteracy, and quality issues must be resolved.

EdTech in India seems to have a bright future. With ongoing funding, legislative backing, and technical developments, the industry is anticipated to be crucial in changing the face of education. India can use EdTech to build a more competent, knowledge-driven, and inclusive society by tackling current issues and putting good plans into place.

In the end, the EdTech revolution is about empowering people, expanding opportunities, and laying a solid foundation for realizing the goal of Viksit Bharat 2047—a developed, independent, and internationally competitive nation—rather than just using technology in education.

References:

- Government of India. (2020). *National Education Policy 2020*. Ministry of Education.
- Ministry of Electronics and Information Technology. (2021). *Digital India Programme*. Government of India.
- NITI Aayog. (2022). *India's EdTech Landscape Report*. Government of India.
- KPMG. (2023). *Online Education in India: 2023 Report*.
- Deloitte. (2022). *Digital Transformation in Education Sector*.
- BYJU'S. (2024). *Annual Impact Report*.
- SWAYAM. (2024). *Official Platform Data and Reports*. Government of India.
- UNESCO. (2021). *Education in a Post-COVID World*.
- World Bank. (2022). *The State of Global Education: Crisis to Recovery*.

Chapter 13

Sustainable HRM, Green Marketing and Corporate Sustainability: An Integrated Business Perspective

Mr. Jwala Krishna Joshi
Assistant Professor, ITM SLS Baroda University, Gujarat
jkjoshi1985@gmail.com, +91 9998958322

Ms. Anuja Dayal
Assistant Professor, ITM SLS Baroda University, Gujarat
anujadayal25@gmail.com

Abstract

Corporate sustainability has evolved from a peripheral compliance concern into a central strategic imperative that shapes how organizations manage people, position products, and create long-term value. This chapter examines the interconnections between Sustainable Human Resource Management (SHRM), Green Marketing (GM), and Corporate Sustainability (CS), arguing that these three domains function most effectively when treated as an integrated system rather than isolated initiatives. Drawing on the triple bottom line framework and stakeholder theory, the chapter traces the conceptual evolution of each construct, explores the mechanisms through which they reinforce one another, and proposes an integrated business model that aligns internal human capital practices with external market positioning in service of holistic sustainability outcomes. The chapter concludes with a discussion of implementation challenges, organizational enablers, and future research directions.

Keywords: Sustainable HRM, Green Marketing, Corporate Sustainability, Triple Bottom Line, Green Human Resource Management, Stakeholder Theory

1. Introduction

The twenty-first century has witnessed an unprecedented convergence of environmental urgency, social accountability, and economic pressure on business organizations. Climate change, resource scarcity, regulatory tightening, and an increasingly conscious consumer base have collectively repositioned sustainability from a philanthropic afterthought to a core determinant of

competitive advantage. Within this landscape, two functional domains have emerged as particularly influential in operationalizing corporate sustainability commitments: Human Resource Management and Marketing.

Sustainable HRM refers to the adoption of HR policies and practices—recruitment, training, performance management, compensation, and employee engagement—that simultaneously support organizational performance, employee well-being, and broader ecological and social goals. Green Marketing, by contrast, refers to the strategic communication, positioning, and delivery of products and services in ways that minimize environmental harm while appealing to environmentally conscious stakeholders. Although these two functions address different audiences—employees on one hand, customers and markets on the other—they share a common foundation: both are expressions of an organization's underlying commitment to corporate sustainability, and both derive their credibility from consistency with that commitment.This chapter argues that businesses which treat SHRM and GM as disconnected departmental functions forfeit significant value, both in terms of internal coherence and external credibility. When employees are recruited, trained, and rewarded in ways that are misaligned with the environmental claims a firm makes in the marketplace, the result is often what scholars term "greenwashing"—a credibility gap between stated values and lived practice that damages brand trust once exposed. Conversely, when SHRM and GM are integrated around a shared sustainability architecture, organizations achieve what this chapter calls "sustainability coherence": an alignment of internal culture and external positioning that strengthens both employer brand and consumer brand simultaneously.

The chapter is organized as follows. Section 2 reviews the conceptual foundations of corporate sustainability. Section 3 examines Sustainable HRM as an internal enabler of sustainability. Section 4 examines Green Marketing as an external expression of sustainability. Section 5 develops an integrated framework connecting the two. Section 6 discusses implementation challenges and organizational enablers. Section 7 offers practical implications for managers, and Section 8 concludes with directions for future inquiry.

2. Conceptual Foundations of Corporate Sustainability

2.1 From Corporate Social Responsibility to Corporate Sustainability

Corporate sustainability (CS) traces its intellectual roots to the corporate social responsibility (CSR) movement of the mid-twentieth century, but it has since developed into a distinct and more integrative construct. Where CSR often emphasized discretionary social contributions layered atop core business operations, corporate sustainability treats environmental, social, and economic considerations as embedded dimensions of the business model itself. The widely cited "Triple Bottom Line" framework—People, Planet, Profit—captures this shift: sustainable organizations are those that pursue economic viability without compromising ecological integrity or social equity.

Corporate sustainability can be understood along three interdependent pillars:

- **Economic sustainability**: the capacity of the firm to generate long-term financial returns without depleting the resource base on which future returns depend.
- **Environmental sustainability**: the minimization of ecological footprint through resource efficiency, emissions reduction, waste minimization, and circular economy practices.
- **Social sustainability**: the fair and equitable treatment of employees, communities, and other stakeholders, including considerations of diversity, labor rights, and community welfare.

2.2 Stakeholder Theory as an Organizing Lens

Stakeholder theory provides a useful lens for understanding why sustainability cannot be confined to a single department. It posits that firms are accountable not only to shareholders but to a broader constellation of stakeholders—employees, customers, suppliers, regulators, communities, and future generations. Both HRM and Marketing are, in this sense, stakeholder-facing functions: HRM manages the employment relationship with one of the firm's most critical internal stakeholder groups, while Marketing manages relationships with external stakeholders, particularly customers. A stakeholder-theoretic view of corporate sustainability therefore naturally implies that internal (HR) and external (marketing) sustainability practices must be mutually reinforcing rather than independently designed.

2.3 The Legitimacy and Signaling Function of Sustainability

A further conceptual thread relevant to this chapter is legitimacy theory, which holds that organizations seek to operate within the norms and expectations of the societies in which they are embedded. Sustainability claims—whether directed at employees through HR policy or at consumers through marketing communication—function as legitimacy signals. However, signals are only credible when they are consistent across audiences. An organization that markets itself as environmentally responsible while failing to embed sustainability into its internal talent practices risks signal inconsistency, which stakeholders increasingly detect through social media scrutiny, employer review platforms, and investigative journalism.

3. Sustainable Human Resource Management: The Internal Enabler

3.1 Defining Sustainable HRM

Sustainable HRM (sometimes discussed under the closely related label of Green Human Resource Management, or GHRM) refers to the deliberate design of HR systems that generate positive outcomes across three simultaneous horizons: organizational performance, employee well-being, and environmental/social responsibility. Unlike traditional strategic HRM, which is primarily oriented toward maximizing organizational performance, Sustainable HRM explicitly incorporates ecological and ethical considerations into HR decision-making, recognizing that the long-term viability of the workforce is itself a sustainability concern.

3.2 Core Practice Areas

Several HR functional areas have been identified as key levers for embedding sustainability:

3.2.1 Green Recruitment and Selection. Organizations increasingly use employer branding to attract candidates who hold pro-environmental values, screen for sustainability competencies during selection, and highlight environmental credentials in job postings. This not only improves person-organization fit but also enhances the firm's reputation as an employer of choice among environmentally conscious talent, particularly younger workforce cohorts who report valuing employer sustainability commitments in job choice decisions.

3.2.2 Green Training and Development. Training programs that build employee capability in resource efficiency, waste reduction, and sustainable process design translate organizational sustainability strategy into operational competence. Beyond technical skill-building, such training also serves a normative function, socializing employees into a shared sustainability culture.

3.2.3 Green Performance Management and Appraisal. Incorporating environmental and social indicators into performance appraisal criteria signals that sustainability is not merely aspirational rhetoric but a measured and rewarded dimension of job performance. This might include targets related to energy conservation, waste reduction, or participation in sustainability initiatives.

3.2.4 Green Compensation and Reward Systems. Linking incentives—bonuses, recognition programs, promotion criteria—to sustainability outcomes reinforces desired behaviors and demonstrates organizational commitment beyond symbolic gestures.

3.2.5 Employee Involvement and Green Empowerment. Mechanisms such as green suggestion schemes, sustainability task forces, and employee-led environmental initiatives cultivate a sense of ownership and psychological engagement with organizational sustainability goals, converting employees from passive recipients of policy into active co-creators of sustainable practice.

3.3 Outcomes of Sustainable HRM

Empirical and conceptual literature converges on several outcome pathways. At the individual level, Sustainable HRM has been associated with enhanced employee engagement, organizational commitment, and pro-environmental behavior both inside and outside the workplace. At the organizational level, it contributes to reduced operational costs (through resource efficiency), enhanced employer brand equity, improved talent attraction and retention, and strengthened regulatory compliance. Importantly for this chapter, Sustainable HRM also has an external signaling function: employees who experience authentic sustainability practices become credible brand ambassadors, amplifying external sustainability messaging through word-of-mouth and social media, thereby directly supporting the firm's Green Marketing efforts.

4. Green Marketing: The External Expression

4.1 Defining Green Marketing

Green Marketing encompasses the range of activities designed to generate and facilitate exchanges intended to satisfy human needs and wants while minimizing detrimental impact on the natural environment. It extends beyond product formulation to encompass the entire marketing mix: product design (eco-labels, biodegradable materials, energy efficiency), pricing (reflecting true environmental costs or premium positioning for sustainable attributes), place/distribution (reducing logistics-related emissions, sustainable packaging), and promotion (communicating environmental credentials honestly and persuasively).

4.2 The Strategic Rationale for Green Marketing

Several converging forces drive firms toward green marketing strategies. Regulatory pressures increasingly mandate environmental disclosure and impose constraints on packaging, emissions, and product composition. Competitive pressures arise as rival firms adopt sustainability positioning, creating a need to maintain parity or differentiation. Perhaps most significantly, consumer demand has shifted substantially, with a growing segment of the market—often termed "green consumers" or "LOHAS" (Lifestyles of Health and Sustainability) segments—willing to pay premium prices for products perceived as environmentally responsible, provided the claims are perceived as credible.

4.3 The Greenwashing Risk

The central strategic risk in green marketing is greenwashing: the practice of conveying a false or exaggerated impression of environmental responsibility. Greenwashing can occur through vague claims, irrelevant claims, or claims unsupported by verifiable evidence. The reputational cost of exposed greenwashing can be severe, including consumer backlash, regulatory penalties, and long-term erosion of brand trust. This risk is precisely where the link to Sustainable HRM becomes strategically critical: green marketing claims are far less vulnerable to greenwashing accusations when they are substantiated by verifiable internal practices—including how the organization treats its employees and embeds sustainability into daily operations. A firm that

markets itself as a sustainability leader while offering no evidence of internal green practice is exposed to scrutiny; a firm whose external claims are visibly mirrored by internal HR practice possesses a much stronger legitimacy defense.

4.4 Green Marketing as Relationship Building

Beyond transactional exchange, contemporary green marketing theory increasingly frames sustainability communication as a relationship-building exercise. Consumers are treated not merely as purchasers but as stakeholders in a shared sustainability project, invited to participate through take-back programs, sustainability reporting transparency, co-created product development, and community environmental initiatives. This relational framing mirrors the employee-involvement emphasis found in Sustainable HRM, further underscoring the conceptual symmetry between the two functions.

5. Towards an Integrated Framework

5.1 The Logic of Integration

The preceding sections establish that Sustainable HRM and Green Marketing, while operating in different domains of stakeholder engagement, are structurally analogous and mutually dependent. Both are expressions of the same underlying construct—corporate sustainability—directed at different audiences. This chapter proposes an integrated framework built on three connecting mechanisms:

Mechanism 1: Internal-External Consistency (Authenticity Transfer). Sustainable HRM practices generate authentic internal capability and culture that substantiates external green marketing claims. When marketing communicates environmental commitment, internal HR practice provides the evidentiary basis that protects the claim from greenwashing accusations. This is an authenticity transfer mechanism: credibility generated internally through employee experience transfers outward through employee advocacy, transparent reporting, and consistent organizational behavior.

Mechanism 2: Employee-as-Brand-Ambassador (Internal Marketing). Employees who are recruited, trained, and rewarded around sustainability values naturally become carriers of the

sustainability brand narrative in their interactions with customers, suppliers, and the public. This "internal marketing" function means that Sustainable HRM does not merely support Green Marketing indirectly through legitimacy but actively participates in the delivery of the marketing message through frontline employee behavior, particularly in service industries where employee-customer interaction is a primary brand touchpoint.

Mechanism 3: Feedback Loop from Market to Workforce (Reverse Signaling). Consumer response to green marketing initiatives—whether positive reception or backlash—provides feedback that shapes internal HR priorities. Strong market reception of sustainability positioning reinforces organizational investment in Sustainable HRM (as sustainability is validated as commercially significant), while market skepticism or backlash can prompt firms to strengthen internal practice to close credibility gaps.

5.2 An Integrated Business Model

Synthesizing these mechanisms, an integrated model of corporate sustainability positions Sustainable HRM and Green Marketing not as parallel initiatives but as two interlocking subsystems within a single sustainability value chain:

1. **Strategic Sustainability Vision** — articulated at the top-management level, defining the organization's sustainability identity and commitments.
2. **Internal Translation (Sustainable HRM)** — the vision is operationalized through recruitment, training, performance management, and reward systems that build employee capability and commitment to sustainability.
3. **Cultural Embedding** — sustained HR practice produces a workforce culture in which sustainable behavior becomes normative rather than exceptional.
4. **External Translation (Green Marketing)** — the same sustainability vision is operationalized externally through product design, pricing, distribution, and promotional strategy.
5. **Stakeholder Perception and Legitimacy** — consistency between internal culture and external claims produces stakeholder trust, reducing greenwashing vulnerability and enhancing both employer and consumer brand equity.

6. **Performance Outcomes** — improved financial performance, reduced environmental footprint, enhanced social contribution, completing the triple bottom line.
7. **Feedback and Renewal** — outcomes and stakeholder feedback loop back into strategic vision, enabling continuous refinement.

This cyclical model emphasizes that corporate sustainability is not a static achievement but an ongoing organizational process requiring continuous alignment between internal and external practice.

5.3 Illustrative Organizational Patterns

While this chapter avoids attributing specific claims to named organizations, several illustrative patterns are commonly observed across industries pursuing integrated sustainability strategies. Firms in the apparel and consumer goods sectors frequently pair sustainable sourcing and circular-economy product lines (a green marketing initiative) with employee training programs on ethical sourcing and supply chain transparency (a Sustainable HRM initiative), enabling frontline staff to credibly answer customer questions about product provenance. Firms in the hospitality and food service sectors often combine sustainability-themed employee recognition programs with public-facing sustainability certifications, allowing customer-facing staff to serve as knowledgeable advocates for the brand's environmental commitments. Firms in the technology and manufacturing sectors frequently link internal green performance metrics for engineering and operations staff with public sustainability reporting, ensuring that externally reported environmental data is grounded in internally incentivized practice. These patterns, while illustrative rather than exhaustive, demonstrate the practical logic of the integrated framework proposed above.

6. Implementation Challenges and Organizational Enablers

6.1 Challenges

Despite its conceptual appeal, integrating Sustainable HRM and Green Marketing faces several practical obstacles.

Departmental Silos. HR and Marketing typically operate as distinct functional silos with separate reporting lines, budgets, and performance metrics, limiting the cross-functional coordination necessary for integration.

Measurement Difficulty. While marketing outcomes (sales, brand equity, market share) and HR outcomes (turnover, engagement scores) are separately measurable, the joint impact of integrated sustainability practice is harder to quantify, complicating the business case for integration to senior leadership.

Short-Termism. Both HR and marketing functions can face pressure to prioritize short-term performance metrics, which may conflict with the longer time horizons typically required for sustainability investments to yield measurable returns.

Resource Constraints. Particularly for small and medium enterprises, the financial and human capital investment required to build robust Sustainable HRM systems alongside sophisticated green marketing capability may exceed available resources.

Greenwashing Temptation. Where genuine internal transformation is costly or slow, organizations may be tempted to pursue green marketing claims without corresponding internal investment, undermining the integration logic and exposing the firm to reputational risk.

6.2 Organizational Enablers

Several organizational conditions facilitate successful integration. **Top management commitment** is foundational; sustainability integration requires visible and sustained leadership sponsorship that transcends departmental boundaries. **Cross-functional governance structures**, such as joint sustainability steering committees involving HR, Marketing, and Operations leadership, provide the coordination mechanism necessary to align internal and external practice.

Integrated sustainability metrics and reporting frameworks that track both internal (employee engagement with sustainability, training completion, green behavior metrics) and external (brand perception, green product sales, customer trust indicators) outcomes support evidence-based decision-making. Finally, **organizational culture** characterized by transparency,

ethical leadership, and long-term orientation provides the underlying soil in which integrated sustainability practice can take root and be sustained over time.

7. Managerial Implications

For practicing managers, this integrated perspective carries several practical implications. First, HR and Marketing leaders should establish regular cross-functional dialogue to ensure that internal talent practices and external brand claims remain mutually consistent, ideally formalized through joint sustainability committees or shared key performance indicators.

Second, organizations should resist the temptation to launch green marketing campaigns that outpace internal capability; sequencing matters, and internal credibility should generally precede or accompany external claims rather than follow them. Third, firms should invest in employee sustainability literacy not only as an operational necessity but as a marketing asset, recognizing that well-informed, values-aligned employees are among the most credible communicators of brand sustainability commitments. Fourth, sustainability reporting should be designed to capture and communicate both internal and external dimensions of practice, providing stakeholders with a holistic view that reduces the perceived risk of selective disclosure. Finally, leadership should treat sustainability integration as a long-term cultural transformation rather than a series of discrete campaigns, recognizing that the credibility benefits of integration compound over time as consistency is demonstrated across multiple stakeholder touchpoints.

8. Conclusion and Future Research Directions

This chapter has argued that Sustainable HRM and Green Marketing, though addressing different stakeholder audiences, are best understood as complementary and mutually reinforcing expressions of a single underlying corporate sustainability strategy. Treating them as integrated rather than isolated functions offers substantial benefits: it reduces greenwashing vulnerability, strengthens both employer and consumer brand equity, and creates a self-reinforcing cycle in which internal culture and external positioning continuously validate one another.

The proposed integrated framework—moving from strategic vision through internal translation, cultural embedding, external translation, stakeholder legitimacy, and performance outcomes—

offers both a conceptual map for scholars and a practical roadmap for managers seeking to embed sustainability more deeply and credibly into organizational life.

Future research would benefit from several avenues of inquiry. Longitudinal studies examining how the internal-external consistency mechanism unfolds over time, particularly in organizations transitioning from fragmented to integrated sustainability practice, would strengthen the empirical basis for the framework proposed here. Comparative research across industry sectors and organizational sizes could illuminate contingency factors that shape the feasibility and form of integration. Additionally, further work is needed on measurement instruments capable of capturing the joint or interactive effects of Sustainable HRM and Green Marketing on firm performance, moving beyond the currently siloed measurement approaches typical of each function. Finally, as artificial intelligence and digital platforms increasingly mediate both employee experience and consumer engagement, research into how digital transformation reshapes the mechanisms of sustainability integration represents a timely and important frontier.

Ultimately, corporate sustainability is not achieved through the sum of isolated departmental initiatives but through the coherence of organizational practice across every point of stakeholder contact. Sustainable HRM and Green Marketing, understood together rather than separately, offer a powerful lens through which this coherence can be pursued.

References

- Elkington, J. (1997). *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. Capstone Publishing.
- Dyllick, T., & Hockerts, K. (2002). Beyond the business case for corporate sustainability. *Business Strategy and the Environment*, 11(2), 130–141.
- Renwick, D. W. S., Redman, T., & Maguire, S. (2013). Green human resource management: A review and research agenda. *International Journal of Management Reviews*, 15(1), 1–14.
- Ehnert, I. (2009). *Sustainable Human Resource Management: A Conceptual and Exploratory Analysis from a Paradox Perspective*. Physica-Verlag.
- Peattie, K., & Crane, A. (2005). Green marketing: legend, myth, farce or prophesy? *Qualitative Market Research: An International Journal*, 8(4), 357–370.

- Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Pitman.
- Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64–87.
- Jabbour, C. J. C., & Santos, F. C. A. (2008). The central role of human resource management in the search for sustainable organizations. *The International Journal of Human Resource Management*, 19(12), 2133–2154.

Chapter 14

Generative AI in Journalism Education: Transforming Teaching, Learning, and Professional Skills in the Digital Age

Dr. Sheweta Gaur,
Associate Professor,
Lingaya's Lalita Devi Institute of Management and Sciences

Abstract

The rapid advancement of generative artificial intelligence (AI) is reshaping journalism education by influencing how students learn, educators teach, and media professionals create and distribute content. AI-powered writing assistants, research tools, transcription software, and content-generation platforms are increasingly part of journalism classrooms and newsrooms. These technologies improve learning efficiency, enhance creativity, support multilingual communication, and strengthen digital storytelling, while simultaneously raising concerns about academic integrity, misinformation, copyright, algorithmic bias, and the erosion of critical thinking and originality. This chapter examines the role of generative AI in journalism education, arguing that it should be treated as a complementary educational resource rather than a substitute for journalistic inquiry, creativity, and editorial responsibility. It proposes a balanced framework—RESPOND—for integrating generative AI into journalism curricula, faculty development, and institutional policy, so that graduates can navigate AI-enabled newsrooms while preserving accuracy, transparency, accountability, and public service.

Keywords: Generative Artificial Intelligence, Journalism Education, Media Literacy, Digital Journalism, AI Ethics, Newsroom Innovation, Curriculum Development

1. Introduction

Journalism education has long centred on developing the ability to gather information, assess credibility, construct meaningful narratives, and communicate them to the public. The emergence of generative artificial intelligence—systems capable of producing text, images, audio, and code in response to human instructions—represents a significant development in this trajectory, introducing new possibilities for journalism practice while raising questions about the

knowledge, skills, and ethical awareness journalism students need.

The relevance of generative AI extends beyond adding another digital tool to the classroom. Reporting, writing, editing, media law, ethics, and production remain the foundations of journalism training, but graduates now enter newsrooms that use AI-assisted systems for transcription, translation, data analysis, and other routine tasks. Journalism education must therefore integrate technological competence with the professional judgment required to evaluate information and make editorial decisions.

The educational value of generative AI is not simply speed or convenience; its potential lies in creating opportunities for experimentation, feedback, and alternative approaches to learning. Students can use AI-assisted tools to compare styles of news writing, examine report structure, practise editing, or develop ideas for multimedia storytelling. When guided appropriately, such activities encourage reflection on the choices involved in journalistic production rather than a narrow focus on the finished product—making the teacher's role, in connecting technology to learning objectives and professional standards, especially important.

At the same time, generative AI presents substantial challenges. AI systems can produce inaccurate statements, fabricated references, and biased content reflecting the limitations of their training data. Because journalistic credibility depends on the accuracy and verifiability of information, these limitations demand serious attention. The use of AI also raises concerns about copyright, privacy, disclosure, authorship, and academic integrity; if students become dependent on automated outputs without developing the capacity to question and verify them, the educational process risks weakening the very skills that responsible journalism requires.

The challenge, therefore, is not whether to accept or reject generative AI in education, but to identify the conditions under which it can be used meaningfully and responsibly. This requires a shift from tool-oriented learning towards critical, reflective, and professionally grounded AI literacy. This chapter brings together the pedagogical possibilities, professional implications, and ethical challenges of generative AI in journalism education, considers the competencies future journalists require, and proposes a balanced framework—placing human judgment, ethical responsibility, and practical learning at the centre of technological adoption—for integrating generative AI into journalism curricula.

2. Understanding Generative Ai in the Context of Journalism

Generative AI refers to systems designed to produce new content—text, images, audio, video, or code—in response to prompts, by identifying patterns from extensive training data rather than following predefined commands. The quality and reliability of these outputs depend on system design, training data, and user instructions. Crucially, an AI system's ability to generate grammatically coherent or visually convincing material does not mean that material is factually accurate, original, or fit for publication. For journalism education, understanding generative AI therefore requires recognising the distinction between content generation and knowledge production: AI may assist in presenting information, but it does not assume the responsibilities of a journalist, such as establishing source credibility, understanding the public interest, or accepting accountability for published content.

Journalism has long used technology—computer-assisted reporting, content-management systems, speech-to-text software, and algorithmic tools for defined tasks governed by fixed rules. Generative AI differs because it produces new language and media from learned patterns rather than executing predetermined instructions. This distinction matters pedagogically: automation and generation are related but not identical, and generative systems' capacity to produce plausible but incorrect information creates particular risks in a field where factual reliability and source accountability are fundamental.

Generative AI applications relevant to journalism education span several stages of practice. In news writing, AI can produce preliminary drafts, suggest headlines, and reorganise paragraphs, though AI-generated drafts must be checked against verified information and should never replace original reporting. In transcription and interview support, AI can convert recordings into text, but names, quotations, and context must be checked against the original recording. In translation and multilingual communication, AI can adapt content for diverse audiences, though students must evaluate cultural meaning and idiomatic accuracy. In research and data journalism, AI can help brainstorm questions and organise information, but students must consult original datasets and credible sources rather than treat AI explanations as evidence. In multimedia and digital storytelling, AI can support scripting, captioning, and production planning, provided

visual and audio material is checked for authenticity.

An important principle is that generative AI should be treated as an assistive system, not an autonomous source of journalistic authority. An AI tool may produce a well-structured report from a short instruction, yet that report may contain an inaccurate date, an invented quotation, or an unverifiable source—its fluency does not guarantee factual validity. Journalism students must learn to recognise that AI-generated content may be factually inaccurate, incomplete, biased, difficult to verify, insufficiently original, or ethically problematic if its synthetic nature is undisclosed. These limitations make AI literacy—the ability to understand how generative systems function, formulate appropriate prompts, critically assess outputs, and use the technology within ethical and professional boundaries—an essential component of contemporary journalism education.

The introduction of generative AI into journalism classrooms also changes the nature of learning activities. Students may increasingly be asked not only to produce a report but to explain how they researched it, which tools they used, how they verified information, and why they accepted or rejected AI suggestions—shifting emphasis toward process rather than only the finished product. For educators, the task is to design activities in which AI strengthens learning rather than encouraging dependence: for example, asking students to analyse an AI-generated draft, identify its factual weaknesses, and produce a corrected version, developing writing, editing, fact-checking, and ethical reasoning simultaneously.

3. The Changing Landscape of Journalism Education

Digital technologies have continuously reshaped how journalism is taught and practised, from desktop publishing and online platforms to mobile journalism, social media, and data-driven reporting. Generative AI represents a further stage in this evolution, but its significance extends beyond new software: it affects the competencies students need, the methods through which they learn, and the responsibilities institutions must address. Journalism education can no longer focus solely on producing technically competent content; it must also help students understand how information is generated, how its credibility can be assessed, and how technological systems may influence narratives and editorial decisions.

Traditional competencies—news gathering, interviewing, reporting, writing, editing, research,

source evaluation, media law, and ethics—remain the foundation of journalism training. However, contemporary media organisations increasingly expect graduates to be comfortable with AI-assisted transcription, translation, content organisation, and data analysis. Journalism programmes must therefore integrate technological fluency without diluting the core professional skills that give journalism its credibility.

Generative AI is also shifting emphasis from pure content production toward verification and editorial judgment. If AI can generate a draft in seconds, the educational and professional value increasingly lies in evaluating, correcting, and contextualising that draft rather than producing it from scratch. This has direct implications for how journalism educators design coursework: assignments can be restructured so that students critique, verify, and improve AI-assisted output rather than simply submitting it. Journalism educators themselves must adapt, moving from being sole transmitters of technical skill toward guiding students in the evaluation of AI-generated material, modelling ethical use, and connecting technological literacy to the enduring social purpose of journalism—accuracy, fairness, accountability, and public service. Preparing students for AI-assisted newsrooms means building both technical familiarity and the judgment to know when AI assistance is appropriate and when it is not.

4. Generative AI Across Journalism Teaching and Practice

Generative AI can meaningfully support several areas of journalism teaching when used with clear educational purpose.

In news writing, AI can generate draft structures and headline alternatives that students evaluate and revise, developing critical judgment about accuracy and tone rather than passive acceptance of machine output. In editing and language development, AI grammar and style tools can support clarity and consistency, but editorial judgment—determining what a story should say and how it should say it—remains a human responsibility. In research and information organisation, AI can help students organise large volumes of material and identify preliminary lines of enquiry, while original sources must still be consulted directly. Fact-checking education benefits from using AI to flag claims requiring verification, provided students treat AI output as a starting point for investigation rather than a conclusion. In data journalism, AI can assist with interpreting datasets and suggesting coding approaches, while data literacy and independent analytical reasoning

remain essential.

Generative AI also has applications in radio and audio production (script assistance and transcription), television and multimedia journalism (storyboarding, captioning, and production planning), and multilingual journalism (translation and localisation), always paired with human review for accuracy, tone, and cultural sensitivity. In the classroom, AI can support personalised learning by adapting explanations to a student's pace, and can enable new forms of classroom activity—such as comparing AI-generated and human-written reports—that make the distinctive value of human judgment visible to students.

These applications also require rethinking assessment. If students can generate a polished report using AI, assessment must evaluate the underlying journalistic competencies—research, verification, judgment, and ethical reasoning—rather than surface polish alone. Assignments that require students to document their process, justify editorial choices, and demonstrate verification are more resistant to unreflective AI use and better aligned with professional practice. Across all these applications, human judgment must remain at the centre of AI-assisted learning: AI can support drafting, editing, translation, and research, but the decision to publish, and responsibility for what is published, must remain with the student or professional journalist.

5. Curriculum Reform for an AI-Enabled Media Environment

The presence of generative AI in professional journalism creates a clear need for curriculum reform. Rather than treating AI as a subject confined to a single course, journalism programmes should integrate AI literacy into core subjects—reporting, editing, ethics, and media law—so that students encounter critical engagement with AI throughout their training rather than in isolation.

At the same time, dedicated modules on AI and journalism can provide deeper technical grounding: how generative systems work, their capabilities and limitations, prompt design, and emerging professional applications. Practical assignments should be redesigned so that AI use is transparent and pedagogically purposeful—for instance, requiring students to submit both an AI-assisted draft and a verified, edited final version with a reflective account of the changes made. Assessment practices must be adapted accordingly, with attention to academic integrity: institutions need clear policies distinguishing acceptable AI assistance (such as brainstorming or language support) from impermissible substitution of independent work, alongside plagiarism

and AI-detection tools used judiciously rather than punitively.

Curriculum reform should also target industry-relevant competencies identified by employers, including AI literacy, verification skills, and comfort working alongside automated systems, and should be paired with faculty development and institutional support so that educators themselves are equipped to teach and model responsible AI use. Throughout this reform process, curriculum innovation must preserve the human core of journalism—critical thinking, ethical reasoning, empathy, and public interest orientation—ensuring that efficiency gains do not come at the cost of the profession's foundational values. A balanced curriculum model treats generative AI as one competency among many, integrated with, rather than substituting for, traditional journalistic training.

6. Competencies for Future Journalists

The emergence of generative AI reshapes the competency profile expected of journalism graduates, though it does not replace the foundations of the discipline. Strong grounding in traditional journalism—reporting, interviewing, writing, and ethics—remains the base on which technological competence is built. To this, students must add AI literacy: a working understanding of what generative systems can and cannot do, and how to use them appropriately. Verification and fact-checking skills become more, not less, important as AI-generated content proliferates, since the ease of producing plausible text increases the risk of unverified claims entering the public record. Critical thinking and editorial judgment allow students to decide not just what AI can produce, but what should be published.

New competencies specific to this environment include prompting and AI-interaction skills—the ability to formulate instructions that produce useful, appropriately scoped output—alongside data and analytical skills for working with AI-assisted data journalism, and multimedia and digital storytelling skills for an increasingly visual and cross-platform media environment. Synthetic media and visual verification skills are increasingly essential, given the growing sophistication of AI-generated images, audio, and video and their potential to mislead audiences. Multilingual and cross-cultural communication skills help students use AI translation responsibly, while ethical reasoning and professional responsibility remain central to every application of the technology.

Students must also develop copyright, authorship, and intellectual-property awareness,

understanding how AI-generated or AI-assisted content intersects with existing legal and ethical norms, and adaptability and lifelong learning, since the pace of technological change means no single skill set will remain sufficient. Collaboration between humans and AI—knowing when to delegate a task and when to retain full control—becomes a professional skill in itself, as do communication, creativity, and originality, qualities that distinguish valuable journalism from generic, formulaic content. Together, these competencies describe an emerging professional profile: a journalist who is technologically capable but fundamentally oriented around human judgment, verification, and public accountability. Journalism programmes seeking to cultivate this profile should treat these competencies as interdependent rather than as a checklist to be covered in isolation. A student who can prompt an AI system effectively but cannot judge the reliability of its output has acquired only a partial skill; similarly, strong traditional reporting ability without any facility with AI-assisted tools may leave a graduate poorly prepared for newsroom environments that increasingly rely on such systems for routine tasks. The goal of competency development is therefore integration—embedding new technological fluencies within, rather than alongside, the profession's existing standards of accuracy, fairness, and accountability.

7. Ethical Challenges and Concerns

The integration of generative AI into journalism education raises significant ethical challenges that curricula must directly address. Accuracy remains the foremost concern: AI systems can generate plausible-sounding errors, and students must be trained to treat all AI output as provisional pending verification. Bias and representation are closely related concerns, since AI systems can reproduce or amplify stereotypes and imbalances present in their training data, with consequences for fairness and inclusion in news coverage. Misinformation and synthetic content—including AI-generated images, audio, and video capable of convincingly depicting events that never occurred—present a growing threat to public trust that journalism graduates must be equipped to identify and counter.

Within educational settings, academic integrity requires careful attention: institutions must distinguish legitimate AI-assisted learning from the submission of AI-generated work as independent effort, and must design assessments accordingly. Related to this are questions of originality and authorship, and of copyright and intellectual property, both of which remain

unsettled in law and practice as AI training data and outputs raise new legal questions. Privacy and the responsible use of information are also implicated, particularly where confidential sources, unpublished material, or sensitive personal data might be entered into AI systems without adequate safeguards.

Transparency and disclosure—informing audiences when AI has played a substantial role in producing content—are emerging as professional norms that journalism education should instil as standard practice rather than an afterthought. A further risk is the loss of human judgment through overdependence on automated systems, which curricula must actively guard against by keeping verification and editorial decision-making squarely with students and professionals. This connects to professional accountability: however AI is used, responsibility for published content cannot be delegated to a machine. Finally, educators must help students understand changing employment and professional roles in an AI-affected media industry, while never losing sight of the human dimension of journalism—empathy, contextual understanding, and public service—that no technology can replace. These ethical concerns are not merely abstract; they surface concretely in classroom exercises whenever a student submits an AI-drafted assignment without disclosure, uses a translation tool that subtly distorts a quoted source's meaning, or relies on an AI summary of a dataset without checking the underlying figures. Addressing such situations as they arise, rather than only through abstract policy statements, helps students internalise ethical reasoning as a practical habit rather than a set of rules imposed from outside.

8. The Role of Educators and Institutions

Meaningful integration of generative AI into journalism education depends substantially on the role played by faculty and institutions. Educators must increasingly act as guides rather than gatekeepers, helping students navigate AI tools critically rather than simply restricting or ignoring their use. This requires developing faculty AI literacy through structured professional development, so that instructors understand the same capabilities and limitations they are expected to teach.

Teaching methods themselves require redesign, incorporating AI-aware activities such as comparative analysis of AI-generated and human-written content, alongside clear classroom guidelines specifying acceptable and unacceptable uses of AI in coursework. Assessment must

be redesigned in parallel, emphasising process, verification, and reflective justification over the polish of a finished product alone.

At the institutional level, clear AI policies are needed to govern classroom use, protect student and institutional data from inappropriate exposure to third-party AI systems, and establish a broader AI-aware academic culture. Strengthening industry-academia collaboration helps ensure that curricula remain aligned with evolving newsroom practice, while establishing AI ethics committees or academic working groups provides institutional mechanisms for addressing emerging dilemmas as they arise. Encouraging responsible experimentation—rather than blanket prohibition—allows students to build practical familiarity under guidance, with faculty serving as ethical role models in their own use of AI tools. Across all these measures, institutions must maintain human-centred education as the organising principle, building a sustainable institutional framework that can adapt as the technology, and its professional applications, continue to evolve.

9. The Respond Framework: A Balanced Approach to AI in Journalism Education

To translate these principles into practice, this chapter proposes the RESPOND Framework—an author-developed model built around seven interconnected principles that distinguish responsible, human-centred journalism education from uncritical automated content production. The framework should not be understood as a rigid sequence in which one stage ends before the next begins; in practice, students move repeatedly between its elements.

Recognise AI capabilities and limitations. Students should become familiar with generative AI's applications—text generation, summarisation, transcription, translation, and data assistance—while understanding that technological capability does not guarantee factual reliability. Before using an AI tool, students should ask whether a task genuinely requires AI assistance and whether its use suits the specific journalistic purpose.

Evaluate information and evidence. AI-generated information should never be treated automatically as evidence. Students must learn to trace the origin of information, identify supporting sources, compare claims against primary material, and verify any statistic, quotation, or reference before incorporating it into journalistic work—connecting AI literacy directly to the traditional foundations of source evaluation and fact-checking.

Strengthen journalistic skills. AI integration should reinforce, not replace, fundamental abilities in reporting, interviewing, writing, editing, research, and multimedia production. AI may assist with transcription, but students should still listen to original recordings; AI may suggest a headline, but students must judge its accuracy. The framework places professional competence before technological convenience.

Protect ethics, privacy, and academic integrity. Responsible AI use requires attention to privacy, copyright, authorship, misinformation, bias, disclosure, and confidentiality. Sensitive or confidential material should not be entered into AI systems without regard for institutional and ethical requirements, and clear distinctions must be maintained between permitted assistance and impermissible substitution of independent academic work.

Own the editorial decision. Regardless of the extent of AI assistance, the student or journalist remains responsible for the final decision to publish. Students should be trained to ask not merely "Can AI produce this?" but "Should this be produced or published in this way?"—keeping editorial judgment connected to evidence, context, ethics, and the public interest.

Navigate technology through continuous learning. Because AI technology develops faster than any single curriculum can track, students and educators need transferable evaluative skills rather than familiarity with one platform alone. This includes ongoing faculty development in AI ethics, pedagogy, and applications.

Develop responsible, human-centred journalism. The framework ultimately returns to journalism's purpose: not to produce more content more quickly, but to provide accurate, contextualised, and socially responsible information, grounded in empathy, cultural understanding, and editorial independence.

This progression—Recognise, Evaluate, Strengthen, Protect, Own, Navigate, Develop—can be applied across journalism courses rather than confined to a dedicated AI module. In a news-writing class, students might generate alternative headlines with AI and then evaluate their accuracy; in radio journalism, compare an AI transcription against the original recording; in television and multimedia courses, investigate whether visual material is authentic or synthetic; in data journalism, verify AI-suggested analyses against the underlying dataset; and in media ethics, debate when audiences should be informed of AI's role in content production. The

RESPOND Framework is intended to become part of the everyday learning culture of journalism education, rather than an isolated technical exercise, moving students from technological awareness toward responsible professional practice.

10. Discussion

The integration of generative AI into journalism education raises questions that extend beyond any single tool or technique: what should journalism students learn, how should they learn it, how should their learning be assessed, and what does it mean to practise journalism responsibly in an increasingly automated information environment. The educational significance of generative AI lies not in its capacity to produce content, but in how it reshapes the relationship between information, technology, learning, and human judgment.

As AI increasingly handles first-draft production, education must shift emphasis from producing content to evaluating it—a shift with implications across curriculum design, teaching methods, and assessment. Efficiency gains from AI use must be weighed against educational value: a shortcut that saves time but bypasses the development of core skills undermines the purpose of journalism training. Verification becomes more important, not less, as the volume of plausible but unverified content grows, reinforcing rather than diminishing the centrality of fact-checking to journalism education.

The role of the teacher is correspondingly transformed, from primary source of technical instruction to guide in critical evaluation and ethical reasoning. Academic integrity requires a more nuanced approach than blanket prohibition, distinguishing between legitimate assistance and improper substitution. For professional preparation and employability, graduates who combine strong traditional skills with AI literacy and ethical judgement are best positioned for AI-affected newsrooms. In contexts such as Indian journalism education, these considerations intersect with rapid media growth, linguistic diversity, and expanding digital access, making responsible AI integration especially consequential for the profession's future development.

Across all these considerations, one principle recurs: human agency must remain at the centre of AI-supported journalism education. This does not mean rejecting technology, but ensuring that technology remains accountable to educational and journalistic purposes—a human-centred approach that the RESPOND Framework operationalises by connecting AI use to verification,

professional skill, ethics, accountability, and continuous learning.

11. Conclusion

Generative artificial intelligence is becoming an important part of the changing media environment, and its influence on journalism education is likely to continue expanding. Its capacity to generate text, process information, support translation, assist transcription, and facilitate multimedia content development creates genuine possibilities for teaching and professional preparation. Yet the educational significance of generative AI should not be measured by the speed or volume of content it can produce.

Journalism is ultimately a human-centred profession, requiring curiosity to ask questions, discipline to investigate, judgment to evaluate evidence, sensitivity to understand people, and responsibility to communicate accurately and fairly—qualities that cannot be reduced to the technical operation of an AI system. Integrating generative AI therefore demands a balanced approach: journalism programmes should give students opportunities to understand and experiment with AI while simultaneously strengthening traditional reporting, research, writing, editing, verification, and critical-thinking skills.

This chapter has argued that AI should occupy three complementary positions within journalism education: as a professional tool, as an object of critical study, and as an ethical challenge. Treating it only as a productivity tool overlooks its wider implications; treating it only as a threat prevents students from developing relevant professional competencies. The proposed RESPOND Framework offers one approach to this challenge, placing human judgment at the centre of technological integration through Recognition, Evaluation, Strengthening of journalistic skills, Protection of ethical and academic values, Ownership of editorial decisions, Navigation through continuous learning, and Development of responsible journalism.

The future journalist must therefore be technologically aware without becoming technologically dependent—knowing how to use AI, but also when to question it, verify it, reject it, or set it aside in favour of a human approach. Educators, correspondingly, should not attempt simply to prevent students from encountering AI, but should help them develop the intellectual maturity to use emerging technologies responsibly. The future of journalism education does not lie in teaching students to compete with artificial intelligence, but in preparing them to work with

technology while preserving the human qualities that give journalism its social value. Generative AI may change how journalism is produced, but the principles of accuracy, verification, accountability, fairness, public interest, and responsible communication must continue to guide why journalism is produced.

References

- Beckett, C. (2019). New powers, new responsibilities: A global survey of journalism and artificial intelligence. London School of Economics and Political Science, Polis.
- Beckett, C., & Yaseen, M. (2023). Generating change: A global survey of what news organisations are doing with AI. JournalismAI, Polis, London School of Economics and Political Science.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.
- Broussard, M. (2018). Artificial unintelligence: How computers misunderstand the world. MIT Press.
- Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. Digital Journalism, 3(3), 416–431.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239.
- Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press.
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. Digital Journalism, 5(7), 809–828.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., et al. (2023). ChatGPT for

good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

- Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO.
- Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO.

Chapter 15

From Gandhian Journalism to the Contemporary Indian Digital News Feed

Reading Gandhiji's Editorial Record Against the Evidence on Algorithmic News Curation in India

Pawan Raj

Abstract

Gandhiji ran two newspapers, *Young India* and *Navajivan*, without carrying advertisements, funding them through subscriptions instead. He wrote about that choice at length, framing it as a way of keeping the reader, rather than the advertiser, at the centre of editorial decisions. A century later, most Indian readers get their news through algorithmic aggregators, where placement follows engagement rather than subscription. This chapter reads Gandhiji's documented editorial record against the growing body of published research on how algorithmically curated news actually behaves, using gatekeeping theory as the shared analytical frame. Rather than building a new dataset, the chapter draws on existing empirical work: a comparative audit of human-and algorithm-curated sections within a major news platform, recent systematic reviews of algorithmic influence on journalism, and Indian-specific research on algorithmic gatekeeping during the 2024 general election. These findings are read against Gandhiji's own writings on journalism and the documented editorial practice of *Young India* and *Navajivan*, to ask what changes, and what does not, when a news outlet's revenue model shifts from a subscriber to an advertiser to an engagement-optimizing algorithm. The aim is not to argue that Gandhiji anticipated the algorithm, or that a subscription-funded press is inherently superior to an advertising or engagement-funded one. It is to make a narrower claim: that revenue structure is a legible variable in gatekeeping behaviour, visible across both records, and that India's own press history offers a documented case for studying it that the existing literature, still dominated by Western and presentist accounts, has not used. The chapter closes by setting out a coding framework through which this comparison could be tested directly against new primary data in future work.

Keywords: Gatekeeping Theory, Algorithmic Curation, Indian Digital Journalism, News Aggregators, Young India, Navajivan, Gandhiji, Revenue Models in Journalism

1. Introduction

In April 1919, Mohandas Karamchand Gandhi took editorial charge of a struggling Bombay weekly called *Young India*. His first decision was not about content. It was about money. He refused to carry advertisements, arguing that a newspaper treated as a means of earning a living would eventually produce serious malpractices, because a paper answerable to profit stops being answerable to anything else (Gandhiji, as cited in Newslaundry, 2019). He ran the paper instead on subscriptions alone, a model he repeated a few years later at *Navajivan*. It was, by his own account, less a business decision than a discipline. The reader paid, so the reader was owed the truth.

A century later, a reader in Patna opens a news app on their phone. That particular app costs nothing. However, what they see next, ranked by an algorithm trained feed to hold their attention as long as possible, is chosen by an entirely different logic. The advertiser pays for the space in the app, and what keeps the advertiser paying is engagement. Gandhiji's discipline and the aggregator's algorithm sit on opposite sides of the same question: who is the newspaper actually working for.

Most writing that notices this contrast stops at the observation. This chapter tries to close that gap without inventing new evidence to do it. It reads Gandhiji's own documented editorial record against the growing body of published, peer-reviewed research on how algorithmically curated news actually behaves, using a shared gatekeeping framework, so the comparison rests on what has already been established about each system rather than on how persuasively the comparison can be argued.

The chapter proceeds in five parts. The literature review traces gatekeeping theory from its origins in the human editor's decision-making through to its recent extension to platform algorithms, and situates the growing but narrowly focused literature on Indian digital gatekeeping within that arc. The theoretical framework sets out revenue structure, rather than technology, as the variable this chapter treats as decisive, and explains why that framing lets a 1920s newspaper and a 2020s aggregator be read inside the same model. The fourth section reads Gandhiji's documented editorial practice against the existing empirical literature on algorithmic

curation. The fifth sets out a coding framework through which this comparison could be tested directly against new primary data in future research.

The argument is not that Gandhiji anticipated the algorithm, or that a subscription-funded press is inherently superior to an advertising-funded one. Both models carry real trade-offs, a point this chapter returns to in its conclusion. The narrower claim is that different revenue structures produce different, and already documented, patterns in what gets surfaced to a reader, and that an Indian historical case, largely absent from a gatekeeping literature still dominated by Western, present-day studies, gives that claim a comparison of its own to draw on rather than one borrowed from elsewhere.

2. Literature Review

2.1 Gatekeeping Theory

Gatekeeping theory begins with a single wire editor. White (1950) tracked the news selections of one Midwestern telegraph editor, "Mr. Gates," over a month, and found that around ninety percent of the wire copy that crossed his desk never made it into the paper, filtered by a mix of professional judgment and personal preference that White treated as broadly representative of how news selection worked. The study's method, one gatekeeper and one set of choices closely observed, became the template the field returned to for decades, even as later work complicated its individualist premise. McCombs and Shaw (1972) shifted attention from what got selected to what effect selection had, showing that the issues media covered most heavily became the issues audiences rated as most important, which reframed gatekeeping as an act with downstream consequences for public attention rather than a purely internal editorial process. Entman (1993) added that selection was never neutral to begin with: how a story is framed, what is emphasized and what is left out, does as much gatekeeping work as the decision to run it at all. Shoemaker and Vos (2009) later synthesized this lineage into a formal theory, describing gatekeeping as a multi-level process operating simultaneously at the level of the individual journalist, the newsroom routine, the organization, and the broader social system, which is the version of the theory most contemporary digital gatekeeping scholarship still builds from.

Bagdikian (1983) while writing about the American press's growing corporate concentration, he argued that gatekeeping power is inseparable from economic power, and that whoever controls a

paper's revenue holds one of the most consequential levers available for shaping what a society sees and does not see. Bagdikian was writing about ownership concentration, not advertising specifically, but the underlying claim, that gatekeeping decisions answer to whoever pays the bills, is the hinge this chapter turns on when it reads Gandhiji's revenue model as a gatekeeping decision in its own right and not merely a financial one.

The theory's most recent extension addresses the algorithm directly. Wallace (2018) proposed a four-archetype model of digital gatekeeping, naming platform algorithms alongside journalists, strategic professionals, and individual amateurs as parallel gatekeepers, each governed by different access rules and selection logics. Thorson and Wells (2016) offered a related but distinct framing, arguing that what a reader eventually sees is the product of several layers of curation performed by different actors, including advertisers, layered on top of one another, a useful reminder that no single actor, human or algorithmic, fully controls what a reader encounters. Van Dalen (2023) pushed the argument further, treating platform algorithms not just as gatekeepers alongside journalists but as a new source of institutional power and legitimacy in their own right, one that professional communicators increasingly have to negotiate with rather than simply route around.

2.2 The Economics of Algorithmic Curation

If gatekeeping power tracks who pays, the relevant question for an algorithmic curator is what it is being paid to optimize for. Bandy and Diakopoulos (2020) offer the closest empirical answer available. Their audit of Apple News compared the app's algorithmically curated Trending Stories section against its human-curated Top Stories section over an extended data collection period, and found the human-curated section drawing on a noticeably wider and more even spread of sources than the algorithmic one, a difference in source diversity, concentration, and evenness measured directly rather than inferred. The difference was not a matter of one system being sloppier than the other. It reflected two different optimization targets: human editors selecting for a mix of newsworthiness criteria, and an algorithm selecting, more narrowly, for what would hold a reader's attention.

Cools et al. (2021) found a related pattern inside newsrooms adopting their own algorithmic recommendation tools, noting that these systems shift not just what gets published but who, or

what, is making the underlying decision, with journalists increasingly working alongside a system optimized for engagement metrics rather than the professional news judgment the gatekeeping literature originally described. Zuboff (2019) supplies the broader economic account of why this optimization target became dominant in the first place: platforms that monetize attention have a structural incentive to maximize the time a user spends engaged, regardless of whether that engagement serves the user's interests, a dynamic she terms surveillance capitalism. O'Neil (2016) makes a parallel point about the mathematical models underneath these systems, arguing that models built to optimize a narrow metric, engagement, ad clicks, watch time, will systematically favour whatever maximizes that metric over other values a more deliberative process might have weighed, even when no one involved intends that outcome.

Read together, this literature describes an algorithmic curator that is not neutral by default but structurally oriented toward a specific outcome, attention capture, because that is what its revenue model rewards. This is precisely the structural claim Gandhiji made about advertising a century earlier, applied to a different technology. Section 4 reads the two claims directly against each other.

2.3 The Indian Digital News Ecosystem

Indian digital gatekeeping scholarship has grown quickly in the last two years, but almost entirely in one direction: fact-checking and misinformation response around elections. Roy et al. (2025) studied Indian fact-checkers during the 2024 general election and described their role as “retroactive gatekeeping,” intervening after false or misleading content had already gone viral across multilingual, mobile-first platforms, rather than before publication in the way a traditional editor would. Their interviews with fact-checkers from six organizations found professionals working under conditions the classical gatekeeping literature, built around a single editor with time to deliberate, did not anticipate: content moving faster than verification, across languages the fact-checking organization may not have staff to cover, on platforms whose amplification mechanics sit entirely outside what a fact-check can reach.

This is genuinely important work, but it addresses a narrower question than the one this chapter asks. It studies what happens after an algorithm has already surfaced something, not the revenue logic that shaped the surfacing decision in the first place. Hastuti et al. (2025), in a systematic

review of 78 empirical studies on algorithmic influence in journalism, name this narrowness as a limitation of the wider field, not just of Indian scholarship specifically: the literature is dominated by Western, English-language studies, with few longitudinal designs and little historical grounding.

India's own press history sits, largely untouched, on the other side of that gap. Jeffrey (2000) documented the enormous expansion of the Indian-language press from the late 1970s onward, an expansion substantially funded by advertising revenue, and noted the tension this created between commercial growth and the press's self-understanding as a public institution. His later work (Jeffrey, 2010) traces this tension back further, to the period when, as he puts it, Gandhiji succeeded Bal Gangadhar Tilak as India's most prominent journalist and drew a sharp distinction between a “newspaper,” run for profit, and a “views paper,” run to persuade and serve. That distinction, made in the 1920s, is functionally a claim about gatekeeping logic: what a publication selects and emphasizes depends on what its revenue model asks it to select and emphasize for. Section 4 tests that specific, documented distinction against the empirical literature on algorithmic curation reviewed above.

3. Theoretical Framework

This chapter treats revenue structure as the primary variable shaping gatekeeping behaviour, rather than treating editorial and algorithmic curation as two categorically different kinds of decision-making. This is a deliberate departure from how much of the digital gatekeeping literature frames the comparison, which tends to treat the shift from human to algorithmic gatekeeping as primarily a technological change, an editor replaced by codes (Wallace, 2018; Van Dalen, 2023). That framing is not wrong, but it brackets a prior question: technological change aside, what is each gatekeeper, human or algorithmic, being asked to optimize for, and by whom.

Bagdikian's (1983) claim that gatekeeping power tracks economic power supplies the theoretical anchor. A gatekeeper answerable to a subscriber optimizes, at least in principle, for what the subscriber values enough to keep paying for. A gatekeeper answerable to an advertiser optimizes for what keeps the advertiser's audience assembled and attentive. A gatekeeper answerable to an engagement-driven ranking system optimizes for whatever the ranking model has been trained to

maximize, typically time on platform or click-through, a more extreme and more automatable version of the same advertiser-facing incentive (Zuboff, 2019; O'Neil, 2016).

Gandhiji's editorial practice at *Young India* and *Navajivan* is legible inside this framework as a specific, documented case of the first model. Both papers ran without carrying a single advertisement, printed as plain, full-scape weeklies funded through subscription alone ("As Mahatma Gandhi Turns 150," 2019). Gandhiji described this as a structural choice rather than an moral flourish: journalism treated as a means of earning a living would, in his words, produce serious malpractices, because a paper's obligations shift once its survival depends on profit rather than readership (Gandhiji, as cited in Newslaundry, 2019). He is widely quoted as holding that "the sole aim of journalism should be service" (Gandhiji, as cited in Newslaundry, 2019), a line that reads, inside this chapter's framework, less as a slogan than as a statement about who the gatekeeper answers to. Jeffrey's (2010) account of Gandhiji's "views paper" model, run to persuade rather than to turn a profit, describes the same structural choice from the historian's side of the archive.

This chapter reads India's contemporary algorithmic news aggregators as occupying a third position in this framework, answerable to an engagement-optimizing system that is itself answerable, several steps removed, to advertising revenue. The comparison this chapter builds are therefore not "old media versus new media" or "human editor versus algorithm." It is a comparison between two gatekeepers built on two different theories of who the gatekeeper is actually working for, one made explicit by its editor in the 1920s, the other made explicit, if at all, in a platform's investor filings rather than its user-facing interface.

4. Reading the Record: Gandhiji's Editorial Practice Against the Evidence on Algorithmic Curation

4.1 Gandhiji's Editorial Record

Two features of Gandhiji's editorship are documented closely enough to work with directly, rather than through inference. The first is structural: *Young India* and *Navajivan* were each eight-page, full-scape weeklies that carried no advertising at all, printed in a plain double-column layout, funded through subscription alone ("As Mahatma Gandhiji Turns 150," 2019). Gandhiji

wrote explicitly that treating a newspaper as a means of earning a living would, in his words, produce serious malpractices, because a paper's obligations shift once its survival depends on profit rather than readership (Gandhiji, as cited in Newslaundry, 2019). He held to this position even when it cost him: circulation of both papers dropped sharply, to under eight thousand copies, after his 1922 imprisonment on sedition charges, and he did not respond by opening the papers to advertising to recover the lost revenue ("As Mahatma Gandhiji Turns 150," 2019). The second documented feature is normative rather than financial. Gandhiji described the aim of journalism, across more than two decades of editorship spanning Indian Opinion, Young India, and Navajivan, as service rather than profit or persuasion for its own sake (Gandhiji, as cited in Newslaundry, 2019). He paired this with an explicit theory of restraint: a newspaper wielded power comparable to an uncontrolled river, capable of devastating whatever it touched unless the editor exercised control from within rather than waiting for control to be imposed from without (Gandhiji, as cited in Newslaundry, 2019). Jeffrey (2010) reads this combination, service as the stated aim and restraint as the operating discipline, as the defining feature that separated Gandhiji's "views paper" from the commercially driven "newspaper" model that came to dominate Indian-language print after his editorship ended.

Read together, these two features describe a gatekeeper whose selection criteria were, by design, insulated from a specific external pressure: the need to hold an advertiser's audience. This is not a claim that Gandhiji's editorial choices were free of bias or agenda; his papers were openly partisan instruments of the independence movement, and he never pretended otherwise. It is a narrower, more checkable claim about the mechanism his revenue model closed off, which the next section finds open, and structurally load-bearing, in algorithmic news curation.

4.2 What the Existing Evidence Shows About Algorithmic Curation

The empirical literature on algorithmic news curation, though built almost entirely outside India, gives a reasonably consistent account of what that open mechanism produces in practice. Bandy and Diakopoulos's (2020) audit of Apple News compares an algorithmic and a human-curated section inside the same platform rather than comparing across different outlets with different resources. Over an extended data collection period, the algorithmically curated Trending Stories section drew on a narrower and less even spread of sources than the human-curated Top Stories section running alongside it. The two sections were subject to the same editorial standards, the

same platform, the same technical constraints; the difference tracked the selection logic itself, human editors weighing a mix of newsworthiness criteria against an algorithm optimizing more narrowly for what would hold attention.

Hastuti et al.'s (2025) systematic review of 78 empirical studies situates this finding as part of a broader, consistent pattern rather than a one-platform anomaly. Across the studies they reviewed, algorithmic systems reconfigure gatekeeping toward what they term "shareworthiness," and platform business models intensify this dependence on engagement metrics in ways that measurably limit investigative depth. Their review also names the field's own blind spot: a dominance of Western, English-language studies, with little historical grounding and few longitudinal designs, meaning that even this consistent pattern has rarely been tested against a case old enough, or documented enough, to isolate revenue structure as the variable doing the work.

India's contribution to this literature so far comes from a different angle. Roy et al.'s (2025) study of Indian fact-checkers during the 2024 general election found professionals operating as what they call "retroactive gatekeepers," intervening only after algorithmically amplified misinformation had already gone viral across multilingual, mobile-first platforms. Their interviews describe fact-checkers working after the fact, correcting individual claims once a platform's own ranking and sharing mechanics have already carried them to a wide audience, a position that leaves the amplification mechanism itself outside what any single fact-check can address. Taken together, these three bodies of evidence, a controlled within-platform audit, a synthesis across dozens of studies, and an India-specific account of retroactive correction in practice, converge on the same underlying description: engagement-optimized curation systematically narrows source diversity and depth relative to human curation, in ways that leave the professionals working around it, not just the readers inside it, working after the decision has already been made.

4.3 Reading the Two Records Together

Placed side by side, Gandhiji's editorial record and the empirical literature on algorithmic curation describe two gatekeepers built on opposite answers to the same design question: what should the selection mechanism be insulated from. Gandhiji's model closed off dependence on an external audience-holder, the advertiser, and replaced it with dependence on the subscriber, a

narrower but more directly accountable relationship. The algorithmic model documented by Bandy and Diakopoulos (2020) and synthesised by Hastuti et al. (2025) does the reverse: it opens the selection mechanism directly to an engagement signal that is, several steps removed, an advertiser's proxy for audience attention, and the narrowing of source diversity that follows is not a flaw reviewers happened to catch but the visible trace of what the mechanism was built to optimize.

This reading does not require treating Gandhiji as a prophet of platform design, and the chapter is not making that claim. His papers operated in a media environment with no equivalent of a ranking algorithm, no comment sections, no share counts. What travels across the century is narrower: the structural insight that a gatekeeper's selection pattern is legible from its revenue model, and that this pattern shows up in what actually gets published or surfaced, not only in what an editor says about their own values. Bagdikian (1983) made a version of the same argument about American press ownership six decades after Gandhiji made it about advertising; Bandy and Diakopoulos (2020) demonstrated it empirically about a single app's two sections within the same platform. Gandhiji's editorship offers a documented case where the same underlying claim was tested, deliberately, by a working editor, a century before the literature reviewed in Section 2 arrived at a version of it independently. Section 5 sets out how that historical case and the algorithmic case could be tested against each other directly, using new primary data, as a next step this chapter's conceptual work makes possible rather than completes.

5. A Proposed Method for Empirical Comparison

The reading offered in Section 4 rests on published evidence about algorithmic curation in general and a documented historical record of Gandhiji's editorial practice. It does not yet rest on a direct, matched comparison between the two, coded on the same terms. That comparison is a natural next step, and this section sets out how it could be done.

5.1 Archival Sample

A matched historical sample would draw on issues of *Young India* and *Navajivan* from Gandhiji's editorship, accessible through the Gandhi Heritage Portal, coded for three variables: front-page topic category, the proportion of page space given to reader correspondence, and the

presence or absence of any content resembling sponsored placement. A sample of ten to fifteen issues drawn across a defined multi-year window would be sufficient to establish a baseline pattern without requiring exhaustive coverage of the full editorship.

5.2 Aggregator Sample

A matched contemporary sample would log trending and top-story placements on Indian news aggregator platforms such as Dailyhunt and InShorts, coded on the same three variables adapted to the platform context: topic category, source diversity across sampled placements, and the extent to which placement tracks engagement signals such as share counts or repeated resurfacing rather than other editorial criteria. A two-week logging window, comparable to the extended collection period Bandy and Diakopoulos (2020) used for their single-platform audit, would allow the pattern to stabilise beyond any single day's anomalies.

5.3 Coding and Analysis

Both samples would be coded using an identical scheme adapted to each medium's format, so the two datasets remain genuinely comparable rather than merely parallel, with ambiguous cases resolved against clear, pre-defined category boundaries and, ideally, a second coder for a reliability check on a subset of each sample. The analysis would be descriptive and comparative, appropriate to the scale of both samples, rather than a claim to statistical generalisability: the contribution of such a study would be the method and the historical case it makes available, not a definitive measurement of either system.

Executed carefully, this design would let a future study state directly what this chapter can only argue from existing evidence: whether Gandhiji's documented, revenue-insulated gatekeeping pattern and today's engagement-optimised aggregator pattern differ in the specific, measurable ways the theoretical framework in Section 3 predicts, using primary data collected and coded for that comparison alone.

6. Conclusion

Gandhiji's decision to keep advertising out of *Young India* and *Navajivan* was not a nostalgic gesture available only to a moment before commercial media matured. It was an argument, made

in print and repeated across more than two decades of editorship, about who a newspaper is accountable to and how that accountability shows up in what actually gets published. This chapter has tried to take that argument as seriously as the parallels drawn between Gandhiji's press and today's algorithmic aggregators usually allow, by reading it directly against the empirical literature on algorithmic curation rather than treating the comparison as a rhetorical flourish.

The two records, read together, converge on a structural claim rather than a nostalgic one: a gatekeeper's selection pattern tracks its revenue model, whether that gatekeeper is a person exercising editorial restraint by design or an algorithm optimising for engagement by design. Neither record proves the other. What they do is make the same underlying claim independently, a century apart, in two entirely different media systems, which is itself a reason to take the claim seriously rather than dismiss it as a coincidence of framing. None of this argues that a subscription-funded press is straightforwardly superior to an advertising or engagement funded one. Subscription models carry their own distortions, including a tendency to serve existing subscribers' preferences over a genuinely public interest, a limitation Gandhiji's own model did not escape and did not claim to. The contribution this chapter offers is narrower and, it hopes, more useful: a documented Indian case, currently absent from a literature still dominated by Western and presentist accounts, for studying revenue structure as a legible variable in gatekeeping behaviour, and a concrete method, set out in Section 5, for testing that case directly against new primary data. Where Bagdikian (1983) asked who controls the gate, and Wallace (2018) asked what kind of gatekeeper controls it, this chapter has tried to ask a third question: what is the gatekeeper being paid to do, and does the answer look any different when the gatekeeper is a person than when it is code. The evidence assembled here suggests, provisionally, that it does not look very different at all.

References

- As Mahatma Gandhiji turns 150, his news magazines Navjivan and Young India turn 100. (2019, October 2). ThePrint. https://theprint.in/opinion/as-mahatma-Gandhiji-turns-150-his-news-magazines-navjivan-and-young-india-turn-100/298484/
- Bagdikian, B. H. (1983). The media monopoly. Beacon Press.

- Bandy, J., & Diakopoulos, N. (2020). Auditing news curation systems: A case study examining algorithmic and editorial logic in Apple News. Proceedings of the International AAAI Conference on Web and Social Media, 14(1), 36–47. https://doi.org/10.1609/icwsm.v14i1.7277
- Cools, H., Van Gorp, B., & Opgenhaffen, M. (2021). When algorithms recommend what's new(s): New dynamics of decision-making and autonomy in newsgathering. Media and Communication, 9(4), 198–207. https://doi.org/10.17645/mac.v9i4.4173
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x
- Gandhi Heritage Portal. https://www.gandhiheritageportal.org/
- Hastuti, H., Maulana, H. F., Lawelai, H., & Suherman, A. (2025). Algorithmic influence and media legitimacy: A systematic review of social media's impact on news production. Frontiers in Communication, 10, Article 1667471. https://doi.org/10.3389/fcomm.2025.1667471
- Jeffrey, R. (2000). India's newspaper revolution: Capitalism, politics and the Indian-language press, 1977–99. Oxford University Press.
- Jeffrey, R. (2010). Mission, money and machinery: Indian newspapers in the twentieth century (ISAS Working Paper No. 117). Institute of South Asian Studies, National University of Singapore.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. Public Opinion Quarterly, 36(2), 176–187. https://doi.org/10.1086/267990
- Meet Mahatma Gandhi, the journalist: “When a newspaper is a means of making profit, serious malpractices are likely.” (2019, October 2). Newslaundry. https://www.newslaundry.com/2019/10/02/meet-mahatma-gandhi-the-journalist-when-a-newspaper-is-a-means-of-making-profit-serious-malpractices-are-likely
- O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.
- Roy, E., Laskar, K. A., & Reyaz, M. (2025). Retroactive gatekeepers: Fact-checkers, electoral disinformation, and audience engagement in India. Emerging Media, 3(4), 582–600. https://doi.org/10.1177/27523543251395190
- Shoemaker, P. J., & Vos, T. P. (2009). Gatekeeping theory. Routledge.

- Thorson, K., & Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. Communication Theory, 26(3), 309–328. https://doi.org/10.1111/comt.12087
- Van Dalen, A. (2023). Algorithmic gatekeeping for professional communicators: Power, trust, and legitimacy. Routledge.
- Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. Digital Journalism, 6(3), 274–293. https://doi.org/10.1080/21670811.2017.1343648
- White, D. M. (1950). The "gate keeper": A case study in the selection of news. Journalism Quarterly, 27(4), 383–390.

Chapter 16

Perceived Credibility of News Content on Instagram Reels among Generation Z: A Conceptual Perspective

Ms. Saumya Grover,
Assistant Professor,
Jagan Nath University, Delhi NCR

Abstract

The rapid expansion of social media has transformed the ways in which news is produced, distributed, consumed, interpreted, and evaluated. Among the various social media platforms, Instagram has emerged as an important environment for news consumption, particularly among younger audiences. The introduction and growing popularity of Instagram Reels have further changed the nature of digital news by presenting information through short, visually engaging, mobile-oriented videos. For Generation Z, who have grown up in an environment characterised by smartphones, social networking platforms, algorithmic content recommendations, and continuous digital connectivity, social media increasingly functions not only as a space for social interaction and entertainment but also as a gateway to information and current affairs.

This chapter examines the concept of perceived credibility of news content on Instagram Reels, with particular attention to Generation Z. It conceptualises credibility as a multidimensional perception involving accuracy, trustworthiness, believability, objectivity, completeness, source expertise, transparency, and authenticity. The chapter discusses the changing nature of news credibility in the social media environment and explores how the characteristics of Instagram Reels may influence users' perceptions of news. It further examines the role of content characteristics, source characteristics, platform characteristics, visual presentation, engagement indicators, personalization, algorithmic exposure, and audience characteristics in shaping perceived credibility. The chapter also situates the discussion within established perspectives such as source credibility theory, media credibility, uses and gratifications, social presence, and the elaboration likelihood perspective. Particular attention is given to Generation Z as a digitally immersed audience whose credibility judgments may differ from those of traditional news

consumers. The chapter concludes by proposing a conceptual framework for understanding perceived credibility of news content on Instagram Reels and identifying important directions for future empirical research.

Keywords: News Credibility, Perceived Credibility, Instagram Reels, Generation Z, Social Media, Digital Journalism, News Consumption, Short-Form Video

1. Introduction

The digital transformation of journalism has fundamentally altered the relationship between news organisations, journalists, audiences, and information. Traditional news media once occupied a relatively dominant position in determining what information reached the public. Newspapers, television channels and radio stations functioned as important gatekeepers that selected, verified and organised information before presenting it to audiences. The growth of the internet disrupted this model by allowing information to circulate rapidly across decentralised networks. Social media subsequently accelerated this transformation by enabling ordinary users, journalists, influencers, media organisations, political actors and institutions to participate simultaneously in the creation and distribution of information.

The contemporary news environment is therefore characterised by abundance rather than scarcity. Audiences are no longer dependent on a limited number of news outlets for information. Instead, they encounter news across search engines, social networking sites, video platforms, messaging applications and digital news websites. This abundance has created an important problem: although access to information has become easier, determining whether the information is reliable has become increasingly difficult.

The question of credibility is consequently central to contemporary journalism. In traditional media, credibility was often associated with the reputation of a newspaper, television channel, journalist or broadcasting organisation. In social media environments, however, credibility may be attributed to a much wider range of actors. A news story may originate from a recognised media organisation but reach users through an individual creator. A journalist may publish a news update directly through a personal social media account. An influencer may interpret a current event for followers without formal journalistic training. A user may encounter a

professionally produced news video without immediately recognising who created or verified it.

This complexity becomes particularly significant in the context of short-form video. Instagram Reels has become an important format for the circulation of short, visually oriented content. Although Instagram was initially associated primarily with photographs and lifestyle-oriented content, the platform has increasingly incorporated video-based communication. Reels allow creators and organisations to present information in a brief, engaging and highly visual format. News organisations have consequently adopted short-form video to reach audiences who may not regularly visit conventional news websites or consume traditional television news.

The transformation raises a significant conceptual question: What makes news content credible when it is encountered in the form of an Instagram Reel?

Credibility in this environment cannot necessarily be understood using only the criteria traditionally applied to newspapers or television news. A user watching a 30-second or 60-second Reel may make an immediate judgment based on the appearance of the creator, presentation style, captions, visuals, language, source attribution, comments, likes, shares and perceived authenticity. The platform itself may also influence credibility perceptions because users encounter content through personalised feeds rather than through deliberate searches for particular news stories.

This issue is especially relevant for Generation Z. Generally understood as the generation that has grown up with digital technologies and social media, Generation Z has experienced news consumption in an environment where information is increasingly mobile, visual, interactive and personalised. Their relationship with news may therefore differ from that of audiences whose primary exposure to news occurred through print newspapers, radio or scheduled television broadcasts.

The study of perceived credibility among Generation Z is consequently important not only for understanding audience behaviour but also for understanding the future of journalism. If younger audiences increasingly encounter news through social media, the credibility judgments they make on these platforms may influence whether they accept, question, share, ignore or challenge information.

2. Understanding News Credibility

Credibility is one of the most established concepts in communication research. At its simplest level, credibility refers to the extent to which a source, message or medium is perceived as believable and trustworthy. However, credibility is not an entirely objective property. It is substantially influenced by audience perception.

A news organisation may consider a report accurate because it has followed professional journalistic procedures, but an audience member may nevertheless perceive the report as biased or incomplete. Similarly, an individual creator without formal journalistic credentials may be perceived as highly credible by followers because of familiarity, perceived authenticity or previous experience. This distinction between actual credibility and perceived credibility is important.

Actual credibility may refer to the objective quality, accuracy or reliability of information, whereas perceived credibility refers to the audience's judgment regarding whether information should be trusted or believed. Since audiences generally cannot independently verify every piece of information they encounter, their decisions frequently depend on cues that help them evaluate credibility.In journalism, credibility has traditionally been associated with several characteristics, including accuracy, fairness, objectivity, completeness, balance, expertise and trustworthiness. These characteristics remain important in digital environments but are now accompanied by additional cues.Social media has introduced a more complex credibility environment because the source, message and medium are intertwined. A user does not simply evaluate "the news." Instead, the user may evaluate the person presenting it, the organisation behind it, the visual design of the post, the platform, audience reactions and the context in which the content appears. Credibility can therefore be conceptualised as multidimensional.

2.1 Accuracy

Accuracy concerns whether the information corresponds with available facts and reality. For news consumers, accurate reporting is one of the fundamental characteristics of credible journalism.

2.2 Trustworthiness

Trustworthiness refers to the extent to which users believe that the source has honest intentions and can be relied upon.

In social media environments, trustworthiness may be associated with the reputation of a news organisation, journalist or creator. Repeated exposure may also influence perceptions. A user who believes that a particular creator has consistently provided useful information may develop greater trust in subsequent content.

2.3 Expertise

Expertise concerns the perceived knowledge or competence of the communicator. On Instagram, expertise may be communicated through professional identity, qualifications, journalistic affiliation, references to credible sources or familiarity with the subject. However, expertise may also be perceived through presentation competence. A confident and articulate creator may appear knowledgeable even when formal credentials are not immediately visible.

2.4 Objectivity and Fairness

News audiences often expect credible journalism to distinguish reporting from personal opinion. Objectivity, neutrality and fairness therefore remain important dimensions of credibility.Instagram Reels may complicate this distinction because short-form video frequently relies on personality-driven communication. The creator may speak directly to the camera, use expressive language or combine news with commentary. Consequently, audiences may need to determine whether they are watching factual reporting, interpretation, commentary or opinion.

2.5 Completeness

Completeness refers to whether sufficient information is provided to understand an issue. This dimension is particularly significant for short-form news. A Reel may accurately communicate one aspect of an event while omitting relevant background or opposing perspectives. Consequently, perceived credibility may depend on whether users believe that the Reel provides enough context for them to understand the issue.

2.6 Authenticity

Authenticity has become increasingly important in digital communication. Audiences often value content that appears genuine, transparent and less artificially constructed.

For Generation Z, perceived authenticity may be particularly significant because social media communication frequently involves direct and informal interaction. A creator who appears relatable and transparent may be perceived as more credible than highly polished content that appears promotional or distant.

3. The Transformation of News Credibility in the Social Media Environment

The emergence of social media has transformed credibility from a relatively institutional concept into a networked and participatory phenomenon.In traditional journalism, credibility was closely associated with established institutional gatekeepers. News organisations possessed editorial structures, professional norms and established reputations. Social media reduces the distance between producers and audiences. News can now be published, commented upon, shared and challenged almost instantaneously.On one hand, social media enables faster access to information and provides audiences with multiple perspectives. Users can encounter eyewitness accounts, expert commentary, local information and direct communication from institutions.On the other hand, the same environment enables misinformation, manipulated media, misleading headlines, fabricated content and contextually distorted information to circulate rapidly.

The credibility problem is therefore no longer simply about whether journalism is credible. It is also about how audiences identify credibility within an environment where professional and non-professional information appear alongside each other.Social media users may employ several heuristics to address this problem. They may look at the source name, profile information, verification status, visual quality, language, supporting evidence, comments and engagement levels. These cues can help users make quick judgments when they do not have sufficient time or motivation to independently verify information.

4. Instagram Reels as a News Environment

Instagram Reels represents a distinctive form of digital communication because it combines video, sound, text, graphics, captions, music, interaction and algorithmic distribution.Unlike a traditional newspaper article, a Reel does not require the audience to commit to extended reading. The information is delivered through a highly compressed audiovisual format. This makes it particularly suitable for mobile-first audiences who consume content while scrolling through personalised feeds.First, visual presentation can influence perceived professionalism. Clear graphics, readable captions, appropriate footage and coherent editing may create an impression of competence.Second, audio and delivery style can influence perceptions. A confident voice, conversational tone or professional narration may strengthen perceived credibility, while exaggerated delivery may reduce it.Third, brevity affects the amount of context that can be communicated. A short Reel may capture attention effectively but may not provide sufficient background for complex issues.Fourth, engagement indicators such as views, likes, comments and shares can act as social signals. Users may interpret high engagement as evidence that others consider the content valuable or credible. However, popularity and credibility are not necessarily equivalent.Fifth, platform algorithms influence what users see. Because users may encounter news content without actively searching for it, the context of exposure becomes an important part of credibility formation.Thus, Instagram Reels should not be understood simply as another channel for distributing traditional news. It represents a distinct communication environment in which the form of presentation, platform characteristics and social interaction influence how information is interpreted.

5. Generation Z as a News Audience

Generation Z represents an important audience for understanding contemporary digital news consumption. Although definitions of Generation Z vary across studies, the generation is generally associated with individuals who have grown up during the expansion of smartphones, social media and digital communication. Their media environment is characterised by continuous connectivity, visual communication, on-demand information and personalised digital experiences.For Generation Z, the boundary between entertainment, social interaction and information is often fluid. A user may open Instagram primarily to communicate with friends or view entertainment content but subsequently encounter a news Reel. News consumption can therefore become incidental rather than intentional.Traditional news consumption often involves

deliberate selection. A person may purchase a newspaper, switch on a television news channel or visit a news website. On Instagram, however, information may appear while the user is engaged in another activity. The credibility judgment may therefore occur within seconds.

Generation Z users may assess:

- Who posted the content?
- Does the information appear believable?
- Does the creator seem knowledgeable?
- Is the information supported by evidence?
- Does the video look professional?
- Are credible sources mentioned?
- What are other users saying in the comments?
- How many people have liked or shared the Reel?
- Does the content appear authentic?
- Does the information agree with what they already know?
- Is the presentation factual or opinion-based?

These judgments may occur simultaneously and often unconsciously.The study of Generation Z therefore requires a broader understanding of credibility that incorporates both traditional journalistic standards and social media-specific cues.

6. Theoretical Perspectives Relevant to Perceived Credibility

6.1 Source Credibility Theory

Source credibility theory provides one of the most important foundations for understanding perceived credibility. The theory traditionally emphasises characteristics such as expertise and trustworthiness. A communicator perceived as knowledgeable and trustworthy is more likely to influence audience attitudes and judgments.In the context of Instagram Reels, source credibility may be associated with:

- journalist identity;
- media organisation;
- professional qualifications;

- previous reputation;
- verification or authentication signals;
- demonstrated knowledge;
- transparency regarding sources; and
- consistency of previous content.

Source credibility is especially important when users encounter information from unfamiliar creators. In such situations, users may rely on visible cues to determine whether the communicator is worthy of trust.

6.2 Media Credibility Perspective

Media credibility shifts attention from individual communicators to the broader medium or institution. Historically, audiences have developed perceptions regarding the credibility of newspapers, television channels and radio stations. In the digital environment, platforms themselves may influence expectations.

Instagram may be perceived primarily as a social and entertainment platform rather than a traditional news institution. This distinction may affect how users evaluate news encountered there.

The credibility of a news Reel may therefore be shaped by an interaction between the credibility of the source and the credibility attributed to the platform.

6.3 Uses and Gratifications Perspective

Uses and gratifications theory emphasises the active role of audiences in selecting media to satisfy particular needs. Generation Z users may consume news Reels for information, entertainment, social interaction, convenience, surveillance of current events or participation in public conversations.

The motivation behind consumption may influence credibility judgments. A user seeking quick information may prioritise brevity and convenience, whereas a user seeking detailed understanding may place greater importance on completeness and source verification. Thus,

credibility should be considered alongside the reasons users engage with news content.

6.4 Elaboration Likelihood Perspective

The elaboration likelihood perspective provides another useful conceptual lens. When users are highly motivated and capable of processing information, they may evaluate arguments, evidence and sources carefully. When motivation or attention is low, users may rely on peripheral cues.Instagram's scrolling environment may encourage rapid judgments. Visual appeal, creator attractiveness, popularity, engagement numbers, production quality and emotional presentation may therefore function as peripheral cues.However, this does not imply that Generation Z is incapable of critical evaluation. Rather, the level of cognitive involvement may vary according to the importance of the news topic, personal relevance and available time.

7. Dimensions of Perceived Credibility of News Reels

A conceptual understanding of Instagram news credibility can be developed through several interconnected dimensions.

7.1 Source-Related Factors

The first group involves characteristics of the communicator. These include expertise, trustworthiness, reputation, transparency and authenticity. A recognised journalist or established news organisation may possess an initial credibility advantage. However, social media creators can also develop credibility through consistent and transparent communication.

7.2 Message-Related Factors

The content itself can influence credibility. Important characteristics include factual accuracy, clarity, evidence, completeness, balance and logical presentation. A Reel that clearly identifies the event, location, date, source and relevant context may appear more credible than a Reel relying exclusively on dramatic statements.

7.3 Visual and Production Factors

Instagram is fundamentally visual. Therefore, production characteristics may influence perceived

credibility.

Relevant cues include:

- quality of visuals;
- captions and typography;
- editing;
- graphics;
- audio quality;
- visual consistency;
- use of supporting footage;
- presentation style; and
- professional appearance.

However, production quality should not be confused with factual credibility. Highly polished misinformation may appear more credible than poorly produced factual information. This distinction is particularly important when studying audience perception.

7.4 Platform-Related Factors

The platform provides the environment within which the message is encountered. The credibility judgment may be influenced by the Instagram interface, recommendation system, profile information, comments, verification indicators and interaction features.

The platform also determines visibility. Content that repeatedly appears in a user's feed may gradually become familiar, and familiarity can influence perceptions.

7.5 Social Validation

Social validation refers to the influence of other users on credibility perceptions. Likes, shares, comments and views can operate as social cues. A user may perceive highly engaged content as more important or trustworthy. Conversely, negative comments may create doubt.

However, engagement should not automatically be interpreted as credibility. Viral content can be

inaccurate, sensational or misleading. Therefore, the relationship between popularity and credibility remains conceptually complex.

7.6 Personal Relevance

Personal relevance may influence how carefully users evaluate information. When a Reel concerns an issue directly affecting a user's life, education, community or interests, the user may devote greater attention to the information. Personal relevance can therefore influence both the motivation to verify information and the importance assigned to credibility.

7.7 Prior Knowledge

Existing knowledge can shape credibility judgments. Users with substantial knowledge about a topic may recognise inconsistencies more easily. Users with limited knowledge may depend more heavily on external cues such as source reputation, engagement and presentation style.

Generation Z should therefore not be treated as a homogeneous audience. Differences in education, media literacy, interests, political awareness, digital skills and news habits can lead to different credibility judgments.

8. The Role of Media and News Literacy

Media literacy is increasingly important in the context of social media news consumption. Media literacy involves the ability to access, analyse, evaluate and create media content. For news consumers, this includes recognising sources, identifying potential bias, distinguishing fact from opinion, checking evidence and understanding how media messages are constructed.

In the Instagram environment, media literacy may influence whether users rely on superficial cues or engage in deeper evaluation.

A media-literate user may ask:

Who created this content?

What evidence is provided?

Where did the information originate?

Is the headline consistent with the actual content?

Is this reporting or commentary?

What information has been omitted?

Can the claim be verified through another credible source?

Such questions can reduce dependence on popularity or visual presentation as indicators of credibility. Media literacy is therefore an important conceptual variable when examining Generation Z's perceptions of news credibility.

9. Authenticity, Influencers and the Changing News Source

One of the distinctive characteristics of social media journalism is the growing importance of individual creators.

Traditional journalism relies heavily on institutional authority. Social media communication can rely more heavily on personal relationships.A creator may develop a loyal audience that perceives the creator as authentic, relatable and trustworthy. Consequently, personal familiarity may sometimes compete with institutional authority.A professionally produced news Reel from a major news organisation may be perceived as credible because of institutional reputation. At the same time, an informal Reel produced by an individual creator may be perceived as credible because the creator appears authentic and personally connected to the audience. Therefore, credibility on Instagram may emerge from two different pathways: **institutional credibility** and **relational credibility**.Institutional credibility is associated with professional reputation, journalistic standards and organisational authority.Relational credibility is associated with familiarity, authenticity, interaction and perceived closeness between creator and audience.The relative importance of these two forms of credibility is an important area for empirical investigation.

10. Algorithmic Personalisation and Credibility

Algorithmic recommendation represents another important dimension.

Social media platforms increasingly personalise users' content feeds according to their behaviour, interests and interactions. This means that two users may receive substantially different information about the same event.

Personalisation may have both positive and negative implications.It can help users discover information relevant to their interests and reduce information overload. However, repeated exposure to similar viewpoints may also contribute to selective exposure and reinforce existing beliefs.From a credibility perspective, repeated exposure may create familiarity. A user may begin to perceive frequently encountered information as more believable simply because it is familiar.This creates an important distinction between being repeatedly exposed to information and being exposed to credible information.Consequently, algorithmic visibility should not be treated as evidence of credibility. Rather, it should be conceptualised as a contextual factor that may influence the process through which credibility perceptions develop.

11. The Credibility Paradox of Short-Form News

Its primary strength is its ability to communicate information quickly and attract attention. Yet the same brevity can limit context. A Reel may successfully communicate the central event but leave unanswered questions about causes, consequences, historical background or competing perspectives. This creates tension between accessibility and completeness.For Generation Z, short-form content may be particularly attractive because it fits existing patterns of mobile media consumption. However, attractive presentation does not guarantee comprehensive understanding.The credibility of short-form news should therefore not be evaluated solely in terms of whether a statement is technically accurate. Researchers should also consider whether the information provides sufficient context to prevent misleading interpretations.

12. Conceptual Framework

Based on the preceding discussion, perceived credibility of news content on Instagram Reels can be conceptualized as the outcome of multiple interacting factors. The conceptual framework may be represented through five broad categories: Source Factors → Message Factors → Platform

Factors → Audience Factors → Perceived Credibility

Source Factors

- Expertise
- Trustworthiness
- Reputation
- Authenticity
- Transparency

Message Factors

- Accuracy
- Clarity
- Completeness
- Objectivity
- Evidence
- Balance

Platform Factors

- Visual presentation
- Engagement indicators
- Verification signals
- Algorithmic exposure
- Comments and social validation

Audience Factors

- Media literacy
- Prior knowledge
- Personal relevance
- News consumption habits
- Motivation

- Digital familiarity

Outcome

Perceived Credibility of News Content on Instagram Reels

The framework recognises that credibility is not generated by a single characteristic. Rather, it develops through the interaction between the source, message, platform and audience.

A useful conceptual representation is:

Source Credibility + Message Credibility + Platform Cues + Audience Characteristics → Perceived Credibility

This framework can subsequently be operationalised and tested empirically among Generation Z respondents in Delhi NCR.

13. Delhi NCR as a Relevant Context

Delhi NCR provides an appropriate contextual environment for examining this phenomenon because of its diverse and digitally connected population.The region represents a complex media environment in which national news, regional issues, political developments, entertainment, lifestyle content and hyperlocal information coexist.Generation Z users in Delhi NCR may encounter news from national media organisations, regional outlets, independent journalists, creators and individual social media users. The diversity of sources creates an environment in which credibility judgments may become particularly important.The region's heterogeneous population also provides an opportunity to examine whether demographic and media-use differences influence perceptions of credibility.Factors such as age, gender, educational background, occupation, frequency of Instagram use, frequency of news consumption and preferred sources may potentially contribute to differences in perceived credibility.

14. Conceptual Propositions for Future Research

The conceptual discussion suggests several propositions that can guide empirical investigation.

Proposition 1: Higher perceived expertise of the news source is likely to be associated with higher perceived credibility of news content on Instagram Reels.

Proposition 2: Higher perceived trustworthiness of the source is likely to positively influence perceived credibility.

Proposition 3: News Reels perceived as accurate, balanced and adequately supported by evidence are likely to receive higher credibility evaluations.

Proposition 4: Perceived authenticity of the creator is likely to influence Generation Z's credibility judgments.

Proposition 5: Visual and production quality may positively influence initial perceptions of credibility, although production quality alone may not determine actual credibility.

Proposition 6: Social engagement indicators such as likes, comments and shares may influence perceived credibility through social validation.

Proposition 7: Higher levels of media literacy may encourage users to rely more heavily on substantive credibility indicators and less heavily on superficial platform cues.

Proposition 8: Personal relevance of a news topic may increase the level of cognitive evaluation applied to the content.

Proposition 9: Prior familiarity with a source may influence perceived credibility.

Proposition 10: Repeated algorithmic exposure may influence familiarity and consequently affect perceived credibility.

These propositions provide a foundation for developing measurable variables and hypotheses in an empirical study.

15. Implications for Journalism and News Organisations

The conceptual discussion has important implications for contemporary journalism. News

organisations cannot assume that institutional reputation alone guarantees credibility among younger social media audiences. They must communicate credibility through the formats in which audiences actually consume news.This includes clearly identifying sources, providing context, distinguishing fact from opinion, using accurate captions and visuals, and maintaining transparency.News organisations should also recognise that short-form news is not merely a promotional tool for longer stories. It has become an important form of news communication in its own right.

Journalists working with Reels therefore need to balance brevity with accuracy and engagement with context.The challenge is not simply to make news more attractive. It is to make credible journalism understandable and accessible without compromising essential information.

16. Implications for Generation Z

For Generation Z, the conceptual issue extends beyond individual news preferences.Young audiences increasingly operate within information environments where news, entertainment, advertising, commentary and user-generated content are mixed together.Consequently, credibility evaluation becomes an everyday digital skill.Generation Z audiences may benefit from developing habits such as checking original sources, comparing information across multiple outlets, examining evidence, distinguishing reporting from commentary and recognising emotionally manipulative content.The ability to critically evaluate short-form news is likely to become increasingly important as digital platforms continue to influence public knowledge and civic participation.

17. Conclusion

The emergence of Instagram Reels represents an important development in the transformation of digital news consumption. Short-form video has created a communication environment in which news is increasingly visual, brief, personalised, interactive and personality-driven. For Generation Z, this environment is particularly significant because social media platforms are integrated into everyday patterns of information seeking and communication.The concept of perceived credibility is therefore central to understanding how Generation Z evaluates news on Instagram Reels.Credibility in this environment is multidimensional. It involves traditional

journalistic characteristics such as accuracy, objectivity, expertise, trustworthiness and completeness, but it also incorporates social media-specific factors such as authenticity, visual presentation, engagement indicators, creator identity, platform context and algorithmic exposure.The central conceptual argument of this chapter is that perceived credibility is not located exclusively within the news message or the news organisation. It is constructed through the interaction of source characteristics, message characteristics, platform cues and audience characteristics. For Generation Z, this construction may occur rapidly and within highly personalised information environments. A Reel's credibility may therefore be influenced by the reputation of its creator, the quality of its information, the appearance of the presentation, the reactions of other users and the user's own knowledge and media literacy. At the same time, the credibility of short-form news involves an important tension between accessibility and completeness. The ability to communicate news quickly is valuable, but brevity can also remove context. Consequently, a credible news Reel should not merely attract attention or present technically accurate information. It should provide sufficient evidence, context and transparency to allow audiences to form informed judgments. The conceptual framework developed in this chapter provides a foundation for examining these relationships empirically. In particular, it can support research investigating how source credibility, message credibility, platform characteristics, social validation and audience characteristics influence the perceived credibility of news content on Instagram Reels among Generation Z in Delhi NCR.Future empirical research can operationalise these constructs through structured scales and examine their relative contribution to credibility perceptions. Such research can contribute to the broader scholarship on digital journalism, social media news consumption and youth media behaviour while also providing practical insights for journalists, news organisations, educators and policymakers.

References

- Abdulla, Rasha A., et al. *The Credibility of Newspapers, Television News, and Online News*. Media Management Center, Northwestern University, 2002.
- Appelman, Alyssa, and S. Sundar. “Measuring Message Credibility in the Contemporary Social Media Environment.” *Communication Research*, vol. 44, no. 1, 2016, pp. 59–79.
- Bucy, Erik P. “Media Credibility Reconsidered: Synergy Effects between On-Air and Online News.” *Journalism & Mass Communication Quarterly*, vol. 80, no. 2, 2003, pp.

247–264.

- Flanagin, Andrew J., and Miriam J. Metzger. “Perceptions of Internet Information Credibility.” *Journalism & Mass Communication Quarterly*, vol. 77, no. 3, 2000, pp. 515–540.
- Hovland, Carl I., Irving L. Janis, and Harold H. Kelley. *Communication and Persuasion*. Yale UP, 1953.
- Kiousis, Spiro. “Public Trust or Mistrust? Perceptions of Media Credibility in the Information Age.” *Mass Communication & Society*, vol. 4, no. 4, 2001, pp. 381–403.
- Metzger, Miriam J., and Andrew J. Flanagin. “Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics.” *Journal of Pragmatics*, vol. 59, 2013, pp. 210–220.
- Sundar, S. Shyam. “The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility.” In *Digital Media, Youth, and Credibility*, edited by Miriam J. Metzger and Andrew J. Flanagin, MIT Press, 2008, pp. 73–100.

Chapter 17

Integrating Agentic AI in the Newsroom Processes: Re-imagining News Delivery in India's Pluralistic Society

Chandramouly Pandey,
Research Scholar,
Department of Mass Communication
Rajiv Gandhi (Central) University,
Arunachal Pradesh, India

Abstract

Artificial Intelligence is a talking point in the media and communication sector, known for its algorithmic systems, personalized audience experience, fact-checking etc. In the realm of this 'agentic AI' emerging as a new actor capable of planning, reasoning and customized experience is taking the centre in the newsroom. While, the existing literatures are more focused on how newsrooms are employing AI for news production, meagre attention has been directed towards the potential of agents in news delivery in pluralistic societies. This chapter explores the role of agentic AI in enhancing news production and dissemination process within the context of India which is characterized by linguistic diversity, cultural plurality, regional identities and other challenges. Further, the study tries to understand various ethical challenges involved in the use of agents from the lens of media and journalistic ethics. Adopting a qualitative research design, the chapter synthesizes multi-disciplinary literature to understand how agents can enhance the efficiency and productivity in the newsroom to augment the user experience thereby enabling more inclusive and responsive journalism. Apart from this, semi-structured interviews were conducted with journalists from 12 news organizations to critically examines concerns related to algorithmic bias, editorial accountability, transparency, over-personalization so that freedom of speech and journalistic autonomy can be preserved. Building upon the discussions, the chapter argues that agentic AI should not viewed mere as a tool for newsroom efficiency but as a strategic enabler for strengthening democratic communication in culturally diverse societies. Furthermore, it contributes towards the emerging discourse on AI-driven journalism by extending the discussion beyond automation towards inclusive, pluralistic and ethically

responsible news ecosystem capable of serving India's pluralist and culturally diverse audience in the age of intelligent media.

Keywords: Artificial Intelligence, Agentic AI, Journalism, News, Digital Media

1. Introduction

Artificial intelligence has become embedded in contemporary journalism, transforming not only news production but also how news is distributed, personalised, and consumed. This includes automated writing, data analysis, fact-checking, translation, content recommendation, and audience analytics. AI is shaping several stages of journalists' work. Recent scholarship suggests that AI integration has moved journalism beyond simple task automation toward more complex human-machine relationships in which algorithms increasingly participate in editorial and communicative processes. Researchers have identified applications such as automated writing, content personalization, fact-checking and format transformation as significant areas of AI integration, while simultaneously highlighting concerns regarding algorithmic bias, transparency, privacy and accountability (Gutiérrez-Caneda et al., 2024). In recent years, the emergence of generative and increasingly agentic forms of AI marks another important stage in this transformation. Unlike conventional AI systems designed to perform pre-defined tasks, agentic AI is increasingly characterized by its capacity to interpret goals, reasons through problems, plan actions, use tools and adapt its responses with comparatively greater autonomy (Baird & Maruping, 2021). This development has significant implications for journalism because AI systems may potentially move from assisting journalists with individual tasks towards coordinating multiple stages of news production and distribution. Recent research on AI and journalism similarly points towards a transformation in professional roles and newsroom practices rather than merely the replacement of isolated manual tasks (Park & Akintade, 2026).

While considerable scholarly attention has been devoted to AI-assisted news production, comparatively more attention is now needed for AI-mediated news delivery and audience experience. AI-powered personalization can analyse users interests and behavioural patterns to recommend and adapt news content, potentially making news consumption more relevant and accessible (Lim & Zhang, 2022). Research on AI-driven personalization demonstrates that such systems can influence user engagement and adoption by tailoring news experiences to individual

preferences. Recent audience research further indicates that AI can facilitate personalized formats, summaries, translations, recommendations and conversational access to news (Arguedas, 2025). However, the increasing autonomy in selecting, prioritizing and presenting information raises important questions about editorial accountability, transparency, algorithmic bias and the creation of filter bubbles due to personalized recommendation system (Garcia & Perry, 2026). Personalization may improve relevance while simultaneously narrowing users exposure to diverse viewpoints that fall outside their established preferences.

Moreover, transparency about AI involvement is increasingly important because audience perceptions or credibility and trust can be influenced by how AI-generated or AI-assisted content is disclosed. These concerns acquire particular significance in India, where journalism operates within a highly pluralistic communication environment characterized by linguistic diversity, cultural heterogeneity, regional identities and unequal access to digital technologies.

In such context, Agentic AI could potentially support multilingual news delivery, personalized information access, biased datasets, opaque recommendation systems and excessive personalization could reproduce existing inequalities or restrict exposure to divers perspectives. Thus, the question is not simply whether AI can make journalism more efficient, but how autonomous AI agents can be integrated without compromising the fundamental ethical and democratic responsibilities of journalism. Accordingly, the chapter examines the emerging role of agentic AI in news production and dissemination in India with particular emphasis on its potential to enhance efficiency, responsiveness and audience experience. It further investigates the ethical challenges associated with algorithmic decision-making and over-personalization. By extending the discussion from AI as a tool of newsroom automation to AI as an increasingly active participant in news delivery, the chapter seek to contribute to the emerging discourse on inclusive, pluralistic and ethically responsible AI driven journalism.

2. Literature Review

The evolution of Artificial Intelligence (AI) in journalism can be understood as a gradual movement from automated content production towards increasingly autonomous systems capable of participating in multiple stages of the journalistic process. In one of the empirical contributions by Clerwall (2014) examined the audience perceptions of software-generated news

in comparison with journalists-produced content, where readers struggled with identifying human and machine produced news, while machine generated stories were perceived as relatively objective, although somewhat more descriptive and less engaging. This early work established an important foundation for subsequent research by demonstrating that the implications of automated journalism extend beyond production efficiency to questions of credibility, authorship and audience perception. Guenat et al. (2022) defined Agentic AI to a category of autonomous artificial intelligence systems designed to independently accomplish complex, multi-step tasks over extended periods with minimal or no human supervision. Further, Maia et al. (2020) added that these systems can interpret and learn from contextual information, make decisions and adapt their actions according to changing circumstances. At its core, agentic AI is characterized by autonomy and adaptability, enabling it to undertake and manage goal-oriented processes with greater independence. The emergence of agentic AI represents a further conceptual shift in this technological trajectory. Agbaeze et al. (2025) conceptualize AI agents through four dimensions – Autonomy, efficacy, goal complexity and generability. Rather than treating agents as binary characteristics, AI systems exist along a spectrum depending on how independently they can operate, how significantly they can affect environment, how complex their goal are and how broadly they can perform tasks. This conceptualization is particularly relevant to journalism because an AI system that merely generates a paragraph from journalist prompt as subsequently different implications from an agent capable of searching multiple sources, evaluating information, constructing a new narrative, monitoring development and adapt its output according to changing circumstances.

The literature on AI-mediated news distribution provides another important foundation for understanding the potential of agentic AI. Lim and Zhang (2022) investigated the adoption of AI-driven personalization in digital news platforms through enhanced user engagement and positive attitudes towards AI-powered news. The authors further added that AI-enabled news delivery can move beyond simply making information available towards dynamically adapting news experiences to individual users. Highlighting the ethical implications of agentic AI, Lim et al. (2025) raise an important question about responsibility when in agent causes harm, particularly while independently, riving or processing data and information from third-party sources. Since such system poses a degree of decision-making capability in contextual autonomy and determining responsibility for potential harmful outcomes become increasingly complex.

Building on this concerns Hahn et al. Situated in within established ethical principle of AI governance, emphasizing the importance of transparency. Justice fairness, non-maleficence, responsibility and accountability along with privacy. These principles become particularly relevant when genetic AI operates with limited human intervention as their autonomous action may affect individuals organisations and why does Society evaluation of Agentic AI requires attention, not only to what the system can accomplish, but also how to decisions are made who is accountable for their consequences and how rights and social values can be protected

3. Objective

- To examine the potential role of Agentic AI in enhancing news production and dissemination.
- To Explore how Agentic AI can improve news delivery among India's pluralistic audience.
- To identify the ethical and journalistic implications associated with Agentic AI.

4. Research Methodology

The study of qualitative research design, combining a multidisciplinary literature review with semi structured interviews. Literature components and synthesises existing scholarly research on AI agents, AI journalism news production, algorithmic personalisation, media ethics and democratic communication to examine the emerging role of AI agents in newsrooms and news examination. In addition, semi-structured interviews were conducted with journalists from 12 news organisations to obtain empirical insights into the adoption and potential applications of Agentic AI in news production and delivery. The interviews particularly examined journalists' perceptions and concerns regarding algorithmic bias, editorial accountability, transparency, over-personalisation, human oversight, freedom of speech and general autonomy. The qualitative finding automatically analyse to identify recurring patterns and concerns and interpreted through the broad framework of media and general cystic ethics with particular attention to the implication of Agentic AI for inclusive pluralistic and democracy-responsible journalism in India.

5. Discussion

The emergence of agentic AI represents a significant development in the relationship between technology and journalism. Unlike conventional AI applications that perform specific, pre-defined functions, agentic AI has the capacity to interpret goals, plan actions, interact with multiple information sources, make decisions and execute complex tasks with a degree of autonomy. The findings of the present study indicate that this transition has implications not only for newsroom efficiency but also for journalistic practices, ethical accountability and delivery of news to linguistically and culturally diverse audiences. The discussion is therefore organized around three interconnected dimensions: the role of agentic AI in enhancing news production, its ethical implications and its potential to expand news delivery across Indian languages.

6. Role of Agentic AI in Enhancing News Production

The findings suggest that agentic AI potentially transform news production by moving beyond the automation of isolated journalistic tasks towards the coordination of interconnected newsroom activities. Traditional AI applications have largely been used for tasks such as transcription, translation, data processing, automated writing, headline generation and content recommendation. Agentic AI in contrast, has the potential to connect several of these functions within a single workflow. An agent could for instance, monitor multiple information sources, identify an emerging news event, collect relevant background information, compare sources, identifying an emerging news event, collect relevant background information, compare sources, organize the available evidence, prepare a preliminary news brief and alert a journalist for editorial verification. In the sense, the value of agentic AI lies not simply in generating content but in its ability to coordinate and execute a sequence of goal-oriented tasks.

This possibility is particularly important in contemporary newsrooms where journalists work under considerable time pressure. The increasing volume of digital information requires journalists to monitor social media, official websites, databases, news agencies and other sources simultaneously. An intelligent agent could assist journalists by continuously monitoring these sources and identifying developments relevant to a particular beat. Such functionality could reduce time spent on repetitive information-gathering activities and allow journalists to devote more attention to investigative work, contextualization, field reporting, and editorial judgement. The findings also indicate that agentic AI should be understood primarily as augmented

intelligence rather than a substitute for journalism. Its greatest contribution may emerge when machine capabilities and human journalistic expertise complement one another. AI agents are more task specific and can process large volumes of information at a speed that is difficult for individual journalists to achieve, whereas journalists provide contextual understanding, ethical judgement, source relationships and editorial responsibility. The newsroom of the future may therefore become increasingly collaborative, with journalists delegating certain operational tasks to AI agents while retaining authority over editorial decisions. At the same time, autonomous systems may alter professional roles. Journalists may increasingly become supervisors, verifiers and editors of AI-generated outputs rather than performing every stage of information processing themselves. This does not necessarily diminish the importance of journalism, instead, it potentially shifts the locus of journalistic experts towards verification, contextual interpretation and accountability. However, such a transformation requires adequate literacy among journalists and clearly defined organizational protocols concerning which tasks can be delegated to AI and which requires human intervention.

Thus agentic AI has the potential to improve newsroom productivity, accelerate information processing and support journalists in managing increasingly complex information environments. However, its effectiveness ultimately depends upon the quality of human – AI collaboration and the institutional structures governing its use.

7. Ethical Implications of Agentic AI in Journalism

While the productive potential of agentic AI is substantial, the findings demonstrate that its increasing autonomy introduces a corresponding set of ethical challenges. The central issue is the question of accountability. When an AI agent independently searches for information, select sources, synthesises content or makes recommendations, determining responsibility for an inaccurate or harmful outcome becomes more complicated. Unlike conventional software, an agentic system may make a sequence of decisions not explicitly specified by an individual journalist at the time of execution. Consequently, responsibility cannot simply be attributed to the machine, it must remain embedded within the newsroom's human and organizational structures. Transparency represents another major concern. Audiences should be able to understand when AI has played a significant role in producing or delivering a news story. Lack

of disclosure may undermine audience trust, particularly when AI-generated information contains error, fabricated information or misleading interpretations. Transparency should therefore extend beyond simply stating that AI was used. News organizations may need clearer policies on the extent of AI involvement, the sources an agent accesses and the degree of human verification before publication. The issue of algorithmic bias is particularly significant in the Indian context. AI agents learn from datasets and operates according to models, algorithms and instructions that may reflect existing social, cultural and linguistic biases. If an agent systematically prioritizes information from dominant languages, metropolitan news organizations or particular social groups, it may reproduce existing inequalities within the media system. Such bias could become more consequential when agents are responsible for selecting groups which information audiences encounter. This concern is further closely connected to over-personalization. Agentic AI could construct highly individualized news environments continuously learning from users preferences and behaviour. Although this may improve relevance and convenience, excessive personalization could reduce exposure to diverse viewpoints and reinforce existing beliefs. In a democratic society, journalism performs a social function that extends beyond satisfying individual preferences. Citizens also requires exposure to information that they may not actively seek but which is socially and democratically significant. Consequently, agentic news delivery must balance socially and democratically significant. Consequently, agentic news delivery must balance personal relevance with editorial diversity and the public interest.

These concerns indicate that human oversight cannot be treated as an optional safeguard. Instead, a human-in-the-loop mechanism should remain central to newsroom deployment. Particularly for sensitive stories, politically consequential information, allegations against individuals and breaking news, AI agents should assist rather than independently determine publication decisions. The ethical principle should therefore be that greater machine autonomy must be accompanied by greater institutional accountability and stronger mechanisms of human oversight.

8. Agentic AI and News Delivery in different Indian Languages

One of the most significant possibilities identified in the study is the capacity of agentic AI to transform news delivery across India's extensive linguistic landscape. India is characterized by substantial linguistic diversity, with news audiences consuming information through multiple regional and vernacular languages. However, the digital news ecosystem remains uneven in terms the availability, quality and timeliness of content across the languages. Agentic AI could could potentially address some of these limitations through automated translation, summarization, localization and conversational news delivery. An AI agent could monitor news development in one language, retrieve information multiple sources, translate and contextualize the information into another language and generate a format appropriate for a regional audience. This could significantly producing separate versions manually for every language, news organization could employ AI agents to facilitate the rapid adaptation of stories across languages while journalists retain responsibility for linguistic and contextual verification.

The potential significance of this capability extends beyond translation. Effective multi-lingual journalism requires localization rather than word for word conversion. Agentic systems could potentially incorporate regional context, terminology and audience preferences while producing language-specific versions of news content. This could strengthen the visibility of regional perspectives within the broader national information ecosystem. Agentic AI could also enhance accessibility through conversational interfaces. Audiences could potentially interact with an AI news agent in their preferred language, ask follow-up questions, request summaries or seek clarification about complex events. Such systems could make news more accessible to audiences who may face barriers to conventional text-based news consumption. However, multilingual AI also presents substantial risks. India's languages vary considerably in terms of available digital datasets and technological resources. High-resource languages may receive substantially better AI performance than low-resource and tribal languages. Poor-quality translation, culturally inappropriate terminology or loss of local meanings could consequently create new forms of linguistic exclusion. Moreover, if AI systems rely predominantly on content generated by major by major national or metropolitan media organization, regional narratives may continue to remain underrepresented despite the availability if multilingual delivery. Therefore the contribution of agentic AI to Indian journalism should not be measured merely by the numbers of languages into which news could be translated. The more important question is whether these technologies can facilitate equitable representation, linguistic inclusion and access to diverse

perspectives. This requires investment in vernacular daatsets, locally relevant AI models, linguistic expertise and human verification mechanism.

Overall, the findings suggest that agentic AI could contribute to a more efficient, responsive and linguistically inclusive Indian news ecosystem, but only when a technological autonomy is balanced with journalistic judgement and ethical governance. Its democratic value will ultimately depend not on how independently an AI agent can operate but on whether that autonomy strengthens rather than weakens the core functions of journalism. In this respect, agentic AI should be conceptualized not merely as another newsroom technology, but as an emerging participant in the communication process whose deployment requires a careful balance between innovation, human agency, ethical responsibility reuires a careful balance between innovation, human agency, ethical responsibility and democratic values.

9. Conclusion

The study concludes that agentic AI represents an important shift in the evolving relationship between AI and journalism, moving beyond task-specific automation towards systems capable of planning, reasoning, adapting and executing interconnected journalistic tasks. The findings suggest that systems can enhance newsroom efficiency by supporting information gathering, source monitoring, content synthesis, translation, verification and audience engagement, thereby enabling journalists to focus more on contextualization, investigative reporting and editorial judgement. However, increased autonomy also introduces significant ethical concerns relating to algorithmic bias, accountability, privacy, over-personalization and the preservation of journalistic autonomy. These challenges show that human oversight must remain central to deploying agentic AI in newsrooms. The study further highlights the considerable potential of agentic AI for multi-lingual news delivery in India. By facilitating translation, localization, summarization and conversational access to information among linguistically disparities in linguistic datasets and the risk of culturally inappropriate or inaccurate translation require careful human supervision. Overall, agentic AI should therefore be viewed not merely as a technological tool for newsroom efficiency, but as a socio-technical actor whose responsible integration can contribute to a more inclusive, pluralistic and democratically responsive Indian news ecosystem.

Reference

- Agbaeze, S. E., Okeke, V. C., Phillips, E. P., Jacobs, A. L., & Oluwakoya, S. T. (2025). Agentic AI in Newsrooms: Towards a multi-dimensional framework for evaluating trust, editorial accountability, and workflow quality. *World Journal of Advanced Research and Reviews*, *28*(2), 1061–1080. https://doi.org/10.30574/wjarr.2025.28.2.3766

- Arguedas, A. R. (2025, June). *How audiences think about news personalisation in the AI era*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/how-audiences-think-about-news-personalisation-ai-era?utm_source=chatgpt.com
- Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artefacts. *MIS Quarterly*, *45*(1), 315–341. https://doi.org/10.25300/misq/2021/15882
- Clerwall, C. (2014). Enter the robot journalist. *Journalism Practice*, *8*(5), 519–531. https://doi.org/10.1080/17512786.2014.883116
- Garcia, I. P., & Perry, S. D. (2026). BrEAKING NeWS: Examining the impact of AI on audience perceptions of journalism. *News Research Journal*, *47*(2), 173–192. https://doi.org/10.1177/30497841261428594
- Guenat, S., Purnell, P., Davies, Z. G., Nawrath, M., Stringer, L. C., Babu, G. R., Balasubramanian, M., Ballantyne, E. E. F., Bylappa, B. K., Chen, B., De Jager, P., Del Prete, A., Di Nuovo, A., Ehi-Eromosele, C. O., Torbaghan, M. E., Evans, K. L., Fraundorfer, M., Haouas, W., Izunobi, J. U., . . . Dallimer, M. (2022). Meeting sustainable development goals via robotics and autonomous systems. *Nature Communications*, *13*(1), 3559. https://doi.org/10.1038/s41467-022-31150-5
- Gutiérrez-Caneda, B., Lindén, C., & Vázquez-Herrero, J. (2024). Ethics and journalistic challenges in the age of artificial intelligence: talking with professionals and experts. *Frontiers in Communication*, *9*. https://doi.org/10.3389/fcomm.2024.1465178
- Hahn, M., Tretter, M., & Dabrock, P. (2026). Ethical perspectives on AI Agents and Agentic AI. *AI And Ethics*, *6*(2). https://doi.org/10.1007/s43681-026-01027-0
- Lim, J. S., & Zhang, J. (2022). Adoption of AI-driven personalization in digital news platforms: An integrative model of technology acceptance and perceived contingency. *Technology in Society*, *69*, 101965. https://doi.org/10.1016/j.techsoc.2022.101965

- Maia, E. H. B., Assis, L. C., De Oliveira, T. A., Da Silva, A. M., & Taranto, A. G. (2020). Structure-Based virtual screening: from classical to artificial intelligence. *Frontiers in Chemistry*, *8*, 343. https://doi.org/10.3389/fchem.2020.00343
- Park, C. S., & Akintade, T. (2026). Artificial intelligence in Nigerian newsrooms: Navigating jurisdictional hybridization, retreat, and maintenance. *New Media & Society*. https://doi.org/10.1177/14614448261429426
- Simon, F. M., Amy, R. A., & Eggert, L. (2025). Reckoning with AI agents. *Oxford University Research Archive (ORA) (University of Oxford)*. https://doi.org/10.5287/ora-ppgrbwpva

Chapter 18

Human Oversight and Control in the Era of AI-enabled Targeting

By Haseeba Sayyed

Abstract

In July 2026, an Indian court was hearing a challenge to the use of facial recognition tools against student protestors in Delhi. This is happening months after Indian and Pakistani forces claimed that they used AI-assisted precision targeting during a border conflict; this is also the same year when a US-Israeli campaign against Iran used generative AI for target selection. These are three different applications of Artificial Intelligence tools. They used algorithm-based identification and targeting of entities, and deployed surveillance data across countries' constitutional law, international humanitarian law, and the largely voluntary norms governing military AI diplomacy.

This chapter examines this using three case studies, including but not limited to India's Jantar Mantar surveillance litigation, Operation Sindoor, and the ongoing US-Israel-Iran military conflict. These case studies were chosen for their contemporaneity and their structural similarity. In each of these, a human decision-maker became a nominal figure in the loop. At the same time, the substantive work of identification, targeting, and (often) risk assessment is being delegated to a system whose error rates and internal reasoning remain undisclosed and within reasonable doubt. These cases expose a shared failure: neither International Humanitarian Law's proportionality doctrine nor domestic privacy jurisprudence is currently equipped to address it alone. It shows how the reduction of human oversight can become just "a stamp of approval."

From this comparison, the chapter derives accountability conditions that current frameworks (the United Nations' Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS), Responsible AI in the Military Domain (REAIM), and India's absent facial recognition statute do not yet jointly enforce:

1. Disclosed error thresholds before a system's output can justify state action,
2. An auditable record independent of the agency deploying the system,
3. And a single standard for "meaningful human control" applied consistently regardless of who (or what) the subject is.

The chapter discusses this standard against all the aforementioned cases, including India's own conduct in both Sindoor and Jantar Mantar, to argue that a workable international norm cannot be built if it applies more to adversaries than to the state proposing and using it.

Keywords: Artificial Intelligence, Military conflict, Facial Recognition, Algorithm, Human control

1. Introduction

In the past year, how many times have you been scared of the fact that you can be replaced by an AI? Too many, right? I have been too. Be it a writer, an engineer, a photographer, or even a researcher, all of us have been, at least once in the past couple of years, down the spiral where we have thought that we are replaceable by AI. But you and I are not the only ones who have dealt with this threat. A person who is standing in a military war room, where all the strategy is being made, and one of the biggest decisions about somebody's life or death is being debated, has also been under threat. Right now, the world is witnessing turmoil where multiple nations are clashing. To name only a few, and not even considering the civil wars happening within countries, right now, the Russia-Ukraine clash, along with the US-Israel-Iran conflict, have all been in the headlines of the international section of every newspaper that cares about what's happening in the world. While I am writing this, next to me lies an issue of Frontline magazine by The Hindu, and on the front page, I see the words read, “Hormuz: No Strait Solutions,” coming back to the point, right now. Many news publications have reported on the use of AI in decision-making and strategy in the ongoing wars.

A machine is not yet pulling a trigger or signing an arrest order, but the boom in Artificial Intelligence surely has changed the way guns are being pointed, and targets are being marked. The meaning of human responsibility has changed, and unfortunately it only defines the reduction of responsibility and the number of people and hours involved in decision-making.

For machines, changing the human sense of accountability does not require pulling a lever or issuing an arrest command. Before reaching that stage, a machine can already identify what

face is classified as a match, what item is classified as a threat, what pattern is seen as suspicious, and what question is removed from the consideration. Especially, what pattern is removed from consideration? The moment the police or army commander is approached with a specific inquiry, the decision might already be narrowed down by an algorithm that is hard to evaluate, unquestionable and even more difficult to legally contest.

This chapter discusses this issue. The theme of increasing use of AI has been a topic of discussion in the academic circles studying militarization of AI. The use of AI is not only restricted to imaginations and VFX in superhero movies. Its happening in real time. The only difference, it's not always a villain on the other side of the target. Sometimes its innocent civilians and citizens.

AI has many functions. Facial recognition, data fusion, drone imagery, or generative systems are just a few. AI systems are being deployed to perform specific tasks in different environments: recognize, categorize, order, forecast, and advise. Institutions are using the technology and justify its use by citing the principle of the "human in the loop" to show that accountability is intact even though it’s non-existent. The important question is whether the human considered as the element of the loop has full control over the matter or is just authorizing the end result without being able to comprehensively evaluate it.

The chapter examines three instances: the 2026 Jantar Mantar surveillance; India’s military action against Pakistan, Operation Sindoor; and the application of artificial intelligence in military operations in the USA-Iran-Israel conflict. These case studies were selected on the grounds of their recency; and the fact that they highlight the same issue across various forms of power employed by states in conducting their affairs. The problem examined in the Jantar Mantar is that of transformation of innocuous assembly into a database of biometric information and weaponisation of civilian dissent against themselves. step further.

The argument made in this chapter is that the true ethics of AI cannot be realized by simply signing the paper at the end of the algorithmic process. The ethics of AI requires three important conditions to be fulfilled: first, independent verification of error thresholds; second, availability of records that can be verified by persons who are not representatives of the operator agency; and third, an autonomous body functioning on an international level that can govern this field.

2. In India, Protest Becomes a Database and Protestors, Target

The reaction of the armed forces during the Jantar Mantar protest in July 2026 threw the issue into sharp focus. Reuters reported that the incident began during a protest with students. The protesting students showed pictures of a mobile surveillance van belonging to the police which was fitted with several cameras that could capture the protesters in 360 degrees (Reuters, 2026). Aishe Ghosh, who had served as president of the JNU Students' Union, sought a ruling that the surveillance of the protesters was unconstitutional, along with the destruction of the data obtained during the protest and a set of guidelines for the usage of such technology in future (Reuters, 2016).

The legal dispute was not limited to the issue of a camera in a public place. In accordance with the reports about the petition, it affected the case of "constant, indiscriminate and intrusive surveillance" of peaceful protesters, with the publications mentioning that monitoring, photography, and videos were taken from protesters without any given statutory authority and procedural safeguards (Indian Express, 2026). The petition also highlighted the problem of women being filmed during heavy rain and in vulnerable physical conditions, making it so that the issue was not only about privacy but also about their bodily dignity and autonomy (Indian Express, 2026).

A petition from MP A. A. Rahim claimed to have identified three methods of biometric monitoring which involve indiscriminate photo and video recording; recognition of facial features in real-time through Ikshana surveillance car and AjnaLens glasses; and collection of fingerprints with the NCRB's Abhigyan app and matching them with the National Automated Fingerprint Identification System (LawBeat, 2026). The petition claimed that this practice operates in "complete legal vacuum"; quoting responses under RTI Act, it stated that no privacy impact assessment was undertaken, that 80% similarity score was taken as a match without public validation and that the data can be stored indefinitely or included in criminal records without any limitations (LawBeat, 2026). These are allegations made in an ongoing legal case.

According to reports, the stance taken by the State is greatly different. The authorities justified the act of spying as legitimate and resulting in the maintenance of law and order. The government's attorney is believed to have made statements about the quest for privacy in public venues that can be called "ironical," as it was said that the acts of monitoring "are

absolutely required" and that the right to interference in the private life of people is "fully justified by the public interests" (Reuters, 2026). Furthermore, the authorities in Delhi Police noted on their Facebook profile that no limitations have been made in terms of people's freedom due to the introduction of the facial recognition system. According to one of the police officers, computer equipment assisted in recognizing offenders and counting demonstrators (Reuters, 2026).

The argument cannot be decided simply by asking if police can document a crowd. Documentation and biometric identification are two different processes. A video can document an incident. On the other hand, a facial-recognition system generates a biometric template, compares it against the database, arrives at a potential match, and connects the individual to further evidence.

The above means that the normal act of being in a public space turns into an official permanent record that can be searched. This is due to the fact that active facial recognition is fundamentally different from standard video recording, as mapped facial data that is provided with a database can suspect everyone being in sight, rather than focusing only on the person under investigation (India Today, 2026).

The constitutional implications are urgent. The petitions filed invoke Articles 14, 19, and 21: the right to equality and protection from arbitrariness, the right to freedom of expression and peaceful assembly, and the right to life, liberty, privacy, and dignity (Indian Express, 2026; LawBeat, 2026). The judgment of the Supreme Court in Justice K.S. Puttaswamy v. Union of India is instrumental in determining whether surveillance can pass the tests of legality, legitimate aim, necessity, and proportionality (Lawful Legal, 2025). Thus, a generic statement of public order is insufficient to address the issues raised by the real-time one-to-many biometric scanning of peaceful protesters.

The key constitutional aspect is that harm does not require any immediate detainment. A student may participate in an entirely legal protest, return home, and never find out if her face has been photographed, identified, registered, or stored. That uncertainty is capable of changing social behavior in itself. Those at the beginning of their careers may understandably be worried that their involvement in protests might be recorded and improperly used by the authorities. The chilling effect is not just a rhetorical statement; it is a reality caused by the ability to process information about people without them knowing about it. And all of this is

happening with the help of AI enabled technologies. AI-enabled systems were deployed against citizens who dared to register their dissent.

The issue is aggravated by the fact that the reliability of technology is unknown. As stated in the research material, the police of Delhi had published very low accuracy levels in previous experiments: 2% in 2018 and less than 1% in 2019 (Internet Freedom Foundation). It should be noted that these figures may refer to particular cases of application and cannot be automatically extended to more recent technologies. Nevertheless, they help to visualize how similarity scores cannot be considered to be absolute evidence. The effectiveness of facial recognition depends upon many factors, including the quality of the photo, lighting, the number of people in the frame, the camera angle, relevant databases, and the population the technology was working with. Therefore, the 80 percent matching ratio (if applied) will not prove that the person is guilty; it is just a matter of the fact that the technology has found some similarity in accordance with its own built-in limitations.

Wider research on facial recognition confirms the concerns; one study cited in the material states that law enforcement systems have a higher likelihood of errors with darker-skin and woman faces. Specifically, 0.8% of white male faces show a possibility of error while dark skin female faces have error rates as high as 34.7% in the data sets (Facial Recognition Technology, 2021). The relevance for India should not be aligned with the overseas categories. In this research it is mentioned that the inaccuracies in facial recognition may also correlate with other socially represented categorizations such as caste, class, gender and many others, as the infrastructure and environment for development of such either does not offer equal opportunities or does not ensure equity (Thomas, 2020).

There is an even more serious issue to be taken into consideration: function creep. In Sadhan Haldar v. State (NCT of Delhi), facial recognition technology was used to fulfill the necessary purpose to find missing children (Internet Freedom Foundation). However, the technology can also be used for different reasons such as screening marches for protests, identifying protesters,

References

- Amnesty International, Access Now, & other signatories. (2026, June). Global: Joint statement on AI in warfare.https://www.amnesty.org/en/wp -

-

content/uploads/2026/06/AMR0411462026ENGLISH.pdf

- Ojha, Y. (2025). Artificial Intelligence in Armed Conflict: Perspectives from International Humanitarian Law. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5162209
- Eastes, I. (2024b). Artificially Armed: The legal implications and limitations of AI-enabled autonomous weapons systems on international humanitarian law and human rights law. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4910932
- Dorsey, J., & Moffett, L. (2026, May 13). Warification and the illusion of precision: AI, targeting, and increasing civilian harm. Lieber Institute, West Point. https://lieber.westpoint.edu/warification-illusion-precision-ai-targeting-increasing-civilian

1. -harm/

- Facial recognition technology: A survey of policy and ethical issues. (2021). PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC8475322/
- Horowitz, M. C., & Scharre, P. (2015). Meaningful human control in weapon systems: A primer. Center for a New American Security. https://www.files.ethz.ch/isn/189786/Ethical_Autonomy_Working_Paper_031315.pdf India Today. (2020, February 18). Delhi, UP police use facial recognition tech at anti-CAA protests; others may soon catch up.
- https://www.indiatoday.in/india/story/delhi-up-police-use-facial-recognition-tech-at-anti-caa-protests-others-may-soon-catch-up-1647470-2020-02-18
- India Today. (2026, July 27). Facial recognition at CJP protest raises privacy and data protection questions.

1. https://www.indiatoday.in/india/law-news/story/facial-recognition-cjp-protest-privacy-data-protection-questions-2957404-2026-07-27

- Indian Express. (2019, December 30). Police film protests, run images through face recognition software to screen crowd. https://indianexpress.com/article/india/police-film-protests-run-its-images-through-face-r ecognition-software-to-screen-crowd-6188246/
- Indian Express. (2026, July 25). Why police surveillance and facial recognition at Jantar Mantar is being challenged in Delhi HC. https://indianexpress.com/article/explained/explained-law/jantar-mantar-surveillance-delh

i-hc-pil-facial-recognition-10802731/

- Internet Freedom Foundation. (n.d.-a). Is the illegal use of facial recognition technology by the Delhi Police akin to mass surveillance? You decide: Project Panoptic. https://internetfreedom.in/is-the-illegal-use-of-facial-recognition-technology-by-the-delhi

1. -police-akin-to-mass-surveillance-you-decide-project-panoptic/

- Internet Freedom Foundation. (n.d.-b). We demand the Delhi Police stop its facial recognition system.

1. https://internetfreedom.in/we-demand-the-delhi-police-stop-its-facial-recognition-system/

- International Committee of the Red Cross. (1977). Article 57: Precautions in attack. https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-57
- Kozlowskyj, C. (2026, March 24). Iran conflict as a testing ground for AI warfare systems. British International Studies Institute. https://bisi.org.uk/reports/iran-conflict-as-a-testing-ground-for-ai-warfare-systems
- LawBeat. (2026, July 28). PIL in Supreme Court challenges Delhi Police's facial recognition surveillance at Jantar Mantar protest. https://lawbeat.in/top-stories/pil-in-supreme-court-challenges-delhi-polices-facial-recogni tion-surveillance-at-jantar-mantar-protest-1616723
- Lawful Legal. (2025, September 22). Supreme Court's perspective on privacy in the era of AI and facial recognition. https://lawfullegal.in/supreme-courts-perspective-on-privacy-in-the-era-of-ai-and-facial-r ecognition/
- Oxford Human Rights Hub. (2020, March 11). Use of facial recognition technology in India A function creep breaching privacy.

Chapter 19

From AI Adoption to Organizational Performance: The Mediating Role of AI-Driven Managerial Decision-Making

Dr. Pratibha Saini,
Assistant Professor,
Department of Management,
Lingayas Lalita Devi Institute of management & Sciences, New Delhi,
Pratibha.saini@lldims.org.in

Abstract

The rapid spread of Artificial Intelligence (AI) is changing business and management practices. It is altering how organizations process information, make decisions, connect with customers, and use their resources. Even with the growing use of AI applications, there is still a limited understanding of how AI adoption leads to better decision-making and improved performance, especially in emerging economies like India. This chapter aims to explore the link between AI adoption and effective management. It will look into how AI capabilities can improve decision quality, operational efficiency, innovation, and overall organizational performance. Using the Technology-Organization-Environment (TOE) framework and the Technology Acceptance Model (TAM), the study will present a conceptual framework which considers AI adoption as the main factor, AI-driven decision-making capability as a mediating factor, and organizational performance as the end result. It will also take into account organizational readiness, employee AI skills, and perceived ethical and privacy risks as contextual factors affecting AI adoption and its success. The study may use a quantitative research design with structured questionnaires sent to managers and business professionals across various sectors. Analysts can evaluate the data using descriptive statistics, reliability and validity checks, correlation analysis, regression, and Structural Equation Modeling (SEM) to assess the proposed connections. This chapter aims to add to the growing literature on AI-driven management. It will provide a picture of the technological, organizational, and human aspects of AI adoption. The findings should offer valuable insights for managers, policymakers, and organizations looking to build AI capabilities while ensuring responsible, transparent, and people-focused decision-making practices.

Keywords: Artificial Intelligence, AI Adoption, Managerial Decision-Making, Organizational Performance, AI Capability, Technology Acceptance Model, TOE Framework, Digital Transformation, Responsible AI, Human-AI Collaboration

1. Introduction

AI has changed from being mainly a technological idea to now being a significant part of current business strategy. Companies in all different sectors are using it more and more for data analysis, forecasting, customer relationship management, fraud detection, marketing personalisation, human resource management, supply chain optimisation, financial analysis, automation, and for providing strategic decision support.

This change is especially clear in India. A backgrounder issued by the Indian Government in February 2026 noted that the country came third in Stanford University's 2025 Global AI Vibrancy Ranking and indicated that 87 per cent of enterprises were actively using AI solutions, according to the NASSCOM AI Adoption Index. At an organizational level, the shift is also becoming more strategic. A report by EY-CII, published in November 2025, found that 47 per cent of the Indian enterprises surveyed had several AI applications running in production, with operations, customer service and marketing being the main areas of focus.

It does not follow that because AI becomes more widely available organisations will automatically perform better.

The fundamental managerial question is therefore no longer:

“Should organizations adopt AI?”

Instead, it increasingly becomes:

What is the way in which organizations can turn their adoption of AI into real managerial abilities and tangible value for the organization?

It is important to make this distinction since the same AI technology can lead to different results in various organizations. For example, one organization might use AI just for automating routine tasks, whereas another could use it in the areas of forecasting, strategic planning, customer analytics, risk management, and managerial decision-making.

AI adoption and AI value creation should not be regarded as the same thing.

Similar research also shows that the use of AI is linked to improvements in efficiency and decision-making, but that the results will vary according to the organization's readiness, the level of trust, strategic alignment, the skills available, and the organization's capability to integrate AI into its current processes.

The chapter focuses on the pathway:

AI Adoption → AI-Driven Decision-Making Capability → Organizational Performance

It also states that this relationship is affected by technological, organizational, environmental, and human factors.

1.1 Background of Artificial Intelligence in Business

Artificial Intelligence (AI) has evolved from a niche area of computer science to one of the most important technologies impacting business and management today. Previous artificial intelligence systems were mainly built to perform rule-based and repetitive tasks. With the advancement of machine learning, deep learning, natural language processing, computer vision, cloud computing and big data analysis, AI has become proficient in carrying out more and more complex activities associated with prediction, pattern recognition, language processing and decision support. Therefore, organisations are gradually moving toward intelligent systems that learn from data and provide suggestions for managerial action as opposed to traditional information processing systems. So, the creation of standard AI may be considered a transition from automation to intelligence.

It can be considered as the evolution of AI from automation to intelligence. The difference is that traditional automation works from specific instructions. Artificial intelligence systems can recognise patterns in large data sets and adapt their output based on experience and the data they have available . This distinction is especially important for modern organisations because business environments generate large amounts of structured and unstructured data every day.

1.2 AI as a Business Strategy Technology

Artificial intelligence is no longer seen as merely an IT tool. It is more and more being recognised as a strategic business technology that can impact organisational efficiency, innovation, customer relations, risk management and competitive advantage. Organisations are using AI to cut processing time, improve forecasting, personalise the customer experience, automate routine activities, and support complex managerial decisions.

For example, an online retailer might analyse a customer’s previous purchases, browsing habits, search history and reaction to promotional offers to forecast what products the customer might be

interested in buying. In the same way, a financial institution might analyse transaction patterns to identify potentially fraudulent activities. In both instances, AI turns large piles of data into valuable information that helps organisations to respond more quickly and accurately.

1.3 AI and Customer Experience

AI is shifting businesses from decision-making based on experience to decision-making based on data and predictions. Customer relationship management is a key example. AI can help companies understand customer preferences, customer segmentation, personalize communication, offer recommendations, and provide automated customer support.

Recommendation systems used in e-commerce and in the video and audio streaming services are the same, but differ in the context of what they are used to recommend. Online stores recommend products based on the customer's previous searches and purchases. Streaming services recommend videos and audio based on a customer's previous engagements with the service. AI-based chatbots are also used to eliminate the wait for human assistance in responses to everyday customer queries on order status and product information and appointment scheduling.

1.4 AI in Marketing and Sales

AI has transformed marketing by providing organizations the ability to analyze customers at a crisper level. With traditional marketing, customers were analyzed along demographic lines. With AI, customers can be analyzed at the level of behaviors. For Example ,the launch of a new smartphone. AI can identify potential upgraders. The AI system may look at prior purchases, internet browsing, how the user interacts with ads, and their level of engagement on the internet. Using the information, marketing managers no longer need to blast promotions on smartphones to the entire customer base.

AI has the potential to support marketing through dynamic pricing, sales forecasting, lead scoring, personalized advertising, sentiment analysis, and prediction of customer churn. Increasingly, marketing is becoming more data driven and predictive.

1.5 AI in Finance and Banking

AI in finance is used heavily because financial services and products are information intensive and time sensitive. AI has the potential to be used in fraud detection, credit analysis, risk management, investment analysis, customer service, and compliance.

By using the above example of the fraud detection system. Let's assume a customer's credit card

or debit card has been recently used to purchase many large value transactions scattered around a wide geographic area. A fraud detection system based on AI may determine the transactions are not normal and therefore represent a potential fraud case. Again, the goal of AI is not automation but to help banking and finance employees prioritize.

1.6 AI in Operations and Supply Chain Management

AI is essential in operations and supply chain management. Many decisions related to operations, such as demand forecasting, planning, inventory management, and logistics require a depth and breadth of data and a degree of uncertainty. Therefore, these are good candidates to leverage the support of AI.

A toy store may leverage AI to monitor the sales of umbrellas during rainy season and may even predict the sales of toys during holiday season. This would help the company avoid both over and under stocking scenarios. Across the logistics chain AI can analyze multiple factors including traffic, available vehicles, delivery schedules, and fuel efficiencies to provide an optimal delivery route. This would cut down costs and improve efficiencies.

1.7 AI and Human Resource Management

HRM is also a much-noted area of focus when discussing AI. Employee management activities ranging from recruitment to training and development to monitoring employee turnover are easily automatable. AI can assist and, in many instances, outperform the traditional methods for the stated employment management activities.An AI system can screen thousands of job applications for an HR manager and can even identify reasons for high employee turnover.

However, AI in HR is also an example of a scenario where technology for its own sake is not enough and a strong framework and management style should also accompany the technology. If bias exists in the recruitment data that an AI system will be trained on, the system may learn to recruit in a bias manner.

1.8 AI Adoption: Beyond Technology Implementation

AI adoption should not simply be understood as purchasing or installing AI software. AI adoption refers to the organizational decision, implementation and continued use of AI-enabled technologies to support or transform organizational activities and processes.

1.9 AI Adoption as an Organizational Capability

The strategic importance of AI does not come only from the technology itself. Competitors can often purchase similar AI tools. What is more difficult to replicate is the organization's ability to combine AI with high-quality data, employee knowledge, managerial expertise, organizational processes, and business strategy.

For example, two banks may purchase similar AI-based fraud detection systems. One bank may have better-quality customer data, skilled analysts, integrated information systems, and effective governance mechanisms. It may therefore obtain greater value from the same technology.

This suggests that:

AI technology + Data + Human Skills + Processes + Managerial Capability = AI-enabled organizational capability

The Resource-Based View provides a useful explanation of this relationship. From this perspective, AI can become valuable when it is combined with organizational resources and capabilities that enable the firm to use it effectively.

2. AI-Driven Managerial Decision-Making

2.1 Concept

Managerial decision-making refers to the process through which managers identify problems, collect information, evaluate alternatives, select courses of action, implement decisions, and monitor results.
AI-driven managerial decision-making occurs when managers use AI systems to support one or more stages of this process.
AI can assist managers by:

- Identifying patterns;
- predicting future outcomes;
- detecting anomalies;
- generating scenarios;
- evaluating alternatives;
- recommending actions;
- automating routine decisions;
- providing real-time information; and
- supporting strategic analysis.

Importantly, AI-driven decision-making does not necessarily mean that machines completely replace managers.

A more realistic model is:

AI provides analytical intelligence; managers provide contextual judgment, ethical reasoning, accountability, and strategic interpretation.

2.2 Transformation of Managerial Decision-Making

There are many ways Artificial Intelligence (AI) can facilitate efficient business practices, we have traditionally relied on experience and intuition, reports, market research, and history when making business decisions. Progressive organizations rely on an integration of these factors with added real-time data, predictive models, and recommender systems.

For example, a traditional sales manager would analyze the previous year's performance to develop expectations for sales target achievement in the upcoming quarter. An AI system looks beyond this sales record to analyze sales history in the context of seasonality, customer preferences, competitor activity and price/service offerings, and other factors to develop predictions on target demand, and identify underlying business opportunities and potential risk exposures. While the responsibility of the final decision remains on the manager, the AI system supplements the available data for the analysis.

2.3 Theoretical Foundations

AI adoption is influenced by more than the availability of technology. An organization may have access to an advanced AI system but may still struggle to use it effectively because of limited resources, employee skills, management support, or regulatory concerns. To understand these factors, this study uses two complementary perspectives: the **Technology-Organization-Environment (TOE) Framework** and the **Technology Acceptance Model (TAM)**.

The TOE framework explains AI adoption from an **organizational perspective**, while TAM focuses on how **employees and managers perceive and accept technology**. Together, they provide a broader explanation of why organizations adopt AI and why some employees use it more effectively than others.

3. Technology-Organization-Environment (TOE) Framework

The TOE framework explains technology adoption through three major contexts:

Technology + Organization + Environment → Technology Adoption

In the context of AI, these three dimensions help explain whether an organization is prepared and motivated to introduce AI into its business activities.

3.1 Technological Context

The technological context refers to the characteristics and capabilities of the AI technology itself. Important factors include **relative advantage, compatibility, complexity, infrastructure, data availability, and security**.

The most important question for managers is often: **"What advantage will AI provide over our existing method?"**

For example, a retail company using traditional spreadsheets to forecast sales may adopt AI if the AI system can analyse customer behaviour, seasonal trends and previous sales to predict demand more accurately. If this reduces excess inventory and stock-outs, the organization sees a clear **relative advantage**.

Similarly, AI should be compatible with existing systems. A bank may find it easier to adopt an AI fraud-detection system if it can be integrated with its existing banking and transaction systems.

Thus, the technological context essentially asks:

Is the AI technology useful, suitable, secure and technically feasible for the organization?

3.2 Organizational Context

The organizational context focuses on the organization's internal ability to adopt and use AI. Important factors include **organizational readiness, financial resources, technical expertise, organizational culture, top-management support, employee skills, and readiness for change**.

AI adoption requires much more than purchasing software. An organization needs:

Technology + People + Data + Processes + Leadership

For example, a manufacturing company may install an AI-based predictive maintenance system to identify when machinery is likely to fail. However, the system will have limited value if employees do not know how to

4. Technology Acceptance Model (TAM)

The **Technology Acceptance Model (TAM)**, proposed by Davis (1989), explains why individuals accept and use a technology. It focuses mainly on two perceptions:

4.1 Perceived Usefulness

It refers to the belief that using a technology will **improve job performance**.

Example:
A sales manager uses an AI forecasting tool because it helps predict customer demand more accurately and supports better sales decisions.

"AI will help me perform my job better."

4.2 Perceived Ease of Use

It refers to the belief that a technology is **easy to understand and use without excessive effort**.

Example:
If an AI dashboard provides simple graphs and clear recommendations, managers are more likely to use it regularly.

"I can use AI without much difficulty."

Thus, when employees perceive AI as **useful and easy to use**, their willingness to accept and use it is likely to increase.

5. Why Combine TOE and TAM?

TOE and TAM explain AI adoption from **two different perspectives**.

TOE explains whether an organization is ready and able to adopt AI, while TAM explains whether its employees and managers are willing to accept and use AI.

Example

Consider a bank introducing an **AI-based fraud detection system**.

- **TOE:** The bank needs good data, IT infrastructure, financial resources, management support, and regulatory compliance.
- **TAM:** Employees must believe that the system is useful and easy to use.
- **Outcome:** When both organizational readiness and employee acceptance are present, AI is more likely to be effectively adopted and used.

Therefore, combining **TOE and TAM** provides a more comprehensive understanding of AI adoption by connecting **organizational readiness with human acceptance of technology**.

6. Empirical Evidence from Recent Research

Recent research increasingly supports the idea that AI's organizational impact occurs through intermediate organizational mechanisms.

A 2024 Management Decision study involving 230 organizations found positive relationships between AI capability, decision-making speed, decision quality, and organizational performance. The research also identified partial mediation through decision-making processes.

Another 2024 study of 346 managers in the United States reported that AI-driven decision-making was positively related to firm performance, while big-data-powered AI positively influenced AI-driven decision-making.

Research published in 2025 similarly found that AI capabilities enhance decision-making performance, particularly decision speed, while also emphasizing the importance of digital capabilities and digital business orientation.

Further evidence suggests that AI adoption can generate business value through multiple complementary mechanisms. Research involving 448 European organizations found that AI adoption influenced organizational performance through decision-making and business-process performance, with process automation, organizational learning, and process innovation acting as additional mediating mechanisms.

This evidence suggests that the relationship between AI and performance is better represented as a **value-creation chain** rather than a simple technology-performance relationship.

Below is a more concise, book-chapter-friendly version with clear headings and short explanatory paragraphs.

7. Managerial Implications

7.1 Align AI Adoption with Business Objectives

Organizations should adopt AI with clear business objectives rather than treating it simply as a technological investment. Managers should identify areas where AI can improve efficiency, reduce costs, enhance customer experience, or support better decisions. For example, AI-based demand forecasting can help retailers reduce stock-outs and excess inventory.

7.2 Strengthen AI-Driven Decision-Making

The value of AI depends largely on how effectively its insights are incorporated into managerial decisions. AI can analyse large volumes of data and generate predictions or recommendations, while managers should interpret these outputs in the context of business conditions and strategic objectives.

7.3 Promote Human–AI Collaboration

AI should complement rather than completely replace managerial expertise. While AI provides analytical capabilities and predictive insights, managers contribute experience, contextual understanding, creativity, ethical judgment, and accountability. Effective human–AI collaboration can therefore improve the quality and reliability of organizational decisions.

7.4 Develop Employee AI and Data Skills

Successful AI adoption requires employees who can understand and effectively use AI-based systems. Organizations should provide training in AI literacy, data interpretation, and analytical skills so that employees can evaluate AI outputs and use them appropriately in their work.

7.5 Ensure Data Quality and AI Governance

Reliable AI-based decisions depend on accurate and relevant data. Managers should therefore establish effective data-management practices along with appropriate policies for privacy, security, transparency, and responsible AI use. Strong governance can also reduce the risks associated with biased or inappropriate AI-based decisions.

7.6 Adopt AI Through a Gradual Approach

Organizations should consider pilot projects before implementing AI on a large scale. Testing AI applications in specific functions allows managers to identify challenges, evaluate outcomes, and refine processes before wider implementation. This approach can reduce implementation risks and improve organizational readiness.

7.7 Measure Business Outcomes

AI adoption should be evaluated through its contribution to organizational performance rather than simply through the number of AI tools implemented. Managers should assess outcomes such as decision-making speed, operational efficiency, cost reduction, customer satisfaction, innovation, and profitability.

7.8 Encourage Continuous Learning and Adaptation

AI adoption is an ongoing process because technologies, markets, and organizational requirements continuously change. Organizations should regularly update AI systems, train employees, monitor outcomes, and modify processes. Continuous learning can help organizations convert AI capabilities into sustained organizational value.

8. Future Research Directions

The relationship between AI adoption and organizational performance provides several opportunities for future research.

Future studies can examine whether the impact of AI differs across industries such as banking, manufacturing, retail, healthcare, education, and services.

Researchers can also compare organizations at different stages of AI maturity to determine whether deeper integration produces greater organizational benefits.

Another important area is the role of **AI literacy**. Employees may have access to sophisticated AI systems, but their ability to understand and use those systems can determine whether organizational value is actually created.

Longitudinal studies could also examine whether AI benefits increase over time as employees gain experience, processes are redesigned, and organizational capabilities develop.

Finally, future research should examine responsible AI, particularly privacy, algorithmic bias, transparency, explainability, and human accountability.

9. Conclusion

Artificial Intelligence is changing organizations not simply by automating tasks but by transforming how information is processed and how managerial decisions are made. However, the adoption of AI technology does not automatically guarantee improved organizational performance.

The central argument of this chapter is that **AI-driven managerial decision-making can serve as an important mechanism connecting AI adoption with organizational performance**.

AI provides organizations with greater analytical and predictive capabilities. Managers transform these capabilities into decisions, and those decisions influence organizational actions and outcomes.

The relationship can therefore be summarized as:

AI Adoption → AI-Driven Managerial Decision-Making → Organizational Action → Organizational Performance

The future of organizational AI is consequently unlikely to depend only on how much AI a firm possesses. It will increasingly depend on **how effectively organizations combine AI technologies, data, human expertise, managerial judgment, organizational processes, and responsible governance**.

In this sense, the real transformation is not merely from **traditional management to AI-enabled management**, but from **data availability to intelligent decision-making and from intelligent decision-making to measurable organizational value**.

References

- Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120. https://doi.org/10.1177/014920639101700108
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008
- Tornatzky, L. G., & Fleischer, M. (1990). The processes of technological innovation.

Lexington Books.

- Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, *26*(7), 1893–1924.

- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

- Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

- Organisation for Economic Co-operation and Development. (2024). *OECD AI principles.* OECD. https://oecd.ai/en/ai-principles

Chapter 20

AI-Driven Business Management: Strategic Innovation, Digital Transformation, and Competitive Advantage

Simran Thakur, Research Scholar

Abstract

Artificial intelligence (AI) has evolved from a specialized computational technology into a strategic resource with the potential to reshape how organizations create, deliver, and capture value. The emergence of machine learning, predictive analytics, natural language processing, generative AI, and increasingly autonomous AI systems has expanded the role of AI beyond operational automation toward strategic decision-making, innovation, customer engagement, and organizational transformation. However, the availability of AI technologies does not automatically produce competitive advantage. The strategic value of AI depends on an organization's ability to combine technological resources with data, human expertise, organizational routines, leadership, innovation capabilities, and responsible governance. This conceptual chapter synthesizes literature on AI, strategic management, innovation, digital transformation, and competitive advantage to develop an integrated AI Capability–Innovation–Digital Transformation–Competitive Advantage framework. Drawing on the resource-based view, dynamic capabilities theory, knowledge-based theory, the technology–organization–environment perspective, and socio-technical systems theory, the chapter proposes that AI capabilities influence competitive advantage both directly and indirectly through strategic innovation, digital transformation, organizational agility, knowledge creation, and customer value creation. The framework further recognizes leadership, organizational culture, digital maturity, firm size, industry turbulence, and regulatory conditions as contextual factors shaping AI outcomes. The chapter contributes to the emerging literature by shifting attention from AI adoption to the organizational capabilities required to transform AI investments into sustainable business value. Managerial implications concerning AI strategy, human–AI collaboration, capability development, and responsible AI governance are also discussed.

Keywords: Artificial intelligence; AI-driven business management; digital transformation; strategic innovation; competitive advantage; dynamic capabilities; business model innovation; responsible AI

1. Introduction

1.1 Background and Emergence of AI in Business

Artificial intelligence (AI) is now one of the most significant technological developments affecting modern business management. Even though AI isn't a new thing, recent progress in machine learning, deep learning, natural language processing, computer vision, generative AI, and large language models has greatly broadened the range of organizational activities that can be supported or transformed by intelligent technologies. It can now be used to help with forecasting, fraud detection, customer service, marketing personalization, supply-chain management, financial analysis, recruitment, product development, and strategic decision-making. The increasing importance of AI is due to its nature as a potentially general-purpose technology that can be applied in different industries and organizational functions (Cockburn et al., 2018; Goldfarb et al., 2023).

The strategic importance of AI therefore goes beyond that of technological automation. While traditional automation is mainly concerned with carrying out pre-determined tasks more efficiently, modern AI is able to detect patterns, make predictions, produce content, assist in decision-making, and interact with users. Generative AI even extends these capabilities by creating text, images, software code, summaries, recommendations, and other kinds of knowledge work. As a result, AI is increasingly being incorporated into managerial processes rather than staying limited to the technical departments.

The organizational implications are significant. AI has the potential to speed up and expand the scale of information processing, to improve forecasting, to reduce repetitive tasks, to assist with employee learning, and to enable more personalized customer experiences. For instance, research carried out in the workplace indicates that generative AI assistance can lead to increased productivity, the extent of the benefit, however, differing depending on the level of employee experience and the nature of the task (Brynjolfsson et al., 2023). These findings support the view that AI should not be seen merely as an information technology but rather as an organizational capability whose value is determined by the way it is integrated with people and processes.

1.2 Progressing from Digitalization to Business Management Driven by AI

The rise of AI should be seen as part of a wider trend moving from digitization to digitalization and finally to digital transformation. Digitization means converting analogue information into digital form. Digitalization consists of using digital technologies to improve current processes. Digital transformation denotes a more extensive change at an organizational level during which digital technologies reshape processes, structures, customer experiences, capabilities, and business models (Verhoef et al., 2021).

The application of AI to business management is just another stage in this process. It takes place when AI is systematically incorporated into strategic planning, organizational decision-making,

operations, innovation, and the creation of customer value. Hence, AI-driven management does not mean simply buying AI software; it means integrating AI into the everyday routines of an organization and making use of its outputs to enhance strategic and operational decisions.

It is important to make this distinction since companies can introduce AI without bringing about real transformation. For example, a company might use a chatbot to handle customer service or generative AI to prepare documents without making any changes to its core processes. On the other hand, what is meant by AI-driven transformation is that organizations redesign their workflows, decision-making structures, employee roles, and business models in light of AI-enabled capabilities. Research into AI implementation stresses that in order to obtain ongoing value from AI, organizations had better go beyond isolated experiments and move towards production, scaling, and integration within the organization (Haefner et al., 2024).

1.3 Strategic Importance of AI

AI can contribute to organizational value through several mechanisms. First, it can enhance operational efficiency by automating repetitive activities and optimizing resource allocation. Second, it can strengthen decision quality by processing large and complex datasets more rapidly than conventional approaches. Third, AI can facilitate customer value creation through personalization, recommendation systems, conversational interfaces, and predictive services. Fourth, AI can stimulate innovation by supporting knowledge discovery, experimentation, product development, and business-model redesign (Bahoo et al., 2023).

The innovation implications are particularly important. AI can identify patterns that may not be readily observable to human decision-makers, generate alternative solutions, accelerate experimentation, and support the analysis of customer needs. Consequently, AI can influence both incremental and more radical forms of innovation. Bahoo et al. (2023) describe AI as capable of reshaping corporate innovation processes by improving the management and interpretation of information throughout innovation activities.

AI can also improve organizational agility. Organizations operating in turbulent markets require the capacity to detect environmental changes, interpret information, and respond quickly. AI-supported analytics can strengthen these processes by providing real-time or near-real-time insights. Nevertheless, technological capability alone does not guarantee agility. Organizations must possess appropriate decision rights, leadership, skills, and routines to act upon AI-generated insights.

1.4 AI and Competitive Advantage

Competitive advantage refers broadly to an organization's ability to create greater value than competitors or to perform activities in ways that competitors cannot easily replicate. From a

resource-based perspective, competitive advantage depends on resources and capabilities that are valuable and difficult to imitate or substitute (Barney, 1991).

AI creates an interesting strategic challenge because many AI technologies are increasingly accessible through cloud platforms and commercial applications. If competitors can purchase the same technology, the technology itself may not constitute a sustainable source of advantage. Instead, differentiation is likely to emerge from complementary organizational resources such as proprietary data, specialized knowledge, employee expertise, organizational routines, customer relationships, and the capacity to integrate AI into distinctive business processes.

1.5 Problem Statement and Research Gap

Although the literature on AI is growing, it is still scattered across technology management, information systems, innovation, strategic management and organizational studies. Some focus on technological adoption, some on productivity, some on innovation, digital transformation, or even on the ethical implications of technology. This disaggregation makes it challenging to understand how AI can be a source of sustainable competitive advantage through an organizational pathway.

The other restriction is the focus on implementation as a technology issue. Even if organizations have sophisticated algorithms, they don't create value due to data quality issues, lack of employee competence, lack of leadership, organizational resistance, cyber security issues, and inappropriate governance. It is, therefore, a technological, organisational, strategic and social phenomenon to implement AI (Haefner et al., 2024).

A third gap is the connection between AI and other aspects of organizational change. AI can enhance specific tasks rather than revolutionizing the organization.AI can optimize specific tasks and not necessarily overhaul the organization itself. The pathways by which the capabilities of AI support strategic innovation, digital transformation, organizational agility, and potential competitive advantage are therefore more important to focus on.

1.6 Purpose and Research Questions

This chapter aims to provide an integrated conceptual framework that offers insights into the potential of AI business management techniques to create value for organizations through strategic innovation and digital transformation.

The chapter addresses five questions:

1. How do AI capabilities influence strategic innovation?
2. How do AI capabilities facilitate digital transformation?
3. Through what mechanisms does AI contribute to competitive advantage?

4. Which organizational and environmental conditions strengthen or weaken these relationships?
5. How can organizations develop responsible and sustainable AI capabilities?

2. Literature Review

2.1 Evolution of Artificial Intelligence in Business

The development of AI in business can broadly be viewed through several stages. Early expert systems relied on explicit rules and structured knowledge. Subsequent developments in machine learning enabled systems to learn patterns from data. Deep learning expanded capabilities in areas such as image and speech recognition, while generative AI introduced systems capable of producing new content and interacting through natural language.

This technological evolution has changed the managerial role of AI. Earlier applications were primarily associated with structured operational problems. Contemporary systems increasingly support knowledge-intensive activities such as writing, coding, analysis, customer interaction, forecasting, and strategic research. AI is therefore moving from an automation-oriented role toward an augmentation and decision-support role (Raisch & Krakowski, 2021).

This does not mean that automation has become irrelevant. Rather, automation and augmentation often coexist. Some activities can be fully automated, while others benefit from human–AI collaboration. Raisch and Krakowski (2021) argue that organizations need to manage the tension between automation and augmentation rather than assuming that one approach will dominate all organizational tasks.

2.2 AI-Driven Business Management

AI-powered business management can be explained as the thorough incorporation of AI technologies, data sources, individual skills, and business processes into management operations to enhance decision-making, productivity, innovation, and value generation.

AI can be applied throughout key business operations. It's useful in marketing for recommendation, personalization and demand prediction, as well as segmentation. AI can support financial institutions in fraud detection and prevention, credit scoring, financial forecasting, and risk management. Intelligent automation and predictive maintenance can increase productivity in operations. AI can aid in talent analytics and workforce planning in HRM. Predictive Analytics can boost demand forecasting and inventory management in supply chains. Generative AI can support research and development by helping with ideas and knowledge acquisition.

The strategic value of these applications depends on integration. Isolated applications may produce localized efficiency gains, but organization-wide value requires interoperability, data governance, employee capability, and managerial coordination. AI implementation research emphasizes that scaling AI involves moving from proof of concept to production and organizational platformization (Haefner et al., 2024).

2.3 AI and Strategic Innovation

Strategic innovation refers to the development or implementation of new products, services, processes, technologies, organizational practices, or business models that create value and strengthen competitive positioning. AI can influence innovation by expanding the organization's ability to process information, identify opportunities, generate alternatives, and experiment rapidly.

AI can support product innovation by identifying customer preferences and market patterns. It can facilitate process innovation through automation and optimization. It can enable service innovation through intelligent customer interfaces and personalized experiences. It can also support business model innovation by enabling new forms of digital service delivery, subscription models, predictive services, and platform-based value creation.

The relationship between AI and innovation is not merely technological. Innovation requires organizational learning, experimentation, cross-functional collaboration, and willingness to challenge existing assumptions. Therefore, AI is better understood as an innovation-enabling capability than as an innovation outcome in itself (Bahoo et al., 2023).

2.4 AI and Digital Transformation

Digital transformation is the process of fundamental changes in organizational processes, structures, capacity, customer experiences and business models that are facilitated by digital technologies (Verhoef et al., 2021). AI can help speed up this transition since it can turn vast amounts of digital information into predictions, recommendations, classifications, created content, or automated responses.

There are at least four ways that AI is helping drive digital transformation. First, it facilitates process transformation through automation or optimization of activities. Second, it enables the transformation of the customer with personalization and intelligent interaction. Third, it supports changes in organizational structures, role changes, decision-making processes, and knowledge flows, or organizational transformation. Fourth, it can help in the transformation of the business model by facilitating new products and services with digital capabilities.

Digital transformation isn't simply about installing digital tools. It is about an organizational change and strategic renewal. In the context of digital transformation, the ability to sense technological and market changes, seize opportunities and reconfigure organisational resources is vital as highlighted by Warner and Wäger (2019).

2.5 AI and Competitive Advantage

AI can contribute to competitive advantage through four major pathways.

Cost advantage

AI-supported automation can reduce processing time, minimize errors, improve resource allocation, and increase employee productivity. However, cost savings alone may not create sustainable advantage because competitors can potentially acquire similar technologies.

Differentiation

AI can enable personalized products and services, responsive customer support, improved product quality, and more sophisticated customer experiences. Differentiation is more sustainable when it is based on proprietary data, unique organizational knowledge, or deeply embedded processes.

Agility

AI can strengthen organizational responsiveness by improving environmental scanning, forecasting, and decision-making. Agility is especially important in turbulent markets where customer preferences, technologies, and competitive conditions change rapidly.

Resilience

AI can support predictive risk management, supply-chain monitoring, anomaly detection, and scenario analysis. These capabilities can help organizations anticipate disruptions and respond more rapidly

Empirical and conceptual literature therefore suggests that AI can contribute to business performance, but the relationship is conditional rather than automatic. AI becomes strategically valuable when technological resources are combined with complementary organizational capabilities (Mikalef & Gupta, 2021; Åström et al., 2022).

2.6 Theoretical Foundations

Resource-Based View

The resource-based view (RBV) proposes that firms achieve competitive advantage by controlling valuable and difficult-to-imitate resources and capabilities (Barney, 1991). Applied to

AI, the theory suggests that access to an algorithm is unlikely to be sufficient because algorithms can increasingly become standardized or commercially available. More distinctive resources may include proprietary datasets, specialized AI expertise, organizational routines, customer knowledge, and effective AI governance.

Knowledge-Based View

The knowledge-based view conceptualizes the firm as an institution for integrating specialized knowledge (Grant, 1996). AI can contribute to knowledge integration by processing large information sets, identifying patterns, generating summaries, and making knowledge more accessible across organizational boundaries.

Generative AI may be particularly significant because it can support knowledge-intensive tasks. Nevertheless, AI-generated outputs require human evaluation because systems can generate inaccurate, biased, or contextually inappropriate information. Human expertise remains central to validating and applying organizational knowledge.

Technology–Organization–Environment Perspective

The technology–organization–environment (TOE) framework emphasizes that technology adoption is influenced by technological characteristics, organizational readiness, and external environmental conditions (Tornatzky & Fleischer, 1990). Applied to AI, technological factors include data availability, system compatibility, and infrastructure. Organizational factors include leadership, skills, financial resources, culture, and digital maturity. Environmental factors include competitive pressure, regulation, industry characteristics, and stakeholder expectations.

Socio-Technical Systems Perspective

AI implementation changes both technological and social systems. Employees may experience changes in responsibilities, decision authority, work processes, skills, and performance expectations. Therefore, effective AI implementation requires alignment among technology, people, tasks, processes, and organizational structures.

This perspective is particularly important for generative AI because the greatest benefits may emerge through human–AI collaboration rather than complete human replacement. Recent reviews similarly emphasize that successful AI integration requires attention to both technical and social elements (Haefner et al., 2024).

3. Conceptual Review Methodology

This chapter adopts a conceptual integrative literature review rather than an empirical research design. The objective is not to estimate causal effects statistically but to synthesize established

theoretical perspectives and recent research in order to develop an integrated conceptual framework.

The literature was identified through targeted searches of scholarly publications relating to artificial intelligence, AI capabilities, digital transformation, strategic innovation, competitive advantage, dynamic capabilities, generative AI, and responsible AI. Priority was given to peer-reviewed journal articles, seminal theoretical publications, systematic reviews, and authoritative institutional frameworks. Recent literature was emphasized because of the rapid development of generative AI, while foundational publications were retained where they provide essential theoretical foundations.

The review was organized around five concept areas: AI capability, strategic innovation, digital transformation, competitive advantage, and organizational context. Literature was then compared conceptually to identify recurring relationships, mechanisms, and boundary conditions.

Rather than treating all AI adoption studies as equivalent, the analysis distinguishes between technological adoption and organizational capability. It also differentiates direct effects of AI from indirect effects occurring through innovation, transformation, agility, knowledge creation, and customer value creation.

The conceptual synthesis produced five major relationships:

1. AI capabilities contribute to strategic innovation.
2. AI capabilities contribute to digital transformation.
3. Strategic innovation contributes to competitive advantage.
4. Digital transformation contributes to competitive advantage.
5. AI capabilities may influence competitive advantage directly, but this relationship depends on complementary organizational capabilities.

4. Conceptual Results and Framework Development

4.1 AI Organizational Capability

The central construct of the framework is AI organizational capability. It refers to the organization's ability to acquire, integrate, deploy, govern, and leverage AI-related resources to achieve strategic objectives. Consistent with the RBV, AI capability is broader than technological infrastructure because it includes complementary resources and routines (Barney, 1991; Mikalef & Gupta, 2021).

4.2 AI Capabilities and Strategic Innovation

AI can strengthen innovation by increasing access to information, accelerating experimentation, supporting customer insight, and reducing the time required to evaluate alternatives. AI-enabled innovation may occur at product, process, service, and business-model levels.

The relationship can be explained through the organization's improved ability to identify market opportunities, generate alternatives, test solutions, and learn from outcomes. AI can therefore strengthen both exploratory and exploitative innovation.

4.3 AI Capabilities and Digital Transformation

AI can deepen digital transformation by moving organizations from digitally supported processes toward intelligent and increasingly adaptive processes. For example, conventional digital systems may record transactions, whereas AI systems can analyze those transactions, predict future conditions, and recommend actions.

AI can consequently influence process design, customer experience, organizational decision-making, and business models. The transformation effect becomes stronger when AI is integrated across functions rather than implemented as isolated applications.

4.4 Strategic Innovation and Competitive Advantage

Strategic innovation contributes to competitive advantage by creating new sources of customer value, improving products and services, reducing process costs, and enabling new revenue models. Innovation is particularly important when competitors can rapidly imitate technologies. Organizations that continuously innovate can renew their value propositions and reduce dependence on existing sources of advantage.

4.5 Digital Transformation and Competitive Advantage

Digital transformation can strengthen competitive advantage by improving operational efficiency, customer responsiveness, organizational agility, and market reach. However, transformation is not automatically beneficial. Poorly coordinated digital initiatives may increase complexity and costs. The strategic value of digital transformation therefore depends on organizational alignment, leadership, employee capabilities, and the ability to redesign processes around digital technologies.

4.6 AI Capabilities and Competitive Advantage

AI capabilities may also have a direct relationship with competitive advantage through improved forecasting, personalization, automation, knowledge processing, and decision support. However, this direct relationship is expected to be weaker when competitors have similar access to AI technologies.

5. Discussion

5.1 Theoretical Contributions

The proposed framework contributes to strategic management and technology research in four ways.

First, it extends the resource-based view by conceptualizing AI as a bundle of technological, human, informational, and organizational resources. AI technologies alone may not satisfy the conditions necessary for sustainable competitive advantage because many technologies are widely available. The strategic resource lies in the organization's distinctive ability to orchestrate these resources.

Second, the framework extends dynamic capabilities theory by positioning AI as an enabler of sensing, seizing, and reconfiguring. AI can improve sensing through predictive analytics, seizing through decision support, and reconfiguration through process and business-model transformation (Teece, 2007).

Third, the framework integrates the knowledge-based view by recognizing AI as a mechanism for processing and distributing knowledge. However, organizational knowledge remains dependent on human interpretation and contextual understanding (Grant, 1996).

Fourth, the framework adopts a socio-technical perspective by emphasizing that AI transformation requires simultaneous changes in technology, people, tasks, and organizational processes. This prevents an overly deterministic view in which technological adoption automatically produces performance improvements.

5.2 Managerial Implications

Managers should treat AI as a strategic transformation programme rather than a collection of disconnected technology projects. AI initiatives should begin with clearly defined organizational problems and strategic objectives. The central question should not be "Where can AI be used?" but rather "Which organizational problems can AI help us solve better?"

Second, organizations should invest in data capability. Poor-quality, fragmented, inaccessible, or biased data can undermine AI performance. Data governance should therefore be considered a strategic capability rather than merely an IT responsibility.

Third, managers should develop AI literacy across the workforce. Employees do not necessarily need advanced programming skills, but they increasingly need to understand AI's capabilities, limitations, appropriate applications, and risks. Continuous upskilling and reskilling are therefore central to AI-driven transformation.

Fourth, organizations should redesign workflows rather than simply automate individual tasks. AI creates greater value when processes are reconsidered as integrated systems. This requires collaboration between technical teams, managers, and employees.

Fifth, managers should measure AI outcomes through business-oriented indicators such as productivity, customer satisfaction, innovation speed, decision quality, revenue growth, risk reduction, and employee learning rather than focusing exclusively on the number of AI applications deployed.

5.3 Human Capital and Human–AI Collaboration

AI is changing the composition of work rather than simply eliminating work. The automation of routine tasks can allow employees to focus on activities requiring judgment, creativity, empathy, relationship management, and strategic reasoning. However, this requires deliberate job redesign and capability development.

The distinction between automation and augmentation is therefore central. Organizations should identify which activities are suitable for automation and which require human judgment. In many knowledge-intensive tasks, the strongest configuration may be human–AI collaboration rather than full substitution (Raisch & Krakowski, 2021).

Generative AI provides an important example. Evidence suggests that AI assistance can improve productivity and help less experienced employees learn from more experienced practices (Brynjolfsson et al., 2023). The managerial implication is that AI may function as a mechanism for organizational learning if employees are trained to critically evaluate and apply AI-generated outputs.

6. Future Research Agenda

The proposed framework provides several directions for future empirical research.

First, researchers can test whether strategic innovation and digital transformation mediate the relationship between AI capability and competitive advantage.

Second, longitudinal research is needed to determine whether AI-generated performance improvements persist over time or diminish as technologies become standardized.

Third, future studies should examine differences between large organizations and SMEs because resource constraints, skills, infrastructure, and leadership structures can substantially influence AI adoption.

Fourth, emerging economies deserve greater attention. Institutional conditions, digital infrastructure, access to skilled labour, regulation, and financing may produce AI outcomes different from those observed in developed economies.

Fifth, future research should distinguish between conventional AI, generative AI, and autonomous or agentic AI. These technologies may create different organizational opportunities and risks.

Finally, responsible AI deserves greater integration with strategic management research. Future studies should examine whether trust, transparency, fairness, and governance can themselves become sources of organizational legitimacy and competitive differentiation.

8. Conclusion

Artificial intelligence should not be conceptualized merely as a technological investment or automation mechanism. Its strategic value emerges when organizations integrate AI with data, human expertise, organizational learning, innovation capabilities, and digital transformation. The central proposition of this chapter is therefore that AI capability creates competitive advantage through organizational orchestration rather than technological ownership alone.

The proposed AI Capability–Innovation–Digital Transformation–Competitive Advantage framework explains this process by positioning AI organizational capability as the foundation, strategic innovation and digital transformation as major value-creation mechanisms, and organizational agility, knowledge integration, and customer value as important intermediate pathways. Competitive advantage is consequently understood as an outcome of the organization's ability to convert AI resources into distinctive capabilities and continuously renew those capabilities.

The framework also highlights the importance of context. Leadership, organizational culture, digital maturity, firm size, industry turbulence, competitive intensity, and regulatory conditions influence the extent to which AI produces value. Equally, responsible governance is essential for ensuring that AI-driven growth does not compromise privacy, fairness, security, transparency, or accountability.

Ultimately, organizations are unlikely to gain sustainable advantage simply by accessing the newest AI model. The more enduring source of advantage lies in developing distinctive combinations of technology, data, people, knowledge, processes, innovation, and governance. AI-driven business management should therefore be viewed as a continuing process of strategic renewal through which organizations learn, innovate, transform, and create superior value in increasingly digital markets.

References

- Bahoo, S., Cucculelli, M., & Qamar, D. (2023). Artificial intelligence and corporate innovation: A review and research agenda. *Technological Forecasting and Social Change, 188*, 122264. https://doi.org/10.1016/j.techfore.2022.122264
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108
- Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly, 24*(1), 169–196. https://doi.org/10.2307/3250983
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161
- Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 115–146). University of Chicago Press.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–160. https://doi.org/10.2307/2095101
- Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review, 97*(4), 62–73.
- Goldfarb, A., Taska, B., & Teodoridis, F. (2023). Could machine learning be a general-purpose technology? A comparison of emerging technologies using data from online job postings. *Research Policy, 52*(2), 104662. https://doi.org/10.1016/j.respol.2022.104662
- Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17*(S2), 109–122. https://doi.org/10.1002/smj.4250171110
- Haefner, N., Parida, V., Gassmann, O., & Wincent, J. (2024). Implementing and scaling artificial intelligence: A review, framework, and research agenda. *Technological Forecasting and Social Change, 197*, 122878. https://doi.org/10.1016/j.techfore.2023.122878

- Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Review Press.
- Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. *Business & Information Systems Engineering, 63*, 5–20. https://doi.org/10.1007/s12599-020-00676-7
- Keding, C. (2021). Understanding the interplay of artificial intelligence and strategic management: Four decades of research in review. *Management Review Quarterly, 71*, 263–300. https://doi.org/10.1007/s11301-020-00181-x
- Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management, 58*(3), 103434. https://doi.org/10.1016/j.im.2021.103434
- Mikalef, P., Conboy, K., & Krogstie, J. (2022). Artificial intelligence as an enabler of B2B marketing: A dynamic capabilities micro-foundations approach. *Industrial Marketing Management, 98*, 80–92. https://doi.org/10.1016/j.indmarman.2021.08.003
- Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S. O., Torvatn, H. Y., Gupta, M., & Niehaves, B. (2022). Enabling AI capabilities in government agencies: A study of determinants for European municipalities. *Government Information Quarterly, 39*(4), 101596. https://doi.org/10.1016/j.giq.2021.101596
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072
- Sjödin, D., Parida, V., Palmié, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research, 134*, 574–587.
- Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1
- Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350. https://doi.org/10.1002/smj.640

- Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003
- Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research, 122*, 889–901. https://doi.org/10.1016/j.jbusres.2019.09.022
- Warner, K. S. R., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning, 52*(3), 326–349. https://doi.org/10.1016/j.lrp.2018.12.001
- Wu, R. M. X., Vakilian, J., Shafiabady, N., Hadjinicolaou, N., Ud Din, F., & Bhandari, B. (2023). Using artificial intelligence (AI) to predict organizational agility. *PLOS ONE, 18*(5), e0283066. https://doi.org/10.1371/journal.pone.0283066

Chapter 21

Computational Transparency in the AI Era: Ethical Divides in Digital Research Oversight

Sourav Sen,
Graduate Instructor,
Basic Training Centre, Kaliganj,
SCERT, Assam
Monali Debnath,
Research Scholar,
Assam University, Silchar, Assam

Abstract

Artificial intelligence has developed into a driving force in the investigative procedure, guiding research design, influencing how data is handled, outcomes are circulated; though the moral value framework accountable for supervising this transition has fallen behind the curve through the learning architecture itself. This study intends to investigate algorithmic transparency characterised by the unveiling of how automated systems sense and respond to data, as a fundamental but inconsistently enforced digital research governance protocol. The study integrates the thematic-synthetic approach into the theoretical one, thus, covering the latest trends in public opinion about the AI system, international models of managing artificial intelligence, and related empirical data, rather than focusing on the methodological aspect. The analysis notes three interconnected issues affecting accountability mechanisms, as a rift within the current system of the reviewing panels presenting a technical incompetence, territorial division concerning the comprehensive regulations implying the need for adaptation in multinational organizations, and social governance relying on artificial intelligence. These problems aggravate in combination with each other and not on individual grounds, presenting a gap for technical solutions to be efficient. The study suggests that these issues should be addressed through the organized and unitary multinational framework rather than the application of separate measures, thus providing the researchers; ethics review panels, and legislators with practical directions for enhanced accountability toward emerging artificial intelligence technologies.

Keywords: Artificial intelligence, accountability mechanisms, reviewing panels, technical incompetence, territorial division, multinational organizations, social governance.

1. Introduction

Artificial intelligence has evolved into an integral part of contemporary research instead Of just a supplementary tool. It assists investigators to formulate research inquiry, evaluate data, survey literature, compose text and authoring documents (Khalifa & Albadawy, 2024). AI also sways how studies get distributed, sourced, and investigated. Recommender engines boost the permanence of some studies, while ranking signal systems influence investigation to engage communities (Färber et al., 2023). Nonetheless AI has progressed more swiftly than governance frameworks to supervise Research Institutional Review Boards (IRBs) and Research Ethics Communities (RECs) were primarily designed for investigation-driven research where policies guiding actions are clear-cut and transparent (Bouhouita-Guermech et al., 2023).

Computational transparency has accordingly turned into a model quandary. It denotes that the argumentation, data providers, suppositions, and problem-solving process of an AI model should be comprehensible and amenable to scrutiny (Díaz-Rodríguez et al., 2023). In conventional research, protocols can typically originate gradually. Several AI models, meanwhile, function as "algorithmic opacity" making it complicated to formulate a sound judgement. Occasionally even its creators are unable to completely describe the outcomes (Gryz & Rojszczak, 2021). This deficit of openness sparks worry about knowledgeable consent, data protection, risk analysis, peer assessment, composition and answerability post- publication (Resnik & Hosseini, 2025).

This study contains that issues stem from a web of systemic learning deficits as opposed to a solitary opacity (Martinho et al., 2026). These encompass the disparity between ethics panel, algorithmic depth, regulatory divergence and the divide between public faith and the swift uptake of AI (Roberts et al., 2023). Such divides impact investigators since their data may be transformed and computed by platforms beyond the reach of oversight. The paper undertakes a research study on ethical AI, responsibility, and administration, summarizing the main ideas in comparative tables and diagrams. The article suggests modifying the existing approaches in order to enhance computational interpretability in research administration. India and other developing countries are the priority areas in this regard, given the need to integrate AI into the administrative functions while ensuring the highest level of research quality.

2. Literature Review

Studies propose that algorithmic accountability should be exercised throughout the research process. This can be achieved by making sure that the use of artificial intelligence is disclosed, the data or database is provided or designed, consent is obtained, the process by which AI arrives at specific conclusions is explained, appropriate documentation is maintained, and responsibility is assigned in case of failures or errors (Díaz-Rodríguez et al., 2023). Findings on interpretability

suggest that transparency and explainability are not equivalent concepts. Disclosure of design principles does not guarantee that the decisions reached by an AI system will be accessible to a given user. Therefore, explanations need to be provided in a manner that suits the expectations and level of understanding of different groups of stakeholders (Laato et al., 2022). Studies on the ethical issues of AI reveal various factors to consider in order to establish ethical principles of AI. These include privacy, bias, informed consent, data makeup, data governance, and transparency. The reports also stress the need to document the use of automated procedures, disclose the use of synthetic data, and engage communities in research (Resnik & Hosseini, 2025).The issues of AI transparency seem especially relevant to IRBs and Research Ethics Committees as expertise on the matter is limited in oversight bodies. At the same time, evolution boards do not have appropriate methodologies to challenge AI-related research (Bouhouita-Guermech et al., 2023). The recommendations stress the importance of ensuring AI observability at all stages of research as well as responsibility of investigators, institutions and authorities (Batool et al., 2025). The findings of the 2026 study highlight the need to consider transparency in the context of fairness, interpretability, answerability and human surveillance at practical, organizational and regulatory levels; recommend auditing of AI before deployment and after implementation (Martinho et al., 2026).

Transparency can be achieved through audits by enforcing answerability. Audits can be used to assess fairness, privacy, responsibility, clarity, reliability, human oversight, and stakeholder engagement (Laine et al., 2024). Indeed, a comparative legal study revealed that significantly different approaches are being taken in across the Atlantic to govern artificial intelligence. However, the mere presence of policies is not sufficient to ensure legitimacy and trust in the system. At the same time, reliance on AI depends on whether the public deems it objective (Laux et al., 2024). In general, the research under analysis suggests the importance of principles-based disclosure, as well as on going, multidisciplinary, and tailored stewardship, as well as the significance of overcoming organizational, global, and conceptual deficiencies in terms of understanding the interconnections between transparency, dependency, and accountability, from which the current investigation departed (Brey & Dainow, 2024).

3. Objectives

1. To recognise and thematically integrate the fundamental moral aspects of algorithmic transparency and use for AI-driven exploration and computing environment.
2. To differentiate how major legal frameworks (European Union, India, United States and ASEAN) and ethics committees put algorithmic accountability to practice and supervision.

3. To analyse the empirical evidence between public confidence, AI implementation and the efficacy of current accountability frameworks.

Table 1: Fundamental elements of algorithmic transparency in AI assisted inquiry

Elements	Illustration
Data linkage and informed consent	Transparency indicates interpreting the data used to instruct AI and soliciting consent from the outset.
Computational interpretability	Explicatory skill to describe AI options in lucid terms that impact mankind and which controllers can comprehend.
Organizational provenance	Documentation of evaluation procedures, alterations, and reliable groups to assure good governance.
Oversight mechanism	Using audit records and defined roles to assure shared data results in answerability and effective implementation.

Table 1 epitomises the core concepts of algorithmic transparency revealed in the assessed literature. These incorporate data feeds and authorization, algorithmic accountability, organisational transparency and liability. The findings suggest that transparency should commence prior to the design phase, pre-data collection instead of post-deployment of the AI model (Felzmann et al., 2020). They also illustrate that transparency signifies notifying people about how AI is being applied while explicability indicates lucidly describing how AI decides. In research, audit trails and different accountabilities are important to ensure that any form of mismanagement is easily identifiable and can be addressed. All these factors contribute to the promotion of responsible AI (Papagiannidis et al., 2025).

Table 2: Comparative analysis of Major AI and data management model

Structure	Dominion	Governance strategy	Main attributes
EU AI Act	European Union	Human-centered, horizontal, retrospective monitoring.	High stakes AI must divulge training sets, and specifications.
Algorithmic Accountability Act	United States	Spontaneous, forward-thinking, evidence-based.	High stakes AI mandate impact evaluation prior to being used in vulnerable areas.
Digital Personal Data Protection Act 2023	India	Consent governance, surrogate AI regulation	Data privacy mandates are in effect while research exceptions are still evolving.
ASEAN framework for AI governance and ethics	ASEAN region	Collaborative, principles-driven and voluntary	Common values foster discretionary alignment through collaboration and mutual agreement.

Table 2 above highlights the four significant AI operating models that make it clear that there is no general method of regulating artificial intelligence on a global scale. The EU pursues the principle of safeguarding mankind from all risks through the strict risk-differentiated meteorology (Lu & Tie, 2025). The US focuses on ensuring that the impact is evaluated and innovation is facilitated, while India does not intend to create a set of laws for regulating AI, but instead, use the existing laws relating to data privacy and consent for processing (Cheong, 2024). The ASEAN Framework lists the ethical principles; however, it is not mandatory. This heterogeneity impacts the research activities of multinational entities since they have to take into account all the possible jurisdictions that might apply different regulations to the same activity. At the same time, it becomes evident that all the four jurisdictions determined the priorities to address, as well as the conditions of governance and scale (Almada & Radu, 2024).

Table 3: Regulatory hurdles in AI research synthesis

Challenge	Explanation
Technical skills gap	IRBs and RECs are frequently manned by health-related, lawful or social scientist as opposed to data engineers restricting their potential to assess AI models.
Competing interests	Independent Ethics Committee (IEC) may encounter stress to embrace investigation that aids their staffs, prejudiced assessment.
Ethical offloading of AI	Leveraging AI tools to plan and inspect governance vulnerabilities subverting human deliberation that ethics-driven decisions need.
Bureaucratic delay	Assessment timeframe for steady-state analyses refuse adaptive AI models, leaving authorizations based on bygone shots.

Table 3 indicates the pivotal obstacles encountered by research centres in supervising AI. Many ethics panel are amateurs in AI and data engineering. IEC may navigate a clash of interests when validating research (Bernstein et al., 2021). Research also cautions that AI must not substitute human discernment in moral choices. Another obstacle is that AI models vary rapidly while active review processes are modelled for conventional investigation but lag behind (Hassen, 2025).

Table 4: Stakeholders ethical misalignment in digital research governance

Misalignment	Explanation
Capability gap	Regulatory bodies fall short of technical proficiency to carefully gauge proposed AI models.
Demarcation	EU, US, India, and ASEAN Rules vary obstructing compliance for international research.
Resource inequality	Well -endowed organisations have a higher limit for technical governance than modest organisations.
Credibility gap	Open AI deployment is enhancing rapidly than faith in its dependability and oversight.

Table 4 provides a list of the four major ethical cleavages uncovered among the contributors. The cleavages are characterized by the lack of expertise in such areas as technology, guidelines, assets, and public trust. It is noteworthy that the issues are interrelated and may aggravate one another. For instance, the organizations which lack resources and knowledge cannot meet the required standards and regulations (Lahusen et al., 2024). Additionally, the study confirms public

interest in data privacy, ownership, and surveillance. The issues are especially pertinent for those whose data is used for the purposes of advanced data analytics such as AI developments (Knowles & Richards, 2021).

Table 5: Suggested methods for bolstering algorithmic accountability throughout the compliance framework

Supervision stage	Suggested methods
Preliminary design	Prerequisite of AI effective evaluation and reporting prior to data acquisition.
Organisational review	On boarding technical trainers and interdisciplinary specialists to review panel.
Model construction	Incorporation of interpretability criteria during technical framing rather than supplementing them later.
Issuing and revelation	Record of reassessment, modifications, and AI-augmented research papers.
Post-implementation	Deployment of persistent external auditing rather than one-off clearance.

Table 5 outlines the key strategies for enhancing interpretability in the course of the investigation. This incorporate impact analysis prior to data acquisition, AI ethicist, AI transparency and post-implementation reviews. The research indicates that openness should be upheld during every stage of the research, not just at the authorization phase (Madanchian & Taherdoost, 2025). In general, operational visibility needs foresightedness, professional training centres, explicit rules, stakeholder engagement, and constant surveillance. A range of actions must take place simultaneously as no solitary solution is sufficient to assure responsible AI (Rismani et al., 2025).

4. Results and discussion

The outcomes of the research are represented by applying four figures that synthesise the core themes recognised in the research synthesis. These figures illustrate that algorithmic transparency is unattainable via a single fix. Rather it banks on institutional strengthening, bolstering rules across several areas, and developing public confidence (Nieuwenhuizen et al., 2025).

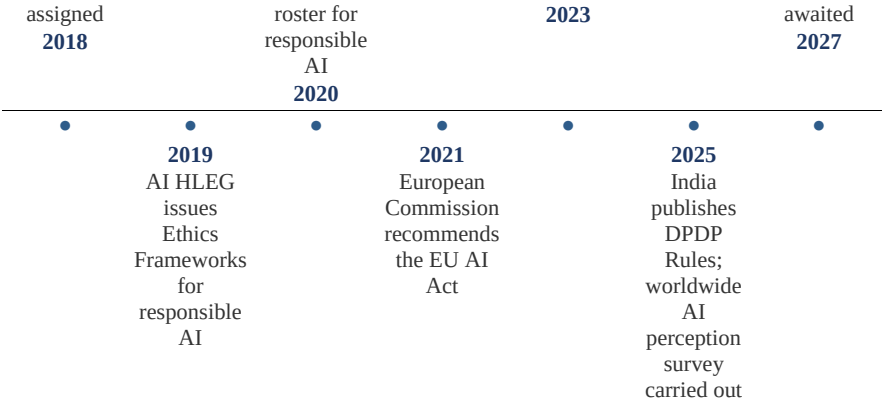


Figure 1: Timeline of AI milestones and data maturity levels relative to Algorithmic accountability (2018 – 2027).

Figure 1 displays the timeline of the legal frameworks explained in table 2. It indicates that algorithmic transparency has evolved in different ways internationally instead of a unified approach (Pua, 2026). Europe's strategy for managing AI is systematic. In contrast, India's approach to data protection and personal data is gradual, even though India has taken a number of measures to ensure data security, for example, the Data Protection and Digital Privacy Act and other measures, however, there are no laws regulating artificial intelligence. Additionally, the graphic includes the 2025 world trust survey, which shows that trust in AI is becoming institutionalized and embedded in the mechanisms and frameworks that govern AI (Laux et al., 2024). Overall, the timeline suggests that governance and trust in AI develop in tandem, although at different rates for different countries.

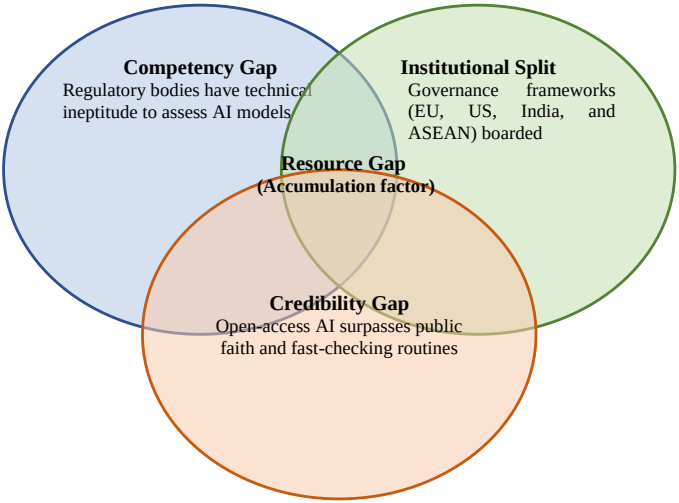


Figure 2: Core framework of ethical gaps in digital research governance.

Figure 2 depicts that the expertise, territorial, asset, and credibility gaps are linked. The resource disparity is centrally located because it intensifies subsequent challenges more significantly. Organisations with technical in aptitude and assets often find it difficult to handle several needs. Trust deficit in public can also limit mankind's eagerness to contribute in responsible AI implementation (Robles & Mallinson, 2025). The figure indicates that fine tuning only one aspect is not sufficient. AI trustworthiness needs effective oversight, high administrative competence, technical proficiency, ample resources, and more public faith working hand-in-hand (Díaz-Rodríguez et al., 2023).



Figure 3: Thematic progression of algorithmic transparency across the research data governance framework.

Figure 3 presents the five phases of algorithmic transparency in online research: data processing, AI computing, research ethics review, revelation and public responsiveness. The feedback loop indicates that supervision should be maintained even after ethical clearance. Rather than being a one-shot procedure, ethical approval should be mandatory by constant surveillance and auditing (Martinho et al., 2026). The figure also indicates the major functions of ethics panel although

they often are devoid of ample assets (Bouhouita-Guermech et al., 2023). On the whole, it illustrates that responsible AI needs both scrutiny, pre-deployment and sustained follow up.

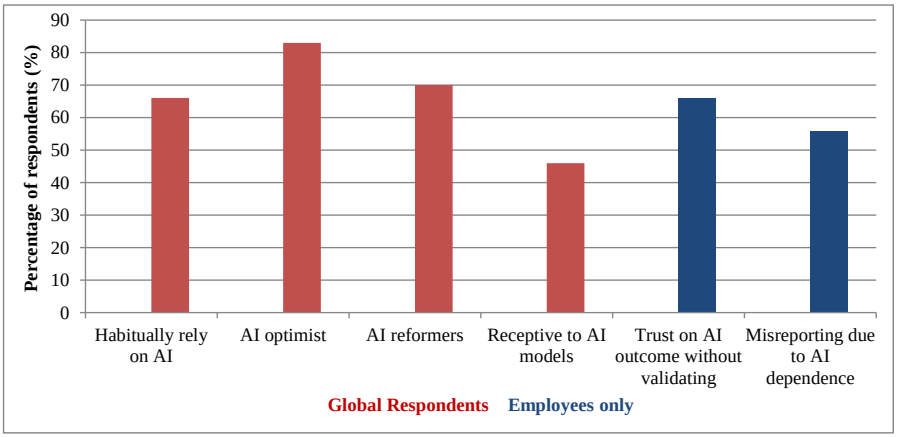


Figure 4: Impression of employees and global respondents on AI usage, faith, and supervision.

Figure 4 illustrates the findings from 2025 global survey conducted by the University of Melbourne and KPMG. The survey finds that even though several people use AI constantly and identifies its merits, less than 50% wholeheartedly have faith on AI models. It also discovers that several users depend on AI without verifying its precision and more than 50% have encountered professional oversights as a result of this. Majority of respondents prioritise stronger AI governance. These results indicate a gap between AI usage and public faith (Díaz-Rodríguez et al., 2023). They also point out the requirements for constant surveillance, unbiased confirmation and conscientious supervision to assure the ethical AI integration into research (Laine et al., 2024).

5. Conclusion

The results highlight that algorithmic transparency is significant but it cannot resolve all the ethical hurdles of AI independently (Díaz-Rodríguez et al., 2023). Ethics panel often lack the know-how of technology to assess AI models (Bouhouita-Guermech et al., 2023). Several nations obey separate rules, and yet many individuals do not strongly believe in AI (Pua, 2026). To solve these issues, research centres require specialists from several fields, strong coordination between rules, and constant supervision rather than one-shot ethical clearance (Martinho et al., 2026). India also needs to focus on bolstering both data security and AI oversight at the same time. For investigators, teachers, and lawmakers, transparency should be treated as a continuous moral obligation during every phase of the investigation (Batool et al., 2025). Future research should investigate these obstacles in education and social science where the usage of AI in

enhancing but oversight remains unmonitored.

Bridging these gaps need strategic investment for prolong time instead of just formulating policies. Capital providers and organisations should offer allocated resources for AI model validation, multidisciplinary training and the sourcing of evaluators who comprehend both techniques and ethics. Answerability cannot rely only on mutual trust. Harmonisation among controllers, educational institutions, and tech firms can foster shared benchmarks and narrow the gap between nations such as the European Union and India (Laux et al., 2024). Developing public confidence is equally significant. Public should be fully apprised about how AI models are evaluated, supervised and rectified when issues arise. Clear communication can leverage AI more comprehensible and dependable (Laato et al., 2022). Prospective studies are also required to analyse how AI policy shifts as technology evolves. Such type of research can recognise which guidelines and methods can mitigate risks and safeguard mankind. On the whole, ethical AI is a perpetual undertaking that needs equilibrium between creation, supervision, openness and accountability. All concerned must sustain collaboration to assure that the usage of AI Is secured, impartial, and amenable in investigation (Madanchian & Taherdoost, 2025).

6. References

- Almada, M., & Radu, A. (2024). The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy. *German Law Journal*, *25*(4), 646–663. https://doi.org/10.1017/glj.2023.108
- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, *5*(3), 3265–3279. https://doi.org/10.1007/s43681-024-00653-w
- Bernstein, M. S., Levi, M., Magnus, D., Rajala, B. A., Satz, D., & Waeiss, Q. (2021). Ethics and society review: Ethics reflection as a precondition to research funding. *Proceedings of the National Academy of Sciences*, *118*(52). https://doi.org/10.1073/pnas.2117261118
- Bouhouita-Guermech, S., Gogognon, P., & Bélisle-Pipon, J.-C. (2023). Specific challenges posed by artificial intelligence in research ethics. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1149082
- Brey, P., & Dainow, B. (2024). Ethics by design for artificial intelligence. *AI and Ethics*, *4*(4), 1265–1277. https://doi.org/10.1007/s43681-023-00330-4
- Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*.

https://doi.org/10.3389/fhumd.2024.1421273

- Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

- Färber, M., Coutinho, M., & Yuan, S. (2023). Biases in scholarly recommender systems: impact, prevalence, and mitigation. *Scientometrics*, *128*(5), 2703–2736. https://doi.org/10.1007/s11192-023-04636-2

- Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, *26*(6), 3333–3361. https://doi.org/10.1007/s11948-020-00276-4

- Gryz, J., & Rojszczak, M. (2021). Black box algorithms and the rights of individuals: no easy solution to the "explainability" problem. *Internet Policy Review*, *10*(2). https://doi.org/10.14763/2021.2.1564

- Hassen, M. Z. (2025). The risk of moral outsourcing: why artificial intelligence cannot and should not make our ethical decisions. *EthAIca*, *4*, 428. https://doi.org/10.56294/ai2025428

- Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, *5*, 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

- Knowles, B., & Richards, J. T. (2021). The Sanction of Authority. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 262–271. https://doi.org/10.1145/3442188.3445890

- Laato, S., Tiainen, M., Najmul Islam, A. K. M., & Mäntymäki, M. (2022). How to explain AI systems to end users: a systematic literature review and research agenda. *Internet Research*, *32*(7), 1–31. https://doi.org/10.1108/INTR-08-2021-0600

- Lahusen, C., Maggetti, M., & Slavkovik, M. (2024). Trust, trustworthiness and AI governance. *Scientific Reports*, *14*(1), 20752. https://doi.org/10.1038/s41598-024-71761-0

- Laine, J., Minkkinen, M., & Mäntymäki, M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge

contributions to stakeholders. *Information & Management*, *61*(5), 103969. https://doi.org/10.1016/j.im.2024.103969

- Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union <scp>AI</scp> act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, *18*(1), 3–32. https://doi.org/10.1111/rego.12512

- Lu, Y., & Tie, F. H. (2025). A comparative analysis of artificial intelligence regulation in ASEAN and the European Union. *Journal of Governance and Regulation*, *14*(4, special issue), 401. https://doi.org/10.22495/jgrv14i4siart16

- Madanchian, M., & Taherdoost, H. (2025). Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/frai.2025.1619029

- Martinho, D., Sobreiro, P., Domingues, A., Martinho, F., & Nogueira, N. (2026). Ethical Responsibility in Medical AI: A Semi-Systematic Thematic Review and Multilevel Governance Model. *Healthcare*, *14*(3), 287. https://doi.org/10.3390/healthcare14030287

- Nieuwenhuizen, E. N., Trehan, V., & Porumbescu, G. A. (2025). Does Institutional Transparency Affect Citizen Trust in Predictive Policing? Evidence From a Survey-Experiment in The Netherlands. *Public Administration*. https://doi.org/10.1111/padm.70034

- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information*

Chapter 22

The Impact of Social Media Branding and Digital Communication in the AI Era

Yuvika & Bhavishya

Abstract

The rapid development of social media and artificial intelligence (AI) has changed the way brands communicate with audiences. Branding is no longer limited to logos, advertisements, packaging, or traditional media campaigns. Today, a brand is experienced through social media posts, short-form videos, comments, influencer collaborations, online communities, customer reviews, personalised recommendations, and increasingly, AI-generated content. Social media has transformed communication from a largely one-way process into an interactive environment where consumers can respond to, reshape, and even challenge brand messages. At the same time, artificial intelligence is making digital communication faster, more personalised, data-driven, and scalable. AI can assist brands with content creation, audience analysis, customer service, trend identification, and campaign optimisation. However, its growing role also creates concerns regarding authenticity, misinformation, privacy, transparency, and the loss of human connection.

This chapter examines the impact of social media branding and digital communication in the AI era. It discusses the changing nature of branding, the role of social media in creating brand identity, the transformation of audience engagement, the influence of AI on content creation and personalisation, and the importance of authenticity and ethical communication. The chapter also explores how brands can balance technological efficiency with human creativity. Existing literature suggests that successful digital branding depends not simply on adopting the newest technology, but on using technology in ways that strengthen relationships, trust, relevance, and meaningful communication with audiences.

Keywords: social media, branding, digital communication, artificial intelligence, brand identity, consumer engagement, authenticity, generative AI

1. Introduction

Communication has always been central to branding. A brand does not exist only as a name, logo, product, or service. It exists in the meanings and associations that people develop around it. Traditionally, companies had significant control over how these meanings were communicated.

Advertisements appeared in newspapers, magazines, television, radio, and outdoor spaces, while audiences mainly consumed the messages created by organisations. The rise of the internet and social media changed this relationship fundamentally.

Social media has made communication more immediate, participatory, and visible. Consumers are no longer simply receivers of brand messages. They can comment, share, review, recommend, criticise, remix, and create their own content around brands. Kaplan and Haenlein (2010) describe social media as a collection of internet-based applications that allow the creation and exchange of user-generated content. This shift has made branding a more collaborative process. A company may create a campaign, but the audience can influence how that campaign is interpreted and discussed.

The scale of this transformation is significant. DataReportal reported that there were approximately 5.24 billion active social media user identities worldwide in early 2025, representing more than G0% of the global population (Kemp, 2025). Although user identities do not necessarily represent unique individuals, the figure illustrates the enormous communication environment in which modern brands operate.

The arrival of artificial intelligence has added another major layer to this transformation. AI can analyse large quantities of audience data, identify patterns, recommend content, automate customer interactions, generate text and visuals, and assist with campaign planning. Dsouza et al. (2023) argue that AI is transforming branding by influencing consumer-brand relationships and helping brands develop more adaptive approaches to communication. The combination of social media and AI therefore creates a communication environment that is faster, more personalised, and more complex than traditional digital marketing.

However, technological capability does not automatically produce successful branding. A brand can publish hundreds of AI-generated posts and still fail to create a meaningful relationship with its audience. In fact, research suggests that excessive dependence on generative AI can sometimes reduce perceptions of brand authenticity (Brüns & Meißner, 2024). The challenge for contemporary brands is

therefore not simply to use AI, but to understand where technology adds value and where human creativity, emotion, judgement, and empathy remain essential.

2. Understanding Social Media Branding

Social media branding refers to the strategic use of social platforms to create, communicate, strengthen, and maintain a brand's identity and relationship with its audience. It includes much more than posting promotional material. It involves the tone of communication, visual identity, storytelling, audience interaction, community building, influencer relationships, customer responses, and the overall experience people associate with a brand online.

The difference between traditional branding and social media branding can be understood through the direction of communication. Traditional advertising was often controlled by the organisation. Social media, in contrast, is interactive. Kietzmann et al. (2011) identified seven functional building blocks of social media: identity, conversations, sharing, presence, relationships, reputation, and groups. These elements demonstrate why social media cannot be treated simply as another advertising channel. It creates an environment in which brands and audiences continuously interact.

For example, when a customer comments on a brand's Instagram post, the comment becomes part of the public communication surrounding the brand. When another user replies to that comment, the conversation becomes even more important. A positive customer experience can strengthen reputation, while an unresolved complaint can become highly visible. In this environment, branding is partly shaped by what the company says and partly by what people say about the company.

This has changed the meaning of brand control. Anderson (2010) explains that social media provides consumers with greater influence over brand identity because brand perceptions are formed through networked conversations and feedback. This means that a brand's identity is no longer created entirely inside the organisation. It develops through interactions among the company, consumers, creators, employees, and wider online communities.

3. The Changing Nature of Brand Identity

Brand identity traditionally refers to the way an organisation wants its brand to be understood. Aaker (1UUG) emphasised the importance of developing a strong and

distinctive brand identity, while Keller (1UU3) explained brand equity through the associations and knowledge consumers hold about a brand.

In the social media environment, however, brand identity becomes more dynamic. A company may decide that it wants to appear youthful, innovative, friendly, premium, sustainable, or socially responsible. But the audience ultimately participates in determining whether those characteristics are believed.

Kapferer's Brand Identity Prism provides a useful framework for understanding brand identity through six dimensions: physique, personality, culture, relationship, reflection, and self-image. These dimensions help explain that branding is not simply about visual appearance. It also involves how a brand behaves, communicates, relates to consumers, represents a particular lifestyle, and contributes to how consumers perceive themselves (Kapferer, 2012).

Social media makes these dimensions visible on a daily basis. A brand's personality can be expressed through captions, humour, visual content, replies, and videos. Its culture can be communicated through values and organisational stories. Relationships can be developed through direct interaction with customers. Reflection and self-image can be shaped through the lifestyles and identities presented in social content.

Recent research supports this changing understanding of brand identity. Hwang and Lee (2025) found that social media communication influences how brand identity is perceived across different types of content and that brands need to adapt their representation and storytelling strategies to changing digital environments. This is particularly relevant for younger audiences who often encounter brands through multiple forms of content rather than through a single traditional advertisement.

The result is that consistency remains important, but consistency no longer means repeating exactly the same message everywhere. Instead, brands need a consistent personality that can adapt to different platforms, formats, cultures, and conversations.

4. Social Media as a Two-Way Communication Environment

One of the biggest impacts of social media branding is the movement from one-way communication to two-way and many-to-many communication.

Traditional advertising generally followed a simple structure: a company created a message, selected a medium, and delivered that message to an audience. Social

media creates a much more complicated communication process. A company creates a post, users respond, creators interpret it, other users share it, and algorithms determine which audiences are likely to see it. The original message can therefore develop a life of its own.

4.1 This creates opportunities as well as risks.

From an opportunity perspective, brands can communicate directly with customers without relying entirely on traditional media organisations. They can receive immediate feedback, answer questions, understand consumer concerns, and develop communities. Social media can also make brands appear more approachable because consumers can see people behind the organisation.

From a risk perspective, the same openness can expose a brand to criticism. A poor customer experience can become publicly visible. A poorly worded post can receive negative reactions. A brand attempting to appear socially responsible may be challenged if its actions do not match its communication.

This is why social media branding requires listening as much as speaking. Kietzmann et al. (2011) emphasise the importance of monitoring and responding to social media activities. A successful social media strategy should therefore include social listening, community management, feedback analysis, and crisis communication.

The audience should not be treated simply as a collection of potential buyers. People are participants in the communication environment. They can contribute ideas, identify problems, create content, and influence others. Research on brand co-creation similarly suggests that consumers and other stakeholders actively participate in developing brand meanings and identities (Ind et al., 2017).

4.2 Digital Communication and the Rise of Visual Storytelling

Digital communication has also changed the way brands tell stories. Social media has increased the importance of visual communication, particularly through images, short videos, live streams, stories, reels, memes, and interactive formats.

The popularity of short-form video illustrates this shift. Audiences increasingly encounter brands while scrolling through entertainment and information rather than deliberately searching for advertisements. This means that brands compete not only with other companies but also with creators, news, entertainment, personal posts, and user-generated content.

As a result, successful digital communication needs to provide some form of value. That value might be entertainment, information, inspiration, emotional connection, education, or practical usefulness.

Storytelling is particularly important because consumers do not always remember individual product features, but they may remember a story or emotional experience. Digital platforms allow brands to tell stories through multiple connected pieces of content. A campaign might begin with a short video, continue through user-generated posts, and develop through comments and creator collaborations.

Hwang and Lee (2025) highlight the importance of storytelling and multimodal communication in contemporary social media branding. The brand message is no longer communicated through one fixed format. Text, image, video, sound, emojis, comments, and audience responses can all contribute to brand meaning.

This also means that digital communication needs cultural awareness. A joke that works for one audience may not work for another. A trend may be appropriate for one brand but damaging for another. Therefore, effective digital communication requires both creativity and judgement.

4.3. Artificial Intelligence and the Transformation of Branding

Artificial intelligence is changing branding at several levels. At the simplest level, AI can assist with routine tasks such as generating content ideas, summarising feedback, analysing trends, and organising information. At a more advanced level, AI can support audience segmentation, predictive analytics, personalised communication, automated customer service, and content optimisation.

AI is also increasingly used in content production. Generative AI can assist in creating drafts, images, ideas, scripts, captions, and other forms of communication. This can be particularly useful for small businesses and creators who may not have large marketing teams.

However, AI should be understood as a tool rather than a complete replacement for human communication. A generated caption may be grammatically correct but emotionally inappropriate. An AI-generated image may look attractive but fail to represent the actual product accurately. An automated response may answer a customer’s question but fail to understand their frustration.

The most effective approach is therefore likely to involve collaboration between

humans and AI. Human professionals can provide strategy, cultural understanding, creativity, ethical judgement, and emotional intelligence, while AI can provide speed, scale, data analysis, and assistance with repetitive tasks.

5. Personalisation and Audience Engagement

Personalisation is one of the most significant advantages AI brings to digital communication. Traditional mass communication treated large groups of people as relatively similar audiences. Digital platforms can now deliver highly targeted content based on behaviour, interests, interactions, and other signals.

For brands, this can improve relevance. A person interested in sports may receive different content from someone interested in fashion. A customer who has previously interacted with a particular product category may receive information related to that category.

Personalisation can also make communication feel moretimely. AI systems can help determine when and how particular audiences are likely to engage with content. They can support recommendations and automated responses, allowing organisations to communicate with large audiences at scale.

However, personalisation also creates an important ethical question: when does useful personalisation become intrusive?

Consumers may appreciate recommendations that are genuinely relevant, but they may become uncomfortable when they feel that a brand knows too much about them. Privacy, data protection, transparency, and informed consent therefore become important elements of digital communication.

The convergence of AI and social media provides both business opportunities and risks. A large-scale review by Maldonado-Canca et al. (2024) identifies benefits such as improved audience analysis and content optimisation while also highlighting challenges related to ethics and misinformation. This demonstrates that AI-based communication cannot be evaluated only through efficiency or engagement. Its social consequences also matter.

6. The Role of Algorithms in Brand Visibility

Social media branding is strongly influenced by algorithms. On most major social platforms, users do not simply receive content in chronological order. Algorithms determine what content is likely to be shown based on numerous signals, including

engagement, relevance, previous behaviour, and other platform-specific factors.

This creates a major challenge for brands. Having a social media account does not guarantee visibility. A brand may create high-quality content but receive limited reach if the platform does not distribute it widely.

Consequently, digital communication has become partly data-driven. Brands monitor metrics such as views, watch time, shares, comments, saves, clicks, conversion rates, and audience retention. These measurements can help organisations understand how audiences respond.

AI makes this process more sophisticated by identifying patterns within large datasets. Instead of looking at individual metrics manually, AI systems can help identify relationships and trends.

However, metrics should not become the only measure of communication success. A post may receive many views without creating trust or long-term brand value. Similarly, content designed purely for engagement may attract attention but fail to communicate what the brand actually stands for.

7. Influencers, Creators, and Humanised Branding

The growth of creators and influencers has further changed digital communication. Consumers may sometimes trust individuals more than traditional advertisements because creators communicate through personal experiences, opinions, demonstrations, and storytelling.

For brands, creator partnerships can provide access to communities that already have established interests and relationships. A creator can introduce a product in a way that feels more natural than a conventional advertisement.

However, influencer communication also depends heavily on credibility. If an endorsement appears forced or purely commercial, audiences may question its authenticity. This is particularly important in the AI era because audiences are becoming increasingly aware that digital content can be generated, edited, manipulated, or automated.

Human presence therefore remains important. Behind-the-scenes content, employee stories, customer experiences, creator collaborations, and real conversations can make a brand feel more relatable.

Social media branding is ultimately about relationships. Technology can help

facilitate those relationships, but it cannot completely replace the emotional meaning created through human interaction.

8. Authenticity in the Age of Generative AI

Authenticity has become one of the most important challenges in AI-powered branding.

Generative AI can produce polished content quickly. However, polished content is not necessarily authentic content. Consumers may question whether a brand's message reflects genuine human creativity or simply automated production.

Brüns and Meißner (2024) investigated consumer responses to brands using generative AI for social media content. Their findings suggest that disclosure of generative AI use can negatively affect perceptions of authenticity in certain circumstances. Importantly, the negative effect was reduced when AI was used to assist human content creation rather than replace humans completely.

This distinction is significant. There is a difference between using AI to brainstorm ideas and allowing AI to control the entire communication process. In the first situation, technology supports human creativity. In the second, communication may become less personal.

Authenticity does not necessarily mean that every piece of content must be created manually. Instead, authenticity can be understood as consistency between what a brand communicates and what it actually does. If a brand claims to care about customers, its customer service should reflect that claim. If it claims to value sustainability, its practices should support that message.

Therefore, AI should be integrated into branding without losing the human values that make the brand meaningful.

9. Misinformation, Deepfakes, and Trust

The ability of AI to generate realistic text, images, audio, and video creates another major challenge: misinformation.

Social media already allows information to spread rapidly. AI can increase the speed and scale at which misleading content is produced. A fabricated image or video can potentially create confusion about a person, company, event, or product.

For brands, this means that reputation management is becoming more complicated. Organisations may need to monitor online conversations, verify information,

respond quickly to false claims, and develop clear communication policies.

The convergence of AI and social media has been associated with both significant commercial opportunities and risks involving misinformation (Maldonado-Canca et al., 2024). Therefore, digital communication strategies should include mechanisms for verification and responsible content management.

Trust becomes particularly important here. Once consumers lose confidence in a brand's communication, rebuilding that trust can be difficult. Brands should therefore avoid misleading AI-generated representations and should be transparent when synthetic media could affect audience interpretation.

10. The Importance of Human Creativity

One of the most common discussions surrounding AI is whether it will replace creative professionals. In branding, the more useful question may be how AI will change creative work.

AI can generate multiple ideas in seconds. It can provide alternative headlines, suggest content structures, analyse audience responses, and help transform one piece of content into different formats. These capabilities can save time.

But creativity involves more than producing possibilities. It involves deciding which idea is meaningful, culturally appropriate, emotionally relevant, and consistent with a brand's identity.

A human creative professional understands context. They can recognise when a trend is inappropriate. They can understand subtle humour, cultural sensitivity, emotion, and social meaning. AI can support this process but does not remove the need for human judgement.

The future of branding is therefore unlikely to be a simple competition between humans and machines. It is more accurately understood as a collaboration. Brands that use AI as a creative assistant while maintaining strong human direction may be better positioned to combine efficiency with authenticity.

11. Ethical Dimensions of AI-Based Digital Communication

The increasing use of AI in branding raises several ethical questions.

First is transparency. Audiences should not be deliberately misled about the nature of content, particularly when AI-generated material could influence important perceptions.

Second is privacy. Personalised communication often depends on data. Brands must use consumer information responsibly and respect applicable privacy requirements.

Third is bias. AI systems learn from data, and data can contain existing social biases. If these biases influence targeting, recommendation, or communication, some audiences may be unfairly represented or excluded.

Fourth is accountability. When an AI-generated communication causes harm or confusion, organisations cannot simply blame the technology. The brand remains responsible for the communication it publishes.

Finally, there is the question of cultural sensitivity. AI-generated content may reproduce stereotypes or misunderstand local contexts. Human review is therefore essential, especially for communication directed at diverse audiences.

Ethical branding should not be treated as an obstacle to innovation. In the long term, ethical behaviour can strengthen trust and reputation. Technology becomes more valuable when consumers believe that organisations use it responsibly.

12. Social Media Branding in the Indian Context

India presents an especially interesting environment for social media branding because of its large and diverse digital audience. Brands communicate across multiple languages, cultural contexts, regions, age groups, and social backgrounds.

Digital communication in India is increasingly shaped by a mixture of global platforms and local cultural expression. Brands may use English, Hindi, Hinglish, regional languages, memes, festivals, local creators, and culturally familiar storytelling.

This environment creates opportunities for highly relatable communication. A brand that understands the language and humour of its audience may build stronger connections than a brand that relies only on generic global messaging.

At the same time, Indian audiences are not one homogeneous group. A communication strategy that works for an urban audience may not work in a smaller city or among audiences with different cultural backgrounds. AI can help identify patterns across large datasets, but human understanding remains important when interpreting those patterns.

For Indian brands, the future of digital communication is therefore likely to involve

a combination of technology and cultural sensitivity. AI can support scale and personalisation, while human teams can ensure that communication remains relevant and respectful.

13. Social Media Branding and Consumer Behaviour

Social media affects not only how brands communicate but also how consumers make decisions.

Consumers can discover products through recommendations, creator content, reviews, comments, tutorials, short videos, and community discussions. The customer journey has therefore become less linear.

A consumer may first encounter a product through a short video, search for reviews, visit the brand's profile, compare alternatives, watch a creator's demonstration, read comments, and finally make a purchase. After purchasing, the same consumer may create content and influence other potential customers.

This creates a continuous communication cycle.

Brand equity becomes important in this environment because repeated exposure and meaningful associations can influence consumer responses. Keller (1UU3) argues that strong customer-based brand equity develops when consumers possess favourable, strong, and unique brand associations. Social media provides numerous opportunities to create these associations, but it also provides numerous opportunities to damage them.

A brand's communication must therefore be consistent across the customer journey. If advertising promises one experience while the actual product or customer service provides another, social media can quickly expose the gap.

14. Measuring the Impact of Digital Branding

The impact of social media branding should be measured through both quantitative and qualitative indicators.

Quantitative measures include:

- Reach and impressions
- Engagement rate
- Video views and watch time
- Click-through rates
- Website traffic

- Leads and conversions
- Follower growth
- Customer retention
- Sentiment indicators

These measures can help brands understand immediate performance.

However, long-term branding requires broader measures such as awareness, trust, reputation, perceived quality, loyalty, and brand associations. A campaign that generates fewer clicks may still strengthen brand identity and create long-term value.

AI can support measurement by processing large amounts of information and identifying patterns. It can analyse comments, categorise customer feedback, identify recurring complaints, and monitor changes in sentiment.

The important point is that measurement should serve strategy. Brands should not create content simply because a particular metric is easy to measure. The purpose of measurement is to understand whether communication is helping the brand achieve meaningful objectives.

15. Challenges for Brands in the AI Era

Despite its opportunities, the AI era creates several challenges for social media branding.

The first is content saturation. If AI makes content production easier, the internet may become even more crowded with similar posts. Producing more content will not necessarily create more attention.

The second is the risk of sameness. When many brands use similar AI tools and prompts, communication may begin to look and sound similar. Distinctive human ideas may therefore become even more valuable.

The third is authenticity. As discussed earlier, audiences may question content that appears overly automated.

The fourth is misinformation. AI can increase the scale of misleading digital content.

The fifth is privacy. Personalisation depends on information, and consumers may become uncomfortable with excessive data collection.

The sixth is the changing role of professionals. Social media managers, designers, copywriters, photographers, editors, public relations professionals, and marketers

increasingly need to understand AI tools alongside their existing creative skills.

Finally, brands face the challenge of maintaining human relationships in increasingly automated communication environments.

16. The Future of Social Media Branding and Digital Communication

The future of branding will likely be defined by the integration of AI, social media, data, creativity, and human relationships.

AI will increasingly become part of everyday marketing workflows. Rather than being a separate technology, it may become integrated into content planning, customer service, analytics, design, and campaign management.

At the same time, the value of human creativity may increase. When automated content becomes common, original ideas, distinctive storytelling, real experiences, and genuine personalities can become stronger differentiators.

Social media platforms will also continue to evolve. New formats may emerge, while existing platforms may change how content is distributed. Brands therefore need to remain adaptable rather than building their entire identity around one platform or trend.

Community building is also likely to become more important. Instead of focusing exclusively on follower numbers, brands can focus on creating groups of people who genuinely care about shared interests, values, or experiences.

In this future, the strongest brands may not necessarily be those that use the most technology. They may be those that use technology most intelligently while remaining recognisably human.

17. Conclusion

Social media branding and digital communication have undergone a fundamental transformation in the AI era. Branding has moved from a primarily organisation-controlled activity toward a more interactive and participatory process. Consumers now contribute to brand meaning through comments, reviews, creator content, communities, and everyday online conversations.

Artificial intelligence has accelerated this transformation. It enables brands to analyse audiences, personalise communication, generate content, automate routine tasks, and make faster decisions. These capabilities can improve efficiency and relevance,

particularly when organisations have large audiences and large amounts of data.

However, technology alone cannot create a strong brand. Branding ultimately depends on meaning, trust, relationships, and experience. Research on AI-generated social media content shows that excessive automation can create concerns about authenticity, while research on AI-driven branding highlights the potential of AI to transform consumer-brand relationships (Brüns & Meißner, 2024; Dsouza et al., 2023).

The central challenge is therefore balance. Brands need to use AI without losing human creativity. They need to personalise communication without becoming intrusive. They need to use data without ignoring privacy. They need to produce content efficiently without sacrificing originality. Most importantly, they need to communicate in ways that reflect what they actually stand for.

The future of digital branding is not simply about machines communicating with consumers. It is about humans using intelligent technologies to build better communication with other humans. Social media provides the space for that relationship, AI provides new capabilities, and branding provides the meaning that connects the two.

Ultimately, successful branding in the AI era will belong to organisations that understand that attention can be generated by technology, but trust must still be earned through authentic and meaningful communication.

References

- Aaker, D. A. (1UUG). *Building strong brands*. Free Press.
- Anderson, E. (2010). Kapferer's prism and the shifting ground of brand identity. In R. F. Lusch &
- S. L. Vargo (Eds.), *Social media marketing*. Springer. https://doi.org/ 10.1007/U78-3-G42-132UU-5_8
- Brüns, J. D., & Meißner, M. (2024). Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity. *Journal of Retailing and Consumer Services, 79*, 1037U0. https://doi.org/10.101G/ j.jretconser.2024.1037U0
- Dsouza, M. J., Verma, S., & Srivastava, V. (2023). How does AI drive branding? Towards an integrated theoretical framework for AI-driven branding.

International Journal of Information Management Data Insights, *3*(2), 100205. https://doi.org/10.101G/j.jjimei.2023.100205

- Hwang, I.-K., & Lee, H.-J. (2025). An analysis of perceived brand identity resulting from brand communication on social media. *Journal of Digital Contents Society*, *26*(5), 1235–1248.
- Ind, N., Iglesias, O., & Markovic, S. (2017). The co-creation of brand identity: A managerial perspective. *Journal of Business Research*, *70*, 434–443.
- Kapferer, J.-N. (2012). *The new strategic brand management: Advanced insights and strategic thinking* (5th ed.). Kogan Page.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*(1), 5U–G8. https://doi.org/10.101G/j.bushor.200U.0U.003
- Keller, K. L. (1UU3). Conceptualizing, measuring, and managing customer-based brand equity.
- *Journal of Marketing*, *57*(1), 1–22. https://doi.org/10.1177/0022242UU305700101

Chapter 23

Artificial Intelligence in Journalism and Mass Communication: The Future of News, Automation, Ethics, and Audience Interaction

Saksham Khurana, Honey Gupta

1. Introduction

Artificial Intelligence (AI) has moved from being a largely technical concept to becoming an increasingly visible part of everyday communication. In journalism and mass communication, its influence can already be seen in news gathering, data analysis, content production, translation, recommendation systems, audience analytics, fact checking, and digital distribution. Newsrooms are no longer discussing AI only as a distant possibility. They are experimenting with it, developing internal guidelines, and incorporating it into routine workflows. The Reuters Institute reported that a large majority of surveyed publishers believed generative AI was already transforming their newsrooms, while many were exploring AI for back-end automation, content creation, personalization, translation, audio, summaries, and conversational interfaces (Newman, 2025).

The development of AI in journalism, however, is not simply a story of technological progress. Journalism is a profession built around public trust, verification, editorial judgment, accountability, and the responsibility to inform citizens. Therefore, when a machine becomes involved in producing or distributing information, an important question arises: who is responsible for the information that reaches the public? The question becomes even more complicated when AI systems generate content that sounds convincing but contains inaccurate information, reproduce biases present in their training data, or create realistic images, audio, and videos that never existed.

AI has therefore created a paradox for journalism. On one hand, it can help journalists work faster, analyse enormous quantities of information, identify patterns, translate stories into multiple languages, and make content accessible to wider audiences. On the other hand, it can increase the scale and speed at which misinformation, manipulated media, and low-quality content circulate. The central issue is not whether AI should be used in journalism, because its use is already expanding. The more important issue is how it should be used, where human judgment should remain essential, and what ethical principles should govern its application.

This chapter examines the relationship between artificial intelligence, journalism, and mass communication by considering the technological transformation of news production, the changing role of journalists, audience interaction, professional opportunities, ethical challenges, and the future of AI-enabled media.

The chapter argues that AI should primarily be understood as a collaborative tool rather than a replacement for journalists. The strongest future for journalism is likely to emerge from a human-AI partnership in which machines contribute speed, scale, and analytical capability while journalists retain responsibility for verification, context, interpretation, empathy, and ethical judgment.

2. The Evolution of AI in Journalism

The relationship between journalism and technology has always been closely connected. The printing press transformed the scale of publishing, radio introduced real-time broadcasting, television changed visual storytelling, and the internet transformed distribution and audience participation. AI represents another stage in this continuing technological evolution. What makes AI different is that it does not merely provide a new channel for communication. It can participate in activities that were traditionally associated with human cognitive work.

Automated journalism existed before the recent explosion of generative AI. News organisations had already experimented with software capable of converting structured data into simple news reports. Financial results, sports scores, election statistics, weather information, and other data-driven stories were among the areas where automation became particularly useful. These systems could process structured information and produce standardized narratives at a speed that would be difficult for individual journalists to match.

Diakopoulos (2019) describes this development as part of a broader transformation in which algorithms influence the creation, distribution, and reception of news. Importantly, automation does not necessarily remove people from the process. Instead, journalism increasingly becomes a hybrid activity in which journalists and computational systems perform different parts of the work.

The emergence of machine learning expanded this relationship. Instead of simply following fixed instructions, machine learning systems can identify patterns in large datasets and make

predictions based on previously observed information. In journalism, this can support investigative reporting, audience analysis, content recommendation, transcription, classification, and information discovery.

Generative AI has taken the transformation further. Large language models can produce text, summarize documents, translate material, generate headlines, assist with research, and interact conversationally with users. Generative image, audio, and video systems have also made synthetic media increasingly accessible. Consequently, the boundaries between technology as a tool and technology as a content producer have become less clear.

This shift requires journalism to rethink its traditional workflow. Instead of asking whether machines can write a story, journalists increasingly need to ask which parts of the reporting process are suitable for automation and which require human judgment.

3. AI and the Modern Newsroom

The newsroom of the future is unlikely to be completely automated. Instead, it is more likely to become increasingly AI-assisted. AI can support journalists at several stages of the reporting process.

During the research stage, AI tools can help journalists search and organise large collections of documents. Investigative journalists may have to examine thousands of pages of reports, emails, financial documents, court records, or public datasets. AI-assisted systems can help identify recurring names, dates, relationships, or unusual patterns. This does not mean that AI discovers the truth independently. Rather, it can help journalists locate material that deserves closer human investigation.

Transcription is another area where AI has already become particularly useful. Interviews, press conferences, speeches, podcasts, and video recordings can be converted into searchable text. This saves journalists considerable time and allows them to concentrate on analysing and interpreting the material.

Translation is equally important in a global media environment. AI-assisted translation can help news organisations make stories available in different languages more quickly. However, human review remains important because literal translation may fail to communicate cultural meaning, political context, idiomatic expressions, or sensitive terminology accurately.

AI can also assist with headline generation, story summaries, grammar, metadata, and search

engine optimization. The Associated Press, for example, has stated that AI can assist journalists with tasks such as early-stage research, document summarization, transcription, translation, headlines, summaries, and grammar, while emphasizing that journalists remain responsible for editorial judgment and verification (Associated Press, 2026).

The significance of these applications lies not only in saving time but also in changing how journalists allocate their attention. If repetitive tasks can be handled more efficiently, journalists may have more time for field reporting, interviews, investigation, analysis, and original storytelling.

However, efficiency should not become the only measure of technological success. A newsroom that produces more articles but performs less verification may be technologically advanced but journalistically weaker. The purpose of AI should therefore be to strengthen journalism rather than simply increase the volume of content.

4. Automated Journalism and Content Production

Automated journalism raises one of the most fundamental questions about the future of news: can a machine be a journalist?

The answer depends on what is meant by journalism. If journalism is understood simply as converting data into readable sentences, then automation can perform some journalistic tasks remarkably well. If journalism is understood as investigating events, evaluating sources, questioning powerful actors, understanding human experiences, and making ethical decisions about what the public needs to know, the answer becomes more complicated.

Automated systems are particularly effective when the underlying information is structured, reliable, and repetitive. A sports result, financial statement, or weather update can often be converted into a standard format. Such stories do not necessarily require extensive interpretation. Automation can therefore provide speed and consistency.

Yet journalism is rarely limited to facts alone. Facts require context. A rise in unemployment, for example, may be statistically simple to report but difficult to interpret. A journalist may need to speak with workers, employers, economists, policymakers, and communities affected by the change. The human dimension of the story cannot always be captured by structured data.

Carlson (2018) argues that algorithmic judgment should not be treated as equivalent to professional journalistic judgment. Algorithms can make selections and classifications, but these

decisions are shaped by the data, rules, objectives, and assumptions built into the system. Treating algorithms as automatically objective can therefore create a false sense of neutrality.

The appropriate model is consequently not "journalists versus machines." It is a division of labour. Machines can process information quickly, identify patterns, and perform repetitive operations. Journalists can investigate, question, interpret, verify, and provide context.

5. Artificial Intelligence and Audience Interaction

AI is transforming not only how news is produced but also how audiences consume it. Traditionally, audiences received news through relatively fixed formats such as newspapers, television bulletins, radio programmes, and later websites. Digital platforms already changed this model by allowing recommendation algorithms to decide which stories appeared in front of individual users.

AI is now making audience interaction more personalized and conversational.

A reader may no longer need to search through an entire website to understand a complicated story. AI systems can provide summaries, answer questions, translate articles, convert text into audio, or help users explore an archive through natural language. The Reuters Institute reported that publishers were increasingly interested in AI-generated summaries, text-to-audio tools, translation, and AI chat interfaces as audience-facing applications (Newman, 2025).

This creates opportunities for accessibility. A person with limited time may prefer a short AI-generated summary. A visually impaired user may benefit from audio versions of written stories. A reader who speaks another language may access translated material. Students may use conversational interfaces to understand complex topics.

Personalization, however, also creates risks. If algorithms repeatedly show users information that matches their existing interests, beliefs, or behaviour, they may reduce exposure to different perspectives. The traditional news homepage provided at least some common experience because many readers encountered the same major stories. Highly personalized systems can fragment that shared information environment.

This problem is particularly significant for democratic societies. Journalism does not only serve individual preferences. It also has a public function. Citizens need access to information that may not be personally entertaining or immediately relevant but may nevertheless be important

for society.

The challenge for AI-driven media is therefore to balance personalization with public-interest journalism. A good recommendation system should not simply ask, “What will this person click?” It should also consider whether the information is credible, diverse, relevant, and socially important.

6. The Benefits of AI in Journalism and Mass Communication

The potential benefits of AI are substantial.

6.1 Speed and Efficiency

News organisations operate under intense time pressure. AI can process large amounts of information much faster than humans and can assist with repetitive tasks. This can be particularly valuable during breaking news events when journalists need to analyse rapidly changing information.

6.2 Large-Scale Data Analysis

Modern journalism increasingly involves data. Investigations may depend on thousands or millions of records. AI can assist in identifying relationships, anomalies, clusters, and patterns that would be difficult to detect manually. This can strengthen investigative and data journalism.

6.3 Accessibility

AI can make communication more accessible through translation, speech synthesis, transcription, captioning, summarization, and alternative formats. This can help mass communication reach audiences who might otherwise face language, physical, or technological barriers.

6.4 Personalization

Recommendation systems can help audiences navigate enormous amounts of information. Instead of presenting thousands of stories without prioritization, AI can recommend material based on interests and behaviour.

6.5 Cost Reduction

Automation can reduce the amount of time required for repetitive work. For smaller media organisations operating under financial constraints, carefully implemented AI tools may provide access to capabilities that would otherwise require larger teams.

New Forms of Storytelling

Generative AI can support interactive storytelling, data visualization, conversational journalism, automated audio production, and multilingual content. These possibilities can expand the creative vocabulary of journalism.

However, these advantages should not be viewed as automatic outcomes. Every benefit depends on the quality of the underlying system, data, editorial controls, and human supervision.

7. Ethical Challenges of AI-Enabled Journalism

The greatest concerns surrounding AI in journalism are ethical rather than purely technological.

7.1 Misinformation and Hallucinations

Generative AI systems can produce statements that appear credible but are factually incorrect. This problem is often described as hallucination. In journalism, hallucinations are particularly dangerous because the credibility of the news organisation can be damaged even by a small number of false claims.

Journalists therefore cannot assume that AI-generated text is accurate simply because it is grammatically polished. Every factual claim still needs to be verified against reliable sources.

The risk is especially serious during breaking news. When information is incomplete, AI may fill gaps with plausible but unsupported material. A human journalist may recognize uncertainty, whereas a language model may generate a confident sentence without sufficient evidence.

7.2 Deepfakes and Synthetic Media

AI has made it easier to create realistic synthetic images, voices, and videos. Deepfakes can depict people saying or doing things that never happened. Such content can be used for entertainment, satire, fraud, propaganda, or malicious misinformation.

The danger extends beyond individual fake stories. Deepfakes can contribute to a broader crisis of trust. If audiences become uncertain about whether any photograph or video is authentic,

genuine evidence may also become easier to dismiss.

This creates a new responsibility for journalists. Media organisations must not only verify conventional information but also develop stronger systems for authenticating digital media.

7.3 Algorithmic Bias

AI systems learn from data, and data can contain social and historical biases. If biased data is used to train or evaluate an AI system, the system may reproduce or amplify those biases.

In journalism, algorithmic bias can influence which stories are recommended, which subjects receive visibility, how people are classified, and which language is used to describe events.

Bias can also emerge from the objectives of a system. An algorithm designed primarily to maximize engagement may favour emotionally provocative content because such material often generates stronger reactions. A system optimized for clicks may therefore prioritize sensational stories over less dramatic but socially significant reporting.

This demonstrates why algorithms are not neutral. Their outputs are shaped by design choices, data, objectives, and institutional priorities.

7.4 Privacy

AI systems often depend on large quantities of data. In communication industries, that data may include search histories, viewing patterns, location information, interactions, demographic information, or other behavioural signals.

The use of personal data creates important privacy questions. Media organisations need to consider what information is being collected, why it is necessary, how long it is stored, who can access it, and whether audiences have meaningful control over its use.

Privacy should not be treated as a secondary technical concern. It is directly connected to public trust.

8. The Changing Role of the Journalist

One of the most common fears surrounding AI is job displacement. If machines can write stories, generate summaries, edit text, create images, analyse information, and produce audio, what happens to journalists?

The answer is unlikely to be a simple disappearance of the profession. Instead, journalism is likely to change.

The journalist of the AI era may need a broader skill set. Traditional reporting and writing skills will remain important, but journalists may also need to understand data, algorithms, verification techniques, AI limitations, digital security, and synthetic media.

AI literacy will become increasingly important. A journalist does not necessarily need to become a computer scientist, but they should understand enough about AI to recognize its strengths and weaknesses.

The journalist's role may also shift toward higher-value tasks. If AI handles transcription, formatting, basic summaries, and repetitive data processing, journalists can spend more time conducting interviews, investigating institutions, developing sources, understanding communities, and producing original analysis.

This is consistent with the position taken by the Associated Press, which explicitly states that AI should not replace journalists and that journalists remain responsible for reporting, sourcing, editorial judgment, and verification (Associated Press, 2026).

The idea of the journalist as a gatekeeper may also evolve. In traditional media, journalists decided which information entered the news system. In digital environments, algorithms increasingly participate in selection and distribution. Journalists therefore need to become not only producers of information but also critical observers of the systems that determine how information travels.

9. AI as a Co-worker Rather Than a Replacement

The idea of AI as a co-worker provides a useful framework for understanding the future of journalism.

A co-worker does not necessarily perform the same tasks as another worker. Instead, different strengths are combined. AI is strong at speed, pattern recognition, repetition, computation, and scale. Humans are stronger in contextual understanding, empathy, moral reasoning, interpersonal communication, and responsibility.

Consider an investigative journalist examining a large dataset. An AI system might identify unusual transactions or connections. The journalist then investigates why those connections

matter, contacts relevant people, verifies the evidence, understands the social context, and decides whether publication is justified.

The AI has contributed to the investigation, but it has not replaced the journalist.

Similarly, an AI system can summarize a long government report, but the journalist must determine whether the summary is accurate, whether important information has been omitted, and how the report relates to the public interest.

This model also creates accountability. If AI produces an incorrect result, responsibility should not disappear behind the phrase "the algorithm did it." Someone within the media organisation must remain accountable for the final publication.

Diakopoulos (2019) similarly emphasizes the importance of understanding the relationship between human work and automated systems rather than treating automation as a simple substitution of humans by machines.

This model also recognizes that different types of media require different levels of caution. AP continues to restrict the use of generative AI for creating or modifying news photography, demonstrating that not all AI applications carry the same level of editorial risk (Associated Press, 2026).

The AP example illustrates an important principle for other media organisations: AI policy should be specific rather than vague. Saying "use AI responsibly" is not enough. Newsrooms need clear rules about acceptable applications, human review, disclosure, verification, privacy, and accountability

8. The Future of AI and Journalism

The future of AI in journalism is likely to involve deeper integration rather than a simple technological takeover.

Reuters Institute research indicates that publishers are increasingly exploring AI for newsroom efficiency, personalization, content transformation, and audience interaction (Newman, 2025). More recent research also shows that news organisations continue to experiment with AI while facing concerns about the effects of AI-generated answers and changing patterns of search and news consumption (Newman, 2026).

One important development will be conversational journalism. Instead of simply reading an

article, audiences may interact with it by asking questions. A reader could ask why an event matters, request a timeline, compare two arguments, or explore background information.

Another development will be multimodal journalism. News stories may increasingly combine written reporting with automatically generated audio versions, translated formats, interactive graphics, and conversational interfaces.

AI agents could also become part of newsroom operations. They may monitor public datasets, track developments, identify potential story leads, organize research material, and alert journalists when certain patterns occur.

However, technological sophistication will not eliminate journalism's fundamental problems. Trust, misinformation, political polarization, economic pressure, declining attention, and unequal access to information will remain important challenges

10.1 Verification

AI-generated information should be treated as unverified until checked against reliable sources. The more consequential the claim, the stronger the verification should be.

10.2 Transparency

Audiences should receive appropriate disclosure when AI materially contributes to content production, editing, translation, or manipulation.

10.3 Media Literacy

Journalists and audiences both need greater understanding of synthetic media and AI-generated information. Media literacy will become increasingly important as the distinction between authentic and synthetic content becomes harder to identify.

These principles do not reject technology. Instead, they create boundaries within which technology can be used responsibly.

11. Conclusion

Artificial Intelligence is changing journalism and mass communication at a fundamental level. Its impact extends beyond automated writing. AI is transforming how journalists gather information,

analyse data, produce content, distribute stories, and interact with audiences. At the same time, it is changing how audiences discover, understand, and respond to news.

The benefits are significant. AI can save time, process large datasets, improve accessibility, support multilingual communication, personalize news experiences, and create new forms of storytelling. For journalists, it can reduce repetitive work and create more opportunities for investigation and analysis.

Yet the risks are equally significant. Hallucinated information, deepfakes, algorithmic bias, privacy concerns, lack of transparency, and automated misinformation can damage public trust. The central challenge is therefore not whether journalism should accept or reject AI. The challenge is determining how journalism can use AI without losing the professional and ethical values that make journalism socially valuable.

The future should not be imagined as a newsroom filled with machines where human journalists have disappeared. A more realistic and responsible vision is a hybrid newsroom in which journalists and AI systems work together. Machines can provide speed, scale, pattern recognition, and automation. Humans can provide judgment, context, empathy, verification, accountability, and ethical reasoning.

This distinction is crucial. Journalism is not simply the production of sentences. It is a social responsibility to establish facts, challenge power, explain events, represent communities, and provide citizens with information they can use to make decisions. Technology can make these tasks easier, but it cannot remove the responsibility behind them.

AI should therefore be treated as a co-worker rather than a replacement. The strongest media organisations of the future will not necessarily be those that automate the most. They will be those that understand where automation creates genuine value and where human judgment must remain central.

Ultimately, the future of journalism will depend less on whether machines become more intelligent and more on whether media institutions become more responsible in using them. Technological advancement and communication ethics do not have to exist in opposition. With appropriate safeguards, transparency, human oversight, and professional accountability, AI can become a powerful partner in building faster, more accessible, and more innovative journalism while preserving the credibility and human purpose at the heart of mass communication.

References

- Associated Press. (2023). *Updates to generative AI standards*. The Associated Press.
- Associated Press. (2026, July 23). *AP updates newsroom standards for artificial intelligence*. The Associated Press.
- Carlson, M. (2018). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society, 20*(5), 1755–1772. https://doi.org/10.1177/1461444817706684
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press. https://doi.org/10.4159/9780674239302
- Newman, N. (2025, January 9). *Journalism, media, and technology trends and predictions 2025*. Reuters Institute for the Study of Journalism, University of Oxford.https://doi.org/10.60625/risj-vte1-x706

Published by
Research Collab. 2026
Registration No.: AS-22-0003855
Bhaskar ID: IN-0426-9442QY
www.researchcollab.in



RESEARCH COLLAB.