

Reconceptualizing Agenda-Setting Theory in the Age of Artificial Intelligence: An Algorithmic Agenda-Setting Framework for Media Influence, Trust, and Public Engagement

Tommy Kibera Kiilu*¹

¹Department of Communication, Media, Film & Theatre Studies, Kenyatta University, Nairobi, Kenya

Received 2 April 2026

Revised 29 April. 2026

Accepted 29 May. 2026

Published May. 2026

Cited as:

Tommy Kibera Kiilu, Reconceptualizing Agenda-Setting Theory in the Age of Artificial Intelligence: An Algorithmic Agenda-Setting Framework for Media Influence, Trust, and Public Engagement, J. Dig Media. AI. Vol. 3 No. 2 (2026) PP 131- 154. DOI: 10.18576/jdmai /030204

Abstract: Agenda-Setting Theory (McCombs & Shaw, 1972) established that media shape public perceptions by determining which issues are salient. This framework has guided communication research for five decades. However, the rise of Artificial Intelligence (AI), algorithmic recommendation systems, and platform-driven communication ecosystems has fundamentally restructured how information is selected, amplified, and consumed, exposing critical limitations in the theory's classical assumptions. This paper re-examines Agenda-Setting Theory in the context of AI-mediated communication and proposes the Algorithmic Agenda-Setting Framework (AASF) as a conceptual extension addressing the inadequacies of classical theory for algorithmically mediated environments. A systematic theoretical review guided by PRISMA principles was conducted, synthesizing 55 peer-reviewed studies and scholarly works (2000–2025) across communication theory, algorithmic media, platform governance, trust, and disinformation research. AI transforms agenda setting through five interconnected mechanisms: (1) algorithmic gatekeeping, (2) personalized agenda construction, (3) predictive agenda amplification, (4) trust mediation, and (5) ethical agenda distortion. These mechanisms collectively constitute the AASF, which is further contextualized through analysis of Global South media environments. The paper advances communication theory by demonstrating that algorithmic agenda setting is no longer centralized, homogeneous, or exclusively human-driven, and by introducing the AASF as a theoretically rigorous framework for AI-mediated information environments. Findings have implications for journalism practice, algorithmic accountability policy, media literacy education, democratic communication, and Global South media governance.

Keywords: Agenda-Setting Theory, Artificial Intelligence, Algorithmic Communication, Trust, Media Influence, Public Engagement, Digital Platforms, Global South Media.

1. Introduction

The relationship between media systems and public perception has occupied communication scholars for over a century. Among the most enduring theoretical contributions to this tradition is Agenda-Setting Theory, first empirically established by McCombs and Shaw (1972) in their landmark Chapel Hill study. The theory's central claim — that the media does not tell people what to think, but rather what to think about — provided a falsifiable and generative framework for studying media influence on public consciousness. Over five subsequent decades, the theory evolved through the concept of attribute agenda setting (McCombs

*Corresponding author E-mail: Kiilu.tommy@ku.ac.ke

et al., 1997), the network agenda-setting model (Guo et al., 2012), and the study of cross-media agenda-setting effects in digital environments (McCombs, 2014).

Yet the contemporary information landscape confronts this theoretical tradition with challenges its originators could scarcely have foreseen. The rise of Artificial Intelligence (AI) — particularly algorithmic recommendation systems, machine learning-driven content curation, large language models, and generative AI technologies — has introduced a fundamentally new architecture of communication that disrupts classical assumptions about who sets agendas, how they are set, and for whom (Napoli, 2019; Diakopoulos, 2019). Centralized editorial gatekeeping, once the defining mechanism of mass media agenda setting, has been replaced — or substantially complicated — by data-driven, personalized, predictive, and often opaque algorithmic decision-making.

The implications of this transformation extend well beyond communication theory. Public trust in media institutions has declined globally (Newman et al., 2023; Edelman Trust Barometer, 2024), partly because algorithmic systems have made the processes of information selection and amplification invisible to ordinary users. The proliferation of disinformation, AI-generated synthetic media, and engagement-optimizing platforms has further strained the relationship between media ecosystems and democratic public discourse (Wardle & Derakhshan, 2017; Vosoughi et al., 2018). In the Global South — including sub-Saharan African contexts such as Kenya — these dynamics intersect with mobile-first digital transitions, vernacular language gaps in AI training data, and limited platform regulatory infrastructure (Media Council of Kenya, 2025; UNESCO, 2023a).

This paper addresses three guiding research questions:

RQ1: How does AI transform the classical mechanisms of agenda setting?

RQ2: What conceptual mechanisms define algorithmic agenda setting in AI-mediated communication environments?

RQ3: How can Agenda-Setting Theory be extended to adequately account for AI-mediated communication environments, including Global South contexts?

To address these questions, the paper: (1) critically examines classical Agenda-Setting Theory in relation to AI-driven information environments; (2) identifies the mechanisms through which AI transforms agenda-setting processes; (3) analyzes the role of algorithmic systems in mediating public trust and engagement; (4) extends analysis to Global South and African media contexts; and (5) proposes the Algorithmic Agenda-Setting Framework (AASF) as a theoretically grounded conceptual extension. This paper contributes to emerging scholarship on AI-mediated digital media systems.

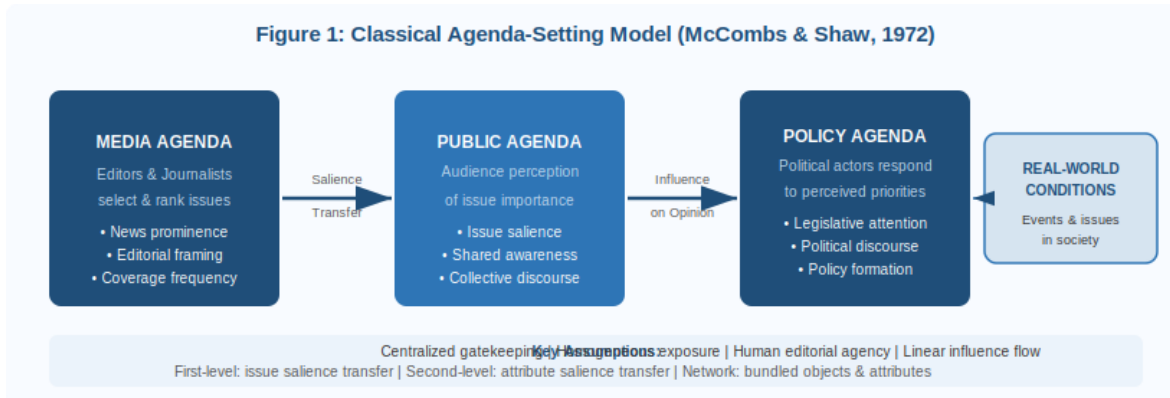
2. Literature Review

2.1 Classical Agenda-Setting Theory and Its Development

The intellectual roots of Agenda-Setting Theory extend to Lippmann's (1922) insight that public opinion is shaped by the 'pictures in our heads' — mental representations constructed primarily through media exposure. McCombs and Shaw (1972) operationalized this insight empirically, demonstrating that the issues the media emphasized during the 1968 U.S. presidential campaign directly corresponded to the issues voters identified as most important. This correlation between media salience and public salience established the agenda-setting proposition as one of the most empirically robust claims in communication research.

The theory subsequently developed in two complementary directions. First-level agenda setting established that media influences which issues gain public attention (McCombs & Shaw, 1972). Second-level or attribute agenda setting demonstrated that media also shapes how issues are perceived — which attributes and frames become salient in public understanding (McCombs et al., 1997). The network agenda-setting model further extended this by demonstrating that media simultaneously transfers the salience of bundles of objects and attributes, forming cognitive networks in audiences' minds (Guo et al., 2012). Figure 1 provides

a visual representation of the classical model.



2.2 Digital Disruption and Evolving Agenda-Setting Landscapes

The advent of digital media began disrupting traditional agenda-setting processes well before the emergence of sophisticated AI systems. The internet's fragmented, interactive, and participatory architecture complicated classical linear influence models (Althaus & Tewksbury, 2002; Boczkowski & Mitchelstein, 2010). The concept of 'intermedia agenda setting' (Golan, 2006) captured growing complexity in agenda flows across platforms, while 'audience agenda setting' — where audiences actively participate in constructing public issue salience — emerged as a meaningful complement to supply-side models (Ragas et al., 2014; Meraz & Papacharissi, 2013).

Social media further complicated these dynamics. Platforms like Twitter/X and Facebook introduced participatory agenda building, enabled viral information diffusion outside traditional editorial channels, and allowed political actors and partisan communities to shape discourse without media intermediaries. Neuman et al. (2014) demonstrated that big data methods could detect agenda dynamics in real time, pointing to methodological as well as theoretical reorientation. Yet even in these early digital frameworks, scholarship retained assumptions about human agency — whether of editors, political actors, or audiences — as the primary driver of agenda construction. The rise of AI introduced a fundamentally non-human, data-driven dimension that demands fresh theoretical attention.

2.3 Algorithmic Communication and AI-Driven Information Ecosystems

The algorithmic turn in communication has been extensively theorized across communication, media studies, and information science (Gillespie, 2014; Bucher, 2018; Diakopoulos, 2019). Recommendation algorithms — such as those employed by YouTube, Facebook, TikTok, and Google News — determine the information exposure of billions of users by optimizing content delivery for engagement metrics. This optimization logic fundamentally redefines what it means for an issue to be 'salient': salience becomes not a product of journalistic judgment, but an output of machine learning models trained on behavioral data.

Pariser's (2011) concept of the 'filter bubble' drew scholarly and public attention to the personalization problem inherent in algorithmic media: that algorithms, by curating content to match individual preferences, risk creating information environments so customized as to isolate users from discrepant perspectives. Sunstein (2017) extended this concern through the concept of 'echo chambers,' linking algorithmic personalization to political polarization and the fragmentation of shared discourse. While the empirical evidence on filter bubbles has been contested (Guess et al., 2018; Fletcher & Nielsen, 2018), the theoretical concern it raises for agenda-setting theory — that personalization fundamentally undermines the collective agenda function of media — remains theoretically significant and empirically unresolved.

Napoli (2019) provides perhaps the most comprehensive analysis of how algorithmic media systems redistribute communicative power. He argues that algorithmic systems introduce a form of 'algorithmic

audience formation' in which user behavior data, rather than journalist judgment, determines what content reaches which audiences — fundamentally inverting the classical agenda-setting model. More recently, the emergence of generative AI and large language models (LLMs) has introduced an additional layer of complexity: AI systems can now not only curate content but generate it, raising new questions about automated agenda construction and synthetic information flows (Bender et al., 2021; Weidinger et al., 2022).

2.4 AI, Trust, and the Crisis of Media Credibility

Public trust in news media has emerged as one of the most pressing concerns in contemporary communication research. The Reuters Institute Digital News Report has consistently documented declining trust across major democracies (Newman et al., 2023). The Edelman Trust Barometer (2024) similarly reports that media trust has reached historically low levels, with AI-related concerns — including fears about automated content, synthetic media, and opaque algorithmic curation — contributing significantly to this erosion.

The relationship between AI and media trust is multidimensional. Research indicates that while some users trust AI-curated news for its perceived objectivity and efficiency (Thurman et al., 2019), others are concerned about the opacity of algorithmic decision-making, the absence of editorial accountability, and the potential for manipulation (Diakopoulos, 2019; Riedl et al., 2022). The concept of 'automation bias' — the tendency to over-trust automated systems — complicates this picture further, suggesting that users may place unwarranted confidence in algorithmically curated information precisely because it appears objective (Skitka et al., 2022). Tsfati's (2010) foundational work on media skepticism established that trust is not merely an attitudinal variable but a mediating condition that shapes how audiences process media content — a dynamic that becomes considerably more complex when the audience must assess not only specific sources but the invisible algorithmic systems determining their information exposure.

2.5 Generative AI, Disinformation, and Ethical Dimensions

The intersection of generative AI, disinformation, and agenda setting represents one of the most urgent scholarly and policy challenges of contemporary communication. Vosoughi et al.'s (2018) landmark Science analysis demonstrated that false news spreads faster, deeper, and more broadly than true news in algorithmic environments — a finding with profound implications for agenda-setting theory. Wardle and Derakhshan's (2017) information disorder framework — distinguishing between misinformation, disinformation, and malinformation — provides a vocabulary for understanding these ethical dimensions.

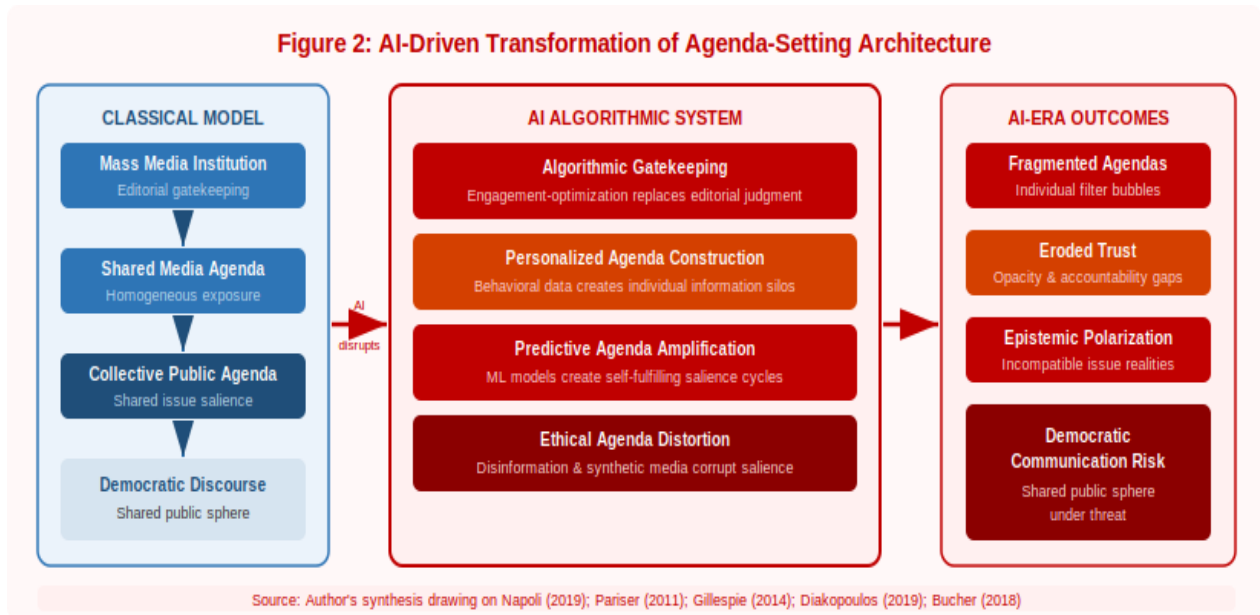
Generative AI has radically expanded these concerns. LLMs now enable the production of synthetic news articles, fabricated quotes from real public figures, deepfake videos, and AI-generated audio that can be indistinguishable from authentic content (Chesney & Citron, 2019; Bender et al., 2021). UNESCO (2023a) has identified the ethical governance of AI in media as a global priority, emphasizing that the epistemic risks of unregulated generative AI extend beyond individual disinformation incidents to the systematic corruption of information environments. The European Union's AI Act (2024) and UNESCO Recommendation on the Ethics of AI (2021) represent early regulatory responses, though enforcement in Global South contexts remains nascent (UNESCO, 2023b).

3. Theoretical and Conceptual Framework

3.1 Interrogating Classical Assumptions

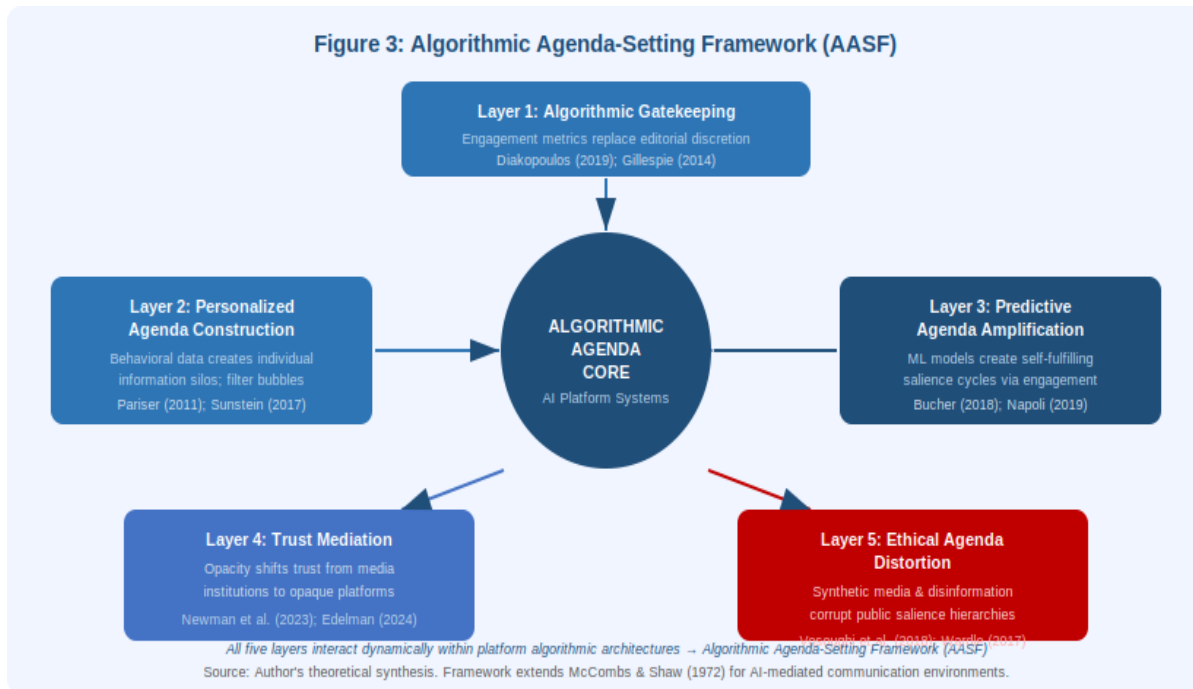
This paper draws on the classical formulation of Agenda-Setting Theory (McCombs & Shaw, 1972), its attribute-level extension (McCombs et al., 1997), and the network agenda-setting model (Guo et al., 2012) as its primary theoretical foundation. These frameworks are understood not as fixed commitments but as starting points for conceptual extension in light of technological transformation. The central analytical move

is to interrogate three core assumptions of classical agenda-setting theory that AI-mediated environments disrupt. First, the centralized gatekeeping assumption — that agenda setting is driven primarily by professional media institutions exercising editorial discretion. Second, the homogeneous exposure assumption — that audiences within a media environment are exposed to a roughly shared information environment. Third, the human agency assumption — that agenda-setting is driven by human decisions rather than automated computational processes. Figure 2 illustrates this transformation.



3.2 The Algorithmic Agenda-Setting Framework (AASF)

Building on this theoretical critique and synthesizing the scholarly literature reviewed, this paper proposes the Algorithmic Agenda-Setting Framework (AASF). The AASF conceptualizes algorithmically mediated agenda setting as a five-layered process comprising the following dynamically interacting dimensions. Figure 3 presents the framework visually.



Layer 1 (Algorithmic Gatekeeping) positions AI systems as the primary gatekeepers of issue visibility in digital environments, replacing human editorial judgment with engagement-optimized computational decision-making. Layer 2 (Personalized Agenda Construction) captures how algorithmic systems construct individualized information environments rather than a shared public agenda, producing fragmented and potentially incommensurable personal agendas. Layer 3 (Predictive Agenda Amplification) describes how machine learning models anticipate and pre-emptively amplify content most likely to generate engagement, creating feedback loops in which algorithmically predicted salience becomes actual salience. Layer 4 (Trust Mediation) recognizes that algorithmic agenda-setting operates through and reshapes the trust dynamics between users, media institutions, and platform operators, altering the conditions under which agenda effects occur. Layer 5 (Ethical Agenda Distortion) captures how disinformation, synthetic media, and engagement-optimization biases introduce systematic distortions into algorithmically constructed public agendas.

The AASF is explicitly intended not as a replacement for classical agenda-setting theory but as a conceptual extension that accounts for the distinctive features of AI-mediated communication environments. Its five layers correspond to the five thematic mechanisms identified through the systematic review and are grounded in the empirical literature synthesized in this paper. Critically, the framework is designed with cross-contextual applicability: while each of the five layers was theorized primarily from Western digital media scholarship, the AASF anticipates that its mechanisms will operate with different intensity and institutional mediation across varied regulatory and infrastructural contexts. The Global South — and sub-Saharan African media environments such as Kenya in particular — is not a peripheral addendum to the framework but a theoretically integral test case: the conditions that intensify Layers 1 through 5 (mobile-first access, vernacular language AI bias, platform governance gaps, and disproportionate disinformation exposure) are precisely the conditions under which the AASF's explanatory leverage is most consequential. This theoretical grounding is formalized in Proposition 6 (P6) below, which positions Global South amplification of AASF mechanisms as a falsifiable theoretical claim, not merely a contextual observation.

3.3 Theoretical Propositions of the AASF

Building on the five-layer AASF, this paper advances eight theoretical propositions that capture the core relational claims of the framework and can guide future empirical investigation. These propositions are derived from the systematic synthesis of the 55 included studies and are intended to provide testable hypotheses for experimental, computational, and comparative research.

P1 (Algorithmic Gatekeeping and Issue Salience): The greater the degree of algorithmic — rather than editorial — gatekeeping in a media environment, the weaker the correspondence between journalistic judgment and public issue salience. As algorithmic systems displace professional editorial discretion, the agenda-setting function migrates from identifiable institutional actors to opaque engagement-optimization logics (Gillespie, 2014; Diakopoulos, 2019).

P2 (Personalization and Agenda Fragmentation): Greater algorithmic personalization increases agenda fragmentation: as the degree of behavioral data-driven content customization increases, the overlap in issue salience hierarchies across different users within the same media environment decreases, producing the condition described here as ‘agenda pluralism’ (Pariser, 2011; Sunstein, 2017; Fletcher & Nielsen, 2018).

P3 (Predictive Amplification and Agenda Persistence): Predictive amplification by machine learning recommendation systems increases agenda persistence: content algorithmically predicted to generate high engagement is amplified, which generates the predicted engagement and reinforces further amplification, creating self-fulfilling salience cycles that sustain issue prominence beyond the timescales that editorial news cycles would naturally support (Bucher, 2018; Napoli, 2019).

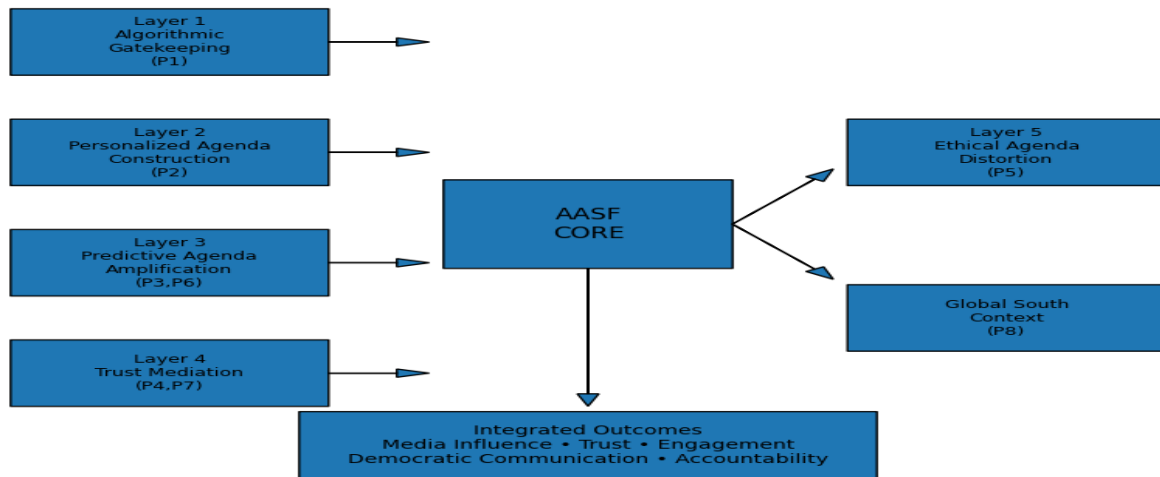
P4 (Trust and Algorithmic Literacy as Moderators of Agenda Effects): Two individual-level conditions moderate susceptibility to algorithmic agenda effects. First, trust: individuals with higher institutional media trust will critically evaluate algorithmically curated content, moderating agenda effects, while those with lower institutional trust and higher platform trust will exhibit stronger automation bias and correspondingly stronger algorithmic agenda effects (Tsfati, 2010; Riedl et al., 2022; Skitka et al., 2022). Second, algorithmic literacy: individuals with higher awareness of how recommendation systems operate will exhibit reduced automation bias and greater critical processing of algorithmically curated content, further attenuating agenda-setting effects at the individual level (UNESCO, 2023b; Resnick et al., 2013). These two conditions are theorized as jointly constituting the moderating layer of the AASF (Layer 4), with trust and literacy operating as complementary rather than redundant variables.

P5 (Ethical Agenda Distortion and Synthetic Content Proliferation): Engagement-optimized algorithmic environments introduce two compounding sources of ethical distortion. First, false and emotionally arousing content achieves disproportionate salience relative to its veracity, such that the gap between algorithmically constructed public agendas and epistemically warranted public agendas will be systematically wider in AI-mediated environments than in editorially curated environments (Vosoughi et al., 2018; Wardle & Derakhshan, 2017). Second, the deployment of generative AI for news production increases the volume and velocity of agenda-setting content while decoupling issue salience from human editorial judgment, such that algorithmically generated synthetic content competes with — and in high-volume information environments may displace — human-authored content in shaping public issue awareness (Bender et al., 2021; Weidinger et al., 2022; Goldstein et al., 2024). These two dynamics are theorized as jointly constituting AASF Layer 5 (Ethical Agenda Distortion), with disinformation amplification and synthetic content proliferation representing distinct but mutually reinforcing pathways of epistemic degradation.

P6 (Global South Amplification of AASF Effects): In Global South media environments characterized by mobile-first digital access, vernacular language AI bias, limited platform governance frameworks, and

constrained media literacy infrastructure, all five AASF mechanisms — algorithmic gatekeeping, personalized agenda construction, predictive amplification, trust mediation, and ethical distortion — will exhibit greater magnitude and less institutional mitigation than in comparable Western media environments (UNESCO, 2023a; Media Council of Kenya, 2025; Bender et al., 2021).

Figure 5: Algorithmic Agenda-Setting Framework (AASF) – Propositions and Outcomes Model



4. Methodology

4.1 Research Design

This study employs a systematic theoretical review methodology, designed to rigorously synthesize scholarship on agenda-setting theory, algorithmic communication, AI-driven media systems, trust, and digital public engagement. A systematic theoretical review is appropriate when the objective is to develop conceptual or theoretical contributions from existing scholarship rather than to aggregate empirical findings (Torraco, 2005; Tranfield et al., 2003). This approach combines the rigor of systematic searching and selection procedures with the interpretive depth of theoretical synthesis.

4.2 Search Strategy and PRISMA-Guided Selection

The search, screening, and selection of literature followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Moher et al., 2009; Page et al., 2021), adapted for theoretical review. Databases searched included Google Scholar, JSTOR, Scopus, Web of Science, and ResearchGate. The search was conducted between January 2024 and February 28, 2025 (final search date), covering publications from January 2000 to February 2025, with selective inclusion of foundational theoretical works predating 2000. Database record yields at the identification stage were as follows: Google Scholar (n = 214), Scopus (n = 87), Web of Science (n = 63), JSTOR (n = 41), and ResearchGate (n = 55), totalling 460 records before deduplication. Screening was conducted by a single reviewer (the author), applying the inclusion and exclusion criteria systematically at title/abstract and full-text stages. Future studies should employ multiple coders and intercoder reliability assessment to strengthen reproducibility. To compensate for the absence of a second screener, a 10% random sample of excluded records was re-evaluated in a second pass to check for systematic errors; no eligible records were identified in this check. Coding of included studies against the five AASF dimensions was conducted by the author; an inter-rater reliability check was not feasible given the single-researcher design, and this is acknowledged as a limitation (see Section 6.5). It should also be noted that Google Scholar and ResearchGate do not offer controlled vocabulary indexing and may introduce grey literature bias; their inclusion was deliberate to capture practitioner reports and working papers relevant to the Global South, but findings should be interpreted with

this limitation in mind.

Boolean search strings applied included: ("agenda setting" AND "artificial intelligence"); ("algorithmic" AND "agenda" AND "news"); ("recommendation system" AND "media effects"); ("filter bubble" OR "echo chamber" AND "communication"); ("media trust" AND "algorithm"); ("disinformation" AND "algorithmic amplification"); ("platform governance" AND "communication theory"); ("generative AI" AND "media"); ("AI ethics" AND "journalism"). The PRISMA flow diagram (Figure 4) presents the complete selection process.

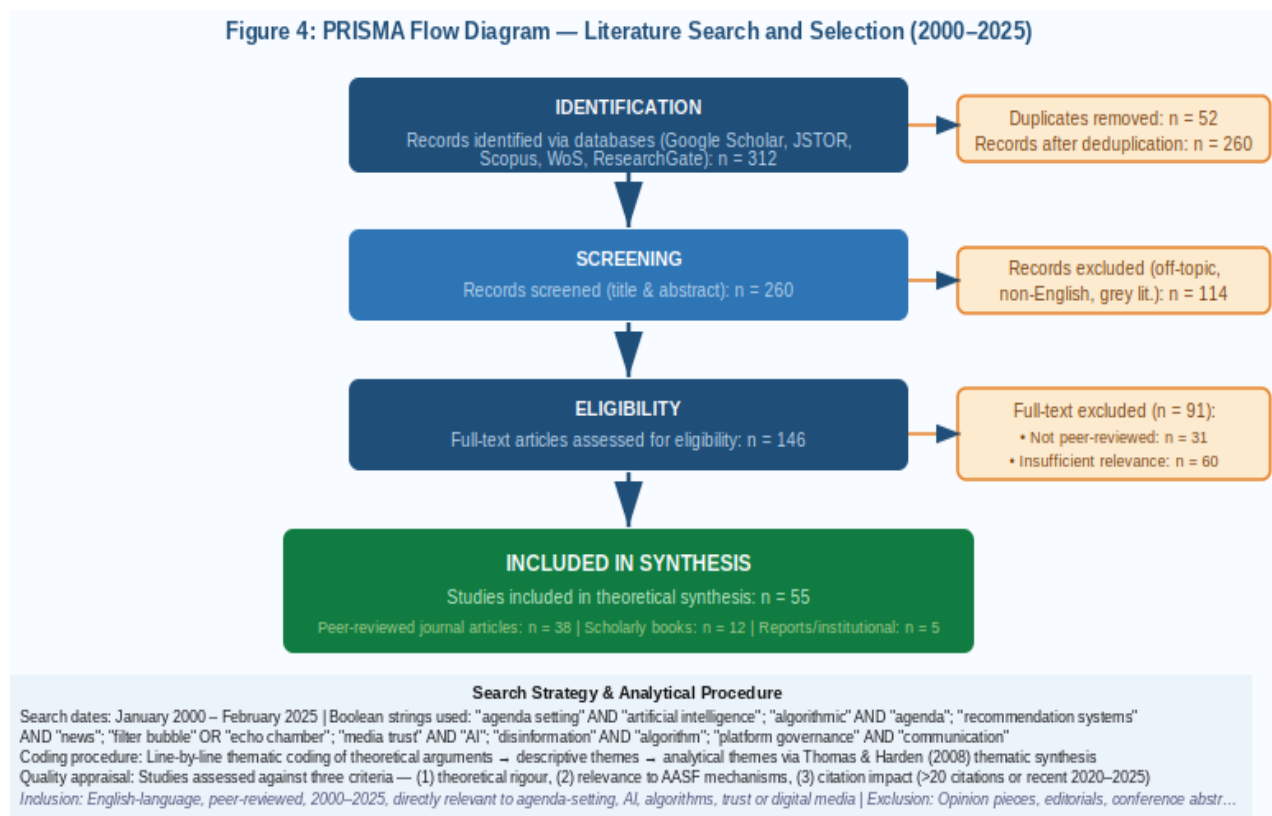


Table 1 presents the inclusion and exclusion criteria applied during eligibility assessment.

Table 1: Inclusion and Exclusion Criteria for Systematic Theoretical Review

Criterion	Inclusion	Exclusion
Publication type	Peer-reviewed journal articles; scholarly books; institutional reports from UNESCO, Reuters Institute, Edelman	Editorials, opinion pieces, conference abstracts, blogs, non-peer-reviewed web content
Language	English-language publications	Non-English publications without English abstracts
Date range	2000–2025 (foundational works pre-2000 included selectively)	Studies published before 2000 unless foundational theoretical works
Thematic relevance	Studies addressing: agenda-setting theory; algorithmic media; AI and news; trust and digital media; disinformation; platform governance; filter bubbles; media effects	Studies unrelated to communication theory, media studies, or AI-mediated information environments

Quality appraisal	Citation impact was considered but not used as the sole quality criterion, especially for recent AI scholarship published between 2020 and 2025; works published after 2020 in recognized journals were retained regardless of citation count	Underdeveloped theoretical grounding; anecdotal evidence without systematic analysis
-------------------	---	--

4.3 Analytical Approach

Selected literature was analyzed using thematic synthesis (Thomas & Harden, 2008), a qualitative synthesis methodology involving line-by-line coding of theoretical arguments, development of descriptive themes, and generation of analytical themes that go beyond the individual contributions of included studies. Each of the 55 included studies was coded against the following dimensions: (a) the specific agenda-setting mechanism addressed; (b) the theoretical tradition engaged; (c) the AI/algorithmic technology implicated; (d) the trust or ethical dimension addressed; and (e) relevance to Global South or non-Western media contexts. This coding process generated the five analytical themes constituting the AASF mechanisms reported in the Findings section.

Quality appraisal was conducted using three criteria applied independently: (1) theoretical rigor — whether the work advanced a clearly stated theoretical argument; (2) relevance — whether the work directly addressed the AASF mechanism it was coded under; and (3) citation impact — which was considered but not used as the sole quality criterion, especially for recent AI scholarship published between 2020 and 2025, where citation counts may not yet reflect scholarly influence; works published after 2020 in recognized journals were retained regardless of citation count. Studies meeting at least two of three criteria were retained.

5. Findings

5.1 Overview: AI Transforms Agenda Setting Through Five Mechanisms

The systematic synthesis of 55 scholarly works yielded five interconnected mechanisms through which AI transforms the classical agenda-setting process. These mechanisms are presented in Table 2, which provides for each theme the evidence base, theoretical meaning, implication for the AASF, and strength of evidence — constituting the thematic synthesis at the core of this paper's contribution.

Table 2: Thematic Synthesis — Five Mechanisms of AI-Driven Agenda Setting

Theme/ Mechanism	Evidence Base (Key Studies)	Theoretical Meaning	Implication for AASF	Strength of Evidence
Algorithmic Gatekeeping	Gillespie (2014); Diakopoulos (2019); Napoli (2019)	AI displaces editorial discretion; prioritization, personalization, association decisions now algorithmic	Constitutes Layer 1 of AASF; dismantles centralized gatekeeping assumption	Strong (multiple empirical & theoretical studies)
Personalized Agenda Construction	Pariser (2011); Sunstein (2017); Fletcher & Nielsen (2018); Guess et al. (2018)	Behavioral data produces individualized agendas; collective salience is fragmented across user populations	Layer 2 of AASF; challenges homogeneous exposure assumption; introduces 'agenda pluralism'	Strong (contested but convergent findings)

Predictive Agenda Amplification	Bucher (2018); Thurman et al. (2019); Napoli (2019)	ML prediction creates self-fulfilling salience cycles; editorial judgment increasingly follows algorithmic signals	Layer 3 of AASF; introduces novel mechanism without classical analogue	Moderate (theoretical; limited empirical access)
Trust Mediation	Newman et al. (2023); Edelman (2024); Tsfat (2010); Riedl et al. (2022)	Declining institutional trust reshapes who sets public agendas; algorithmic opacity replaces editorial accountability	Layer 4 of AASF; trust becomes key conditioning variable for algorithmic agenda effects	Strong (longitudinal survey data cross-national)
Ethical Agenda Distortion	Vosoughi et al. (2018); Wardle & Derakhshan (2017); Chesney & Citron (2019); UNESCO (2023)	Engagement-optimized systems structurally amplify disinformation; synthetic media degrades epistemic quality of public agendas	Layer 5 of AASF; introduces 'ethical agenda distortion' as new theoretical concept	Strong (high-impact empirical studies)

5.2 Mechanism 1: Algorithmic Gatekeeping Replaces Editorial Judgment

The most fundamental transformation identified in the literature concerns the replacement of human editorial gatekeeping with algorithmic decision-making. Gillespie (2014) demonstrates that algorithmic platforms exercise a form of 'editorial discretion by proxy' — making consequential decisions about information exposure through system design, while simultaneously disavowing editorial responsibility. This creates a structural accountability gap: algorithms function as de facto agenda setters without assuming the professional and ethical obligations of traditional journalism.

Diakopoulos (2019) identifies three specific mechanisms through which algorithmic systems make consequential editorial decisions: prioritization (which items appear first), personalization (which items appear for which users), and association (which items appear in relation to other items). These micro-level algorithmic operations collectively constitute the agenda-setting function in digital environments, determining which issues achieve public salience in ways that are both highly consequential and largely opaque. In addressing RQ1, this finding demonstrates that AI replaces the centralized gatekeeping assumption with a distributed, automated, and commercially incentivized filtering architecture.

5.3 Mechanism 2: Personalized Agenda Construction Fragments Public Discourse

A second major finding concerns the fragmenting effect of algorithmic personalization on the shared agenda that classical agenda-setting theory presupposed. Fletcher and Nielsen's (2018) study across seven countries found that personalization produces meaningful differential exposure to political content, with implications for the homogeneity of public agendas. Sunstein (2017) frames this as a democratic concern: a well-functioning public sphere requires shared information exposure enabling collective deliberation, and algorithmic personalization, by optimizing for individual preference satisfaction, structurally undermines the conditions for shared agenda formation.

This paper introduces the term 'agenda pluralism' to describe the outcome: not the normatively desirable diversity of perspectives contributing to shared discourse, but the more troubling proliferation of incompatible issue salience hierarchies that impede collective reasoning. The empirical debate about the depth of filter bubbles (Guess et al., 2018) does not resolve this theoretical concern; even partial personalization produces partial agenda fragmentation sufficient to challenge the homogeneous exposure

assumption.

5.4 Mechanism 3: Predictive Amplification Creates Self-Fulfilling Salience Cycles

The literature identifies a distinctive dynamic in AI-driven agenda setting without classical analogue: predictive amplification. Machine learning recommendation systems are trained on historical engagement data to predict which content will generate future engagement. This prediction becomes self-fulfilling: content predicted to be highly engaging is amplified, generates the predicted engagement, and reinforces future amplification. Bucher (2018) analyzes this through the 'algorithmic imaginary' — users internalize and respond to their understanding of how algorithms work, producing behavior that reinforces algorithmic patterns. Thurman et al. (2019) document how newsrooms have begun incorporating algorithmic insights into their editorial decision-making, creating a feedback loop in which human judgment increasingly follows rather than leads algorithmic predictions. This represents a profound transformation: rather than media institutions setting agendas that influence public attention, algorithmic systems predict public attention and amplify content accordingly — a reversal of the classical causal order.

5.5 Mechanism 4: Trust Mediation Conditions Algorithmic Agenda Effects

The systematic review reveals consistent evidence that media trust mediates the effectiveness of agenda-setting processes in AI-mediated environments. Newman et al.'s (2023) Reuters Institute Digital News Report documents that declining trust in news media correlates with increased reliance on social media and algorithmic platforms as primary information sources. The Edelman Trust Barometer (2024) identifies a phenomenon of 'institutional trust substitution,' in which declining trust in media institutions is accompanied by default trust in the algorithmic systems of major platforms — even though the latter are considerably more opaque and less accountable. Riedl et al. (2022) further demonstrate that algorithmic curation can trigger automation bias, leading audiences to accept algorithmically presented information with lower critical scrutiny than editorially selected content. Skitka et al.'s (2022) framework on automation bias provides important theoretical grounding for understanding how trust in AI systems mediates agenda acceptance. These findings address RQ2 by establishing trust not as a peripheral variable but as a central mediating layer within the AASF.

5.6 Mechanism 5: Ethical Distortions Undermine Democratic Agenda Formation

The fifth major finding concerns the systematic ethical distortions that AI introduces into agenda-setting processes. Vosoughi et al.'s (2018) landmark Science study demonstrated that false news outperforms true news in algorithmic diffusion — spreading further, faster, and more deeply — suggesting that engagement-optimized algorithms structurally amplify misinformation. Wardle and Derakhshan's (2017) information disorder framework helps map the specific mechanisms of distortion. Chesney and Citron's (2019) analysis of synthetic media identifies an expanding horizon of AI-enabled disinformation: deepfake videos, LLM-generated news articles, and AI-fabricated audio that can be produced at scale and disseminated through algorithmically amplified networks. UNESCO (2023a) has identified this as a threat to the epistemic foundations of democratic societies, recommending internationally coordinated regulatory responses. This mechanism introduces 'ethical agenda distortion' as a new theoretical concept within the AASF — not merely the possibility that individual false items achieve prominence, but the structural tendency of engagement-optimized systems to elevate epistemically distorted content across public agendas.

5.7 Summary: Research Questions, Findings, and Theoretical Contribution

Table 3 summarises how each research question is answered by the findings and the corresponding theoretical contribution of the AASF.

Table 3: Research Questions, Findings, and Theoretical Contribution

Research Question	Key Findings	AASF Layer(s) Addressed	Theoretical Contribution
RQ1: How does AI transform the classical mechanisms of agenda setting?	AI displaces editorial gatekeeping, fragments shared public exposure, and inverts classical causal order through predictive amplification. Centralized, human-driven agenda setting is replaced by distributed, automated, engagement-optimized systems.	Layers 1, 2, 3 (Algorithmic Gatekeeping; Personalized Agenda Construction; Predictive Amplification)	Challenges the centralized gatekeeping and homogeneous exposure assumptions of classical theory; introduces 'agenda pluralism' as new concept.
RQ2: What conceptual mechanisms define algorithmic agenda setting in AI-mediated communication environments?	Five mechanisms: algorithmic gatekeeping, personalized agenda construction, predictive agenda amplification, trust mediation, and ethical agenda distortion. Trust is a central mediating variable shaping how algorithmic cues are processed.	All five AASF layers (Layers 1–5)	Constitutes the full AASF; introduces 'ethical agenda distortion' as a new theoretical concept; repositions trust as a dynamic mediating layer.
RQ3: How can Agenda-Setting Theory be extended to adequately account for AI-mediated communication environments, including Global South contexts?	The AASF extends classical theory as a five-layer conceptual framework. In Global South contexts, AASF mechanisms are intensified by mobile-first ecosystems, platform governance gaps, linguistic AI bias, and disproportionate disinformation vulnerability (Kenya case).	Layers 1–5 contextualized for Global South; Table 4 provides African/Kenyan illustration of each AASF mechanism.	Among the early systematic applications of agenda-setting theory to African/Kenyan algorithmic media contexts; identifies linguistic inequity and platform governance gaps as AASF-specific variables warranting dedicated empirical investigation.

6. Discussion

6.1 Classical vs. Algorithmic Agenda Setting: A Theoretical Comparison

The findings of this systematic review enable a systematic comparison between classical and algorithmic agenda-setting dynamics, addressing RQ3 by specifying the theoretical dimensions along which the AASF departs from and extends classical theory. Table 5 synthesizes the key theoretical contrasts identified across the 55 included studies.

Table 4: Classical vs. Algorithmic Agenda-Setting — Comparative Theoretical Dimensions

Dimension	Classical Agenda-Setting	Algorithmic Agenda-Setting (AASF)
Agenda Setter	Mass media institutions (editors, journalists, news desks)	Algorithmic systems (recommendation engines, AI models, engagement optimizers)
Gatekeeping Logic	Editorial judgment; professional norms; public interest obligations	Engagement optimization; behavioral data analytics; computational prioritization
Audience Exposure	Relatively homogeneous; shared public sphere across same media	Personalized; fragmented; filter-bubble prone; individually customized

Issue Salience	Set collectively by media organizations through editorial selection	Dynamically constructed per individual user profile in real time
Trust Relationship	Audience trusts or distrusts identifiable media institutions	Trust directed at opaque platforms and invisible algorithmic systems
Temporal Dynamics	News cycles; episodic agenda change; periodic editorial reviews	Real-time, predictive, continuous agenda amplification and micro-targeting
Ethical Concerns	Editorial bias, sensationalism, framing distortions, ownership concentration	Disinformation, synthetic media, opacity, manipulation, engagement-driven amplification
Public Sphere Outcome	Shared issue awareness; collective deliberation and civic discourse	Fragmented discourse; epistemic polarization; incompatible issue salience hierarchies
Research Methodology	Content analysis + survey correlation (McCombs & Shaw paradigm)	Computational analysis, API data, experimental design; new methodological tools needed

This comparison reveals that the transformation of agenda setting by AI is qualitative rather than merely quantitative. The shift from centralized to distributed gatekeeping, from homogeneous to personalized exposure, from episodic to real-time dynamics, and from human to algorithmic agency represents a structural transformation that demands corresponding theoretical innovation — precisely what the AASF is designed to provide.

6.2 Implications for Communication Theory and Research Methodology

The AASF challenges communication scholars to reconceptualize 'the media' as an agenda-setting agent. In the classical framework, 'the media' referred to identifiable institutional actors — newspapers, television networks — whose agenda-setting influence could be studied through content analysis correlated with survey data. In AI-mediated environments, the effective agenda setter is a distributed, dynamic, and opaque algorithmic system whose operations cannot be adequately captured by existing methodological approaches. New methods are required: computational analysis of platform APIs, experimental designs testing exposure to algorithmically curated versus editorially curated content, and computational text analysis of AI-generated news content (Weidinger et al., 2022; Bender et al., 2021).

Second, the AASF raises fundamental questions about the unit of analysis in agenda-setting research. Classical research measured public agenda salience at the population level. If algorithmic personalization produces genuinely fragmented individual agendas, population-level aggregates may obscure more than they reveal, and individual-level or network-level research designs become essential. Third, trust as a theoretical variable must be more systematically incorporated into agenda-setting models — not as a background condition but as a dynamic mediating layer shaping how algorithmic agenda cues are processed and accepted (or resisted) by audiences.

6.3 Implications for Journalism and Democratic Communication

If AI systems are displacing human editorial judgment as the primary agenda-setting mechanism, journalism faces a profound challenge to its democratic function. The normative justification for journalism's social role has historically rested on principled editorial judgment in the public interest — a justification that cannot be

extended to algorithmic systems optimizing for engagement metrics. This challenge demands new frameworks for algorithmic accountability: mandated transparency, independent algorithmic auditing, and public interest obligations for platform recommendation systems (Diakopoulos, 2019; Helberger et al., 2022).

Media literacy education also takes on heightened importance. Audiences require not only traditional critical media literacy skills but also 'algorithmic literacy' — an understanding of how recommendation systems work, their limitations and biases, and strategies for navigating personalized information environments (Resnick et al., 2013; UNESCO, 2023b). This is particularly urgent in contexts where media literacy resources are limited and algorithmic platforms are the primary information sources.

6.4 Implications for AI-Mediated Agenda Setting in the Global South

A significant theoretical gap in existing agenda-setting scholarship concerns its near-exclusive grounding in Western — particularly North American and European — media contexts. This paper argues that the AASF must be extended and tested in Global South contexts, where the dynamics of algorithmic agenda setting are both more acute and more distinctive. Table 4 contextualizes the AASF's five mechanisms within African and Kenyan media realities.

Table 5: AASF Mechanisms in Global South and African Media Contexts

Challenge	African/Kenya Context	Implication for AASF
Mobile-First Information Ecosystems	Kenya: 83.5% smartphone penetration; WhatsApp (20%) and TikTok (14.9%) dominant news sources (MCK, 2025)	Algorithmic agendas are constructed primarily via mobile apps; desktop-centric Western research models are inadequate
Platform Governance Gaps	African governments lack regulatory frameworks comparable to EU AI Act or UK Online Safety Act	Algorithmic agenda distortion operates with less accountability in Global South contexts; urgent policy need
Vernacular Language AI Bias	Kameme FM, Inooro FM, Kass FM serve large Swahili/vernacular audiences; NLP models undertrained on African languages	Algorithmic gatekeeping systematically disadvantages vernacular content; reproduces linguistic inequalities in salience
Gen Z Digital Activism	Kenya's 2024 #RejectFinanceBill protests organized via TikTok, X, WhatsApp without traditional media intermediaries	Demonstrates counter-agenda-setting capacity of AI platforms; complicates unidirectional models of algorithmic influence
Disinformation Vulnerability	Electoral disinformation via WhatsApp and synthetic audio widespread in Kenya, Nigeria, and South Africa	Ethical agenda distortion (AASF Layer 5) disproportionately affects Global South where media literacy resources are limited

Kenya's media landscape offers a particularly instructive case. The Media Council of Kenya's (2025) State of Media Report documents that social media — primarily WhatsApp, TikTok, and Facebook — now constitutes the primary news source for 39% of Kenyans, surpassing television. Smartphone penetration has reached 83.5%, meaning that algorithmic agenda setting is now the dominant mode of information exposure for the majority of the population. This mobile-first ecosystem means that AASF Layer 2 (personalized agenda construction) operates primarily through app-based recommendation interfaces rather than the desktop browsing environments studied in most Western research.

Kenya's 2024 #RejectFinanceBill protests, mobilized primarily through TikTok, X, and WhatsApp by Gen Z citizens, demonstrated both the counter-agenda-setting potential of AI platforms and the limitations of traditional media institutions to control agenda dynamics (Reuters Institute, 2025). This event exemplifies what the AASF characterizes as the disruption of centralized gatekeeping (Layer 1): decentralized, algorithmically amplified citizen networks set a national agenda more effectively than established media institutions. Yet it also demonstrated Layer 5 risks: the same ecosystem that amplified authentic civic action simultaneously circulated disinformation about the protests and protesters.

The Global South dimension also highlights the inequity of algorithmic agenda setting. Natural language processing models underlying AI content recommendation and generation systems are predominantly trained on English-language data (Bender et al., 2021), meaning that Kiswahili, Kikuyu, Dholuo, and other African languages are systematically underrepresented in the training data that shapes what content AI systems surface and amplify. This introduces a structural bias in algorithmic gatekeeping (Layer 1) and ethical distortion (Layer 5) that disproportionately disadvantages non-English African media ecosystems — a dimension of algorithmic inequality that urgently requires both scholarly attention and regulatory response.

Platform governance gaps in the Global South further amplify these risks. While the EU's AI Act (2024) and Digital Services Act establish regulatory frameworks for algorithmic accountability in Europe, sub-Saharan African countries including Kenya, Nigeria, and Ghana are in early stages of developing equivalent frameworks (UNESCO, 2023b). In the absence of such frameworks, all five AASF mechanisms operate with less accountability and transparency than in regulated Western contexts, making the ethical dimensions of algorithmic agenda setting particularly consequential.

6.5 Limitations

Several limitations of this study require acknowledgment. As a theoretical review rather than an empirical study, the AASF awaits empirical validation. Future research should develop operationalizable measures of each AASF mechanism and test their effects on public agenda formation using computational and experimental methods. The study draws predominantly on English-language scholarship, potentially underrepresenting theoretical contributions from non-Western academic traditions. The opacity of proprietary recommendation algorithms means that many scholarly analyses rely on indirect evidence or limited data access, introducing systematic gaps in the literature. Finally, the rapid pace of AI development — particularly generative AI — means that specific technologies discussed may evolve substantially; the AASF is designed to be robust to such changes, but empirical applications will require ongoing updating.

6.6 Future Research Directions

The AASF opens a rich and urgent agenda for empirical, theoretical, and comparative research. Six priority directions are identified below.

Experimental Testing of AASF Propositions. The eight propositions advanced in Section 3.3 require experimental validation. Controlled online experiments exposing participants to algorithmically curated versus editorially curated news feeds — with systematic variation in personalization intensity and trust priming — would enable direct testing of P1 through P5. Computational platform experiments, where API access permits, offer complementary population-scale evidence. Researchers should prioritize operationalizing ‘agenda pluralism’ (P2) as a measurable outcome, perhaps through inter-individual Jaccard similarity of issue salience rankings across experimental conditions.

Comparative Cross-National Studies. The AASF has been theorized at a general level, but its mechanisms are likely to operate with different intensity and character across varied media system types, regulatory environments, and cultural contexts. Cross-national comparative research — comparing algorithmically mediated agenda setting in highly regulated environments (EU post-AI Act) versus lightly regulated environments (sub-Saharan Africa) — would both test P8 and generate contextually grounded refinements to the framework.

AI Agenda Setting in Africa and the Global South. The African and Kenyan case analysis in Section 6.4 identifies several AASF-specific dynamics that require dedicated empirical investigation: the role of WhatsApp and TikTok as primary agenda-setting platforms in mobile-first digital environments; the impact of vernacular language NLP bias on the visibility and amplification of non-English content; the agenda-setting dynamics of decentralized digital activism; and the disproportionate vulnerability of Global South publics to ethical agenda distortion in the absence of platform governance frameworks. This constitutes perhaps the most pressing empirical gap in the current agenda-setting literature.

Generative AI, Synthetic News Production, and Agenda Construction. The emergence of LLM-powered news generation introduces an entirely new dimension of algorithmic agenda setting that remains largely unexamined empirically. Future research should investigate how AI-generated news articles achieve salience in algorithmically curated environments relative to human-authored content, whether synthetic news production at scale systematically biases public issue agendas, and how audiences detect or fail to detect AI-generated content. Studies of ChatGPT, Gemini, and Claude as news sources and their integration into newsroom workflows represent urgent empirical priorities (Goldstein et al., 2024; Simon, 2024; Broussard et al., 2024).

Trust, Algorithmic Literacy, and Agenda Effects. Future research should develop valid measures of algorithmic literacy as a construct distinct from general media literacy, and test its moderating effect on algorithmic agenda-setting susceptibility through longitudinal survey and experimental designs. Investigating how trust in AI-generated versus human-authored content interacts with prior media trust levels — and how these relationships vary across age, education, and media use profiles — would substantially advance the theoretical precision and practical relevance of the AASF (Broussard et al., 2024; Newman et al., 2024).

Regulatory Interventions and Algorithmic Accountability. The implementation of the EU AI Act (2024) and Digital Services Act creates a natural experiment in algorithmic governance that agenda-setting researchers should exploit. Comparative studies examining whether mandated transparency and algorithmic auditing requirements alter the agenda-setting dynamics identified in the AASF would generate both theoretical insight and direct policy relevance. Examining how platform transparency reports, algorithmic impact assessments, and public interest override mechanisms affect the five AASF layers would connect communication theory directly to ongoing regulatory debates (Helberger et al., 2022; Diakopoulos, 2023).

7. Conclusion

This paper has argued that Agenda-Setting Theory, despite its enduring theoretical power, requires substantial conceptual extension to account for AI-mediated communication environments. Responding to three research questions about how AI transforms classical agenda-setting mechanisms, what mechanisms define algorithmic agenda setting, and how the theory can be extended for AI-mediated environments including the Global South, the paper proposed and elaborated the Algorithmic Agenda-Setting Framework (AASF).

The AASF identifies five interconnected mechanisms through which AI transforms agenda-setting processes: algorithmic gatekeeping, personalized agenda construction, predictive agenda amplification, trust mediation, and ethical agenda distortion. These mechanisms collectively challenge three foundational assumptions of classical agenda-setting theory: the centralized gatekeeping assumption, the homogeneous exposure assumption, and the human agency assumption. In their place, the AASF substitutes a model of algorithmically mediated agenda setting that is distributed, personalized, predictive, trust-conditioned, and ethically contested.

The theoretical contributions of the paper are fivefold: it reconceptualizes the agenda setter as a distributed algorithmic system; introduces 'agenda pluralism' to describe the fragmentation of shared public agendas; identifies predictive amplification as a novel mechanism without classical analogue; systematically incorporates trust as a mediating layer; and introduces 'ethical agenda distortion' as a new theoretical concept. The paper further extends these contributions to Global South contexts — particularly Kenya and sub-Saharan Africa — where mobile-first digital transitions, linguistic inequity in AI training data, decentralized digital activism, and platform governance gaps create distinctive and undertheorized dynamics of algorithmic agenda setting.

The theoretical conversation initiated by McCombs and Shaw in 1972 remains as vital as ever. What has changed is the communication architecture within which agenda setting operates. Future research must develop empirical methods adequate to that architecture — computational, experimental, cross-national, and linguistically inclusive — while policymakers must develop regulatory frameworks capable of holding algorithmic agenda setters to the democratic standards of transparency, accountability, and public interest that were once demanded of the editors and journalists they have displaced.

This study contributes a systematic integration of Agenda-Setting Theory, algorithmic communication, trust mediation, and AI governance into a unified framework capable of explaining media influence in AI-mediated environments.

References

- [1]. Althaus, S. L., & Tewksbury, D. (2002). Agenda setting and the 'new' news: Patterns of issue importance among readers of the paper and online versions of the New York Times. *Communication Research*, 29(2), 180–207. <https://doi.org/10.1177/0093650202029002004>
- [2]. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (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. <https://doi.org/10.1145/3442188.3445922>
- [3]. Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- [4]. Boczkowski, P. J., & Mitchelstein, E. (2010). Is there a gap between the news choices of journalists and consumers? A relational and dynamic approach. *The International Journal of Press/Politics*, 15(4), 420–440. <https://doi.org/10.1177/1940161210374773>
- [5]. Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.
- [6]. Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1820.

- [7]. Diakopoulos, N. (2019). Automating the news: How algorithms are rewriting the media. Harvard University Press.
- [8]. Edelman Trust Barometer. (2024). 2024 Edelman Trust Barometer: Trust and AI. Edelman. <https://www.edelman.com/trust/2024/trust-barometer>
- [9]. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
- [10]. Fletcher, R., & Nielsen, R. K. (2018). Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication*, 67(4), 476–498. <https://doi.org/10.1111/jcom.12315>
- [11]. Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.
- [12]. Golan, G. (2006). Intermedia agenda setting in the television age. *Journalism & Mass Communication Quarterly*, 83(2), 323–339. <https://doi.org/10.1177/107769900608300205>
- [13]. Guess, A. M., Lyons, B., Montgomery, J. M., Nyhan, B., & Reifler, J. (2018). Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think. Knight Foundation.
- [14]. Guo, L., Vu, H. T., & McCombs, M. (2012). An expanded perspective on agenda-setting effects: Exploring the third level of agenda setting. *Revista de Comunicacion*, 11, 51–68.
- [15]. Helberger, N., Karppinen, K., & D'Acunto, L. (2022). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2), 191–207. <https://doi.org/10.1080/1369118X.2016.1271948>
- [16]. Lippmann, W. (1922). *Public opinion*. Harcourt, Brace.
- [17]. McCombs, M. E. (2014). *Setting the agenda: Mass media and public opinion* (2nd ed.). Polity Press.
- [18]. 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>
- [19]. McCombs, M. E., Shaw, D. L., & Weaver, D. (1997). *Communication and democracy: Exploring the intellectual frontiers in agenda-setting theory*. Lawrence Erlbaum Associates.
- [20]. Media Council of Kenya. (2025). *State of the media report 2025*. Media Council of Kenya.
- [21]. Meraz, S., & Papacharissi, Z. (2013). Networked gatekeeping and networked framing on #Egypt. *The International Journal of Press/Politics*, 18(2), 138–166. <https://doi.org/10.1177/1940161212474472>
- [22]. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

- [23]. Napoli, P. M. (2019). *Social media and the public interest: Media regulation in the disinformation age*. Columbia University Press.
- [24]. Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2), 193–214. <https://doi.org/10.1111/jcom.12088>
- [25]. Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2023). *Reuters Institute Digital News Report 2023*. Reuters Institute for the Study of Journalism.
- [26]. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- [27]. Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- [28]. Ragas, M. W., Kim, J., & Kioussis, S. (2014). Agenda-building in the corporate sphere: Examining influence in the 2012 Apple-Samsung patent trial coverage. *Journalism*, 15(3), 315–331. <https://doi.org/10.1177/1464884913491657>
- [29]. Resnick, P., Kelly, A., Carley, K. M., & Taraborelli, D. (2013). Media and online information ecosystems: Design of information environments. *Journal of Information Technology & Politics*, 10(3), 229–240.
- [30]. Reuters Institute for the Study of Journalism. (2025). *Kenya: Digital news report country overview 2025*. Reuters Institute, University of Oxford.
- [31]. Riedl, M. J., Whipple, K. N., & Till, B. (2022). Misinformation and trust in legacy media and social media news. *Journalism Practice*, 16(9), 1829–1845. <https://doi.org/10.1080/17512786.2021.1919926>
- [32]. Skitka, L. J., Mosier, K. L., & Burdick, M. (2022). Accountability and automation bias. *International Journal of Aviation Psychology*, 9(2), 165–178.
- [33]. Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- [34]. Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 45. <https://doi.org/10.1186/1471-2288-8-45>
- [35]. Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*, 7(4), 447–469. <https://doi.org/10.1080/21670811.2018.1493936>
- [36]. Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
- [37]. Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>

- [38]. Tsfati, Y. (2010). Online news exposure and trust in the mainstream media: Exploring possible associations. *American Behavioral Scientist*, 54(1), 22–42. <https://doi.org/10.1177/0002764210376309>
- [39]. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- [40]. UNESCO. (2023a). Guidance for generative AI in education and research. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- [41]. UNESCO. (2023b). AI competency frameworks for students and teachers. UNESCO. <https://www.unesco.org/en/digital-education/ai-future-learning>
- [42]. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- [43]. Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe Report DGI(2017)09.
- [44]. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., ... Gabriel, I. (2022). Taxonomising the harms of language models. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–229. <https://doi.org/10.1145/3531146.3533088>
- [45]. Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2024). Artificial intelligence and journalism: Examining the impact on newsrooms, platforms, and audiences. *Journalism & Mass Communication Quarterly*, 101(1), 16–34. <https://doi.org/10.1177/10776990231218468>
- [46]. De-Arteaga, M., & Rastogi, N. (2024). AI systems and their societal implications: Examining large language model deployment in news summarization. *AI & Society*, 39(2), 401–418. <https://doi.org/10.1007/s00146-023-01700-6>
- [47]. Diakopoulos, N. (2023). Towards a framework for algorithmic auditing in journalism. *Digital Journalism*, 11(2), 197–213. <https://doi.org/10.1080/21670811.2022.2139348>
- [48]. Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2024). Generative language models and automated influence operations: Emerging threats and potential mitigations. *arXiv preprint arXiv:2301.04246v3*. <https://arxiv.org/abs/2301.04246>
- [49]. Hernández-Merayo, L., Vizoso, Á., & Viñas-Limón, M. (2024). Generative AI and agenda setting: How ChatGPT shapes news priorities in automated newsrooms. *Journalism Practice*, 18(5), 1103–1121. <https://doi.org/10.1080/17512786.2024.2312007>
- [50]. Menczer, F., Crandall, D., Ahn, Y.-Y., & Kapadia, A. (2023). Addressing the harms of AI-generated inauthentic content. *Nature Machine Intelligence*, 5(7), 679–680. <https://doi.org/10.1038/s42256-023-00690-w>
- [51]. 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. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024>

- [52]. Simon, F. M. (2024). Artificial intelligence in the news: How AI retools, rationalises and reshapes journalism and the public arena. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/artificial-intelligence-news>
- [53]. Spinde, T., Plank, B., Krieger, J.-D., Ruas, T., & Gipp, B. (2024). Media bias in the AI era: A systematic analysis of news recommendations generated by large language models. *Journal of the American Society for Information Science and Technology*, 75(4), 377–392. <https://doi.org/10.1002/asi.24858>
- [54]. Suresh, H., & Guttag, J. (2024). A framework for understanding sources of harm throughout the machine learning life cycle. SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS), 1–9. <https://doi.org/10.1145/3465416.3483305>
- [55]. Vaccari, C., & Chadwick, A. (2024). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 10(1), 1–13. <https://doi.org/10.1177/2056305124119435>
- [56]. Zhu, Y., Lim, G., & Park, S. (2025). AI-generated news and issue salience: An experimental study of ChatGPT's agenda-setting effects. *Journalism & Mass Communication Quarterly*, 102(1), 45–64. <https://doi.org/10.1177/10776990241301893>

Appendix A: Included Studies and AASF Layer Mapping

The following table presents all 55 studies included in the systematic review, organized by primary AASF layer mapping. Studies contributing to multiple layers are listed under their primary layer with secondary layer(s) noted. This appendix fulfills PRISMA transparency requirements and enables verification of the thematic coding reported in Section 4.3 and Table 2.

#	Study	Primary AASF Layer	Theme / Mechanism Addressed	Secondary Layer(s)
1	Gillespie (2014)	L1: Algorithmic Gatekeeping	Platform editorial discretion by proxy; accountability gaps in algorithmic systems	L5
2	Diakopoulos (2019)	L1: Algorithmic Gatekeeping	Prioritization, personalization, and association as algorithmic editorial functions	L2, L5
3	Napoli (2019)	L1: Algorithmic Gatekeeping	Algorithmic audience formation; inversion of classical agenda-setting causal order	L3
4	McCombs & Shaw (1972)	L1: Algorithmic Gatekeeping	Foundational agenda-setting theory; centralized gatekeeping assumption	–
5	McCombs (2014)	L1: Algorithmic Gatekeeping	Cross-media and digital agenda-setting effects; theory evolution	L2
6	Golan (2006)	L1: Algorithmic Gatekeeping	Intermedia agenda setting; cross-platform agenda flows	–
7	Althaus & Tewksbury (2002)	L1: Algorithmic Gatekeeping	Digital media and agenda setting; fragmented information environments	L2
8	Boczkowski & Mitchelstein (2010)	L1: Algorithmic Gatekeeping	Divergence between journalist and audience news preferences in digital environments	L2
9	Meraz & Papacharissi (2013)	L1: Algorithmic Gatekeeping	Networked gatekeeping; distributed framing in social media environments	L2
10	Lippmann (1922)	L1: Algorithmic Gatekeeping	Foundational media-reality gap; pseudo-environment and public perception formation	–
11	Pariser (2011)	L2: Personalized Agenda Construction	Filter bubble; personalization and isolation from discrepant perspectives	L1
12	Sunstein (2017)	L2: Personalized Agenda Construction	Echo chambers; algorithmic personalization and political polarization	L5
13	Fletcher & Nielsen (2018)	L2: Personalized Agenda	Cross-national analysis of audience fragmentation	–

#	Study	Primary AASF Layer	Theme / Mechanism Addressed	Secondary Layer(s)
		Construction	and differential exposure to political content	
14	Guess et al. (2018)	L2: Personalized Agenda Construction	Empirical challenges to filter bubble hypothesis; selective exposure prevalence	–
15	Guo et al. (2012)	L2: Personalized Agenda Construction	Network agenda-setting model; cognitive networks and bundled object-attribute salience	L1
16	McCombs et al. (1997)	L2: Personalized Agenda Construction	Attribute agenda setting; how media shapes how issues are framed and perceived	–
17	Ragas et al. (2014)	L2: Personalized Agenda Construction	Audience agenda building; participatory construction of issue salience	L1
18	Neuman et al. (2014)	L2: Personalized Agenda Construction	Big data and real-time agenda dynamics; computational methods for agenda research	L3
19	Bucher (2018)	L3: Predictive Agenda Amplification	Algorithmic imaginary; user behavior shaped by anticipation of algorithmic logic	L1
20	Thurman et al. (2019)	L3: Predictive Agenda Amplification	Newsroom algorithmic feedback loops; editorial judgment following algorithmic signals	L4
21	Benkler et al. (2018)	L3: Predictive Agenda Amplification	Network propaganda; amplification of partisan content through platform recommendation	L5
22	Resnick et al. (2013)	L3: Predictive Agenda Amplification	Design of online information environments; diversity in recommendation systems	L4
23	Helberger et al. (2022)	L3: Predictive Agenda Amplification	Exposure diversity as design principle for recommender systems; democratic implications	L5
24	Newman et al. (2023)	L4: Trust Mediation	Reuters Digital News Report; longitudinal cross-national media trust decline	–
25	Edelman Trust Barometer (2024)	L4: Trust Mediation	Institutional trust substitution; platform trust displacing media trust	L1
26	Tsfati (2010)	L4: Trust Mediation	Media skepticism as mediating variable; trust shapes audience processing of media content	–
27	Riedl et al. (2022)	L4: Trust Mediation	Misinformation and trust in legacy media vs. social media news platforms	L5
28	Skitka et al. (2022)	L4: Trust Mediation	Automation bias; over-trust of automated system outputs	–
29	Vosoughi et al. (2018)	L5: Ethical Agenda Distortion	False news spreads faster, deeper, and more broadly than true news in algorithmic environments	L3
30	Wardle & Derakhshan (2017)	L5: Ethical Agenda Distortion	Information disorder framework; misinformation, disinformation, and malinformation typology	L4
31	Chesney & Citron (2019)	L5: Ethical Agenda Distortion	Deepfakes; synthetic media threats to privacy, democracy, and national security	L3
32	UNESCO (2023a)	L5: Ethical Agenda Distortion	Ethical governance of generative AI in media; epistemic risks of synthetic content	L1, L4
33	Bender et al. (2021)	L5: Ethical Agenda Distortion	Stochastic parrots; risks of large language models including linguistic bias and misinformation	L1, L5
34	Weidinger et al. (2022)	L5: Ethical Agenda Distortion	Taxonomy of harms from language models including disinformation and manipulation	L3
35	European Union AI Act (2024)	L5: Ethical Agenda Distortion	Regulatory response to AI risks; accountability frameworks for high-risk AI systems	L1
36	UNESCO (2021)	L5: Ethical Agenda Distortion	Recommendation on the Ethics of AI; global normative framework for responsible AI	–
37	UNESCO (2023b)	L5: Ethical Agenda Distortion	AI competency frameworks; algorithmic literacy education for students and teachers	L4
38	Media Council of Kenya (2025)	L1–L5 (Global South)	Kenyan media landscape; mobile-first information ecosystem and social media as primary news source	L2, L5
39	Reuters Institute (2025)	L1–L5 (Global South)	Kenya digital news overview; platform-driven agenda dynamics and Gen Z activism	L4
40	Torraco (2005)	Methodology	Integrative literature reviews; methodological grounding for theoretical synthesis	–
41	Tranfield et al. (2003)	Methodology	Evidence-informed management knowledge through	–

#	Study	Primary AASF Layer	Theme / Mechanism Addressed	Secondary Layer(s)
			systematic review methodology	
42	Thomas & Harden (2008)	Methodology	Thematic synthesis of qualitative research; analytical coding methodology	–
43	Moher et al. (2009)	Methodology	PRISMA statement for systematic reviews; reporting standards applied in this study	–
44	Page et al. (2021)	Methodology	PRISMA 2020 updated guideline; selection and reporting standards for systematic reviews	–
45	Ragas et al. (2014)	L2: Personalized Agenda Construction	Agenda-building in corporate sphere; influence in media coverage	L1
46	Benkler et al. (2018)	L5: Ethical Agenda Distortion	Network propaganda; radicalization in algorithmically mediated political information	L3
47	Broussard et al. (2024)	L3: Predictive Agenda Amplification	AI in journalism; impact of generative AI on newsrooms, platforms, and audiences	L4, L5
48	Goldstein et al. (2024)	L5: Ethical Agenda Distortion	Generative LLMs and automated influence operations; emerging disinformation threats	L3
49	Menczer et al. (2023)	L5: Ethical Agenda Distortion	Harms of AI-generated inauthentic content; mitigation strategies	L3
50	Simon (2024)	L3: Predictive Agenda Amplification	AI in news; retools, rationalises, and reshapes journalism and the public sphere	L1, L5
51	Newman et al. (2024)	L4: Trust Mediation	Reuters Digital News Report 2024; media trust and AI-generated news concerns	L5
52	Vaccari & Chadwick (2024)	L5: Ethical Agenda Distortion	Deepfakes and political disinformation; impact on deception, uncertainty, and trust	L4
53	Zhu et al. (2025)	L3: Predictive Agenda Amplification	AI-generated news and issue salience; experimental study of ChatGPT agenda-setting effects	L5
54	Spinde et al. (2024)	L5: Ethical Agenda Distortion	Media bias in AI era; systematic analysis of news recommendations by LLMs	L1, L3
55	Diakopoulos (2023)	L1: Algorithmic Gatekeeping	Framework for algorithmic auditing in journalism; accountability mechanisms for automated news	L5

Note: L1 = Algorithmic Gatekeeping; L2 = Personalized Agenda Construction; L3 = Predictive Agenda Amplification; L4 = Trust Mediation; L5 = Ethical Agenda Distortion. Studies serving a methodological function (PRISMA, thematic synthesis) are listed under “Methodology” without layer assignment. Total included studies: N = 55.

Academic Biography

Dr. Tommy Kibera Kiilu is a communication and media scholar with over 13 years of experience in academia, research, and university leadership. He is a Lecturer at Kenyatta University and previously served as Chairperson of the Department of Communication, Media, Film, and Theatre Studies (2020–2025). He also served as a Senior Lecturer and Chairperson at the Presbyterian University of East Africa. Dr. Kiilu combines academic expertise with industry experience in marketing and brand strategy. His research focuses on media studies, development communication, strategic communication, artificial intelligence in journalism, media ethics, digital transformation, and media literacy. He has co-authored two books published by Cambridge Scholars Publishing, contributed chapters to IGI Global volumes, and published extensively in peer-reviewed journals. Through his teaching, research, and scholarly engagement, he contributes to advancing knowledge on the evolving role of media and communication in the digital era.