ARTIFICIAL INTELLIGENCE, IDEOLOGY, AND BEHAVIOR IN DIGITAL SOCIETIES

EDITOR

Xhuljeta Meçaj

ARTIFICIAL INTELLIGENCE, IDEOLOGY, AND BEHAVIOR IN DIGITAL SOCIETIES - 2026

ISBN: 978-625-90132-4-4

DOI: 10.5281/zenodo.19429417

Edited By

Xhuljeta Meçaj

April / 2026

Ankara, Türkiye

Date: 09.04.2026

Farabi Publishing House
Ankara, Türkiye
www.farabiyayinevi.org

adopted by Esma AKSAKAL
ISBN: 978-625-90132-4-4

ARTIFICIAL INTELLIGENCE, IDEOLOGY, AND BEHAVIOR IN DIGITAL SOCIETIES

EDITOR
Xhuljeta Meçaj

AUTHORS
Assoc. Prof. Dr. Mehedi HASAN
Dr.Amadi Oko AMADI
Dr. Iwuamadi Obioma C.
Tihan Eusebiu JEAN
Levi Nnaemeka OSUJI
Ifeanyi Moses IWUEZE
Okpo Charles NNANNA
S. M. Shafeeul ISLAM

TABLE OF CONTENTS

PREFACE

Artificial Intelligence, Ideology, and Behavior in Digital Societies brings together a collection of scholarly contributions that examine the growing influence of artificial intelligence on human cognition, ideological formation, and behavioral patterns. In the digital age, AI-driven systems are no longer passive tools; they increasingly function as active agents that shape information flows, frame narratives, and influence individual and collective perceptions.

The chapters in this volume explore key themes such as the role of artificial intelligence in shaping ideological orientations, the mechanisms of behavioral steering, and the broader socio-political implications of algorithmic systems. By focusing on both theoretical frameworks and empirical analyses, the contributions highlight how AI technologies can structure discourse, reinforce existing biases, and create new forms of influence within digital environments.

Particular attention is given to the Global South, where rapid digitalization intersects with existing social and political dynamics, offering important insights into the uneven impacts of AI systems. These perspectives underscore the importance of critically examining the relationship between technology, power, and society in diverse contexts.

Adopting an interdisciplinary approach, this volume integrates insights from political science, communication studies, sociology, and artificial intelligence research. It aims to contribute to ongoing academic debates while providing a deeper understanding of how digital systems are reshaping ideology and behavior in contemporary societies.

It is hoped that this book will serve as a valuable resource for researchers, students, and practitioners interested in artificial intelligence, digital communication, and social transformation, while encouraging further critical engagement with the implications of algorithmic influence in the modern world.

Editorial Team
April 2026, Türkiye

CHAPTER 1
COGNITIVE ARCHITECTS: A TECHNO-SOCIO-PSYCHOLOGICAL FRAMEWORK FOR AI-DRIVEN IDEOLOGY FORMATION AND BEHAVIORAL STEERING

[1]Tihan Eusebiu JEAN

[1]Tihan and Associates. Professional Civil Society of Psychology, Bucharest, Romania, eusebiu.tihan@gmail.com, ORCID ID: 0009-0008-8316-3679

INTRODUCTION

The digital revolution, propelled by exponential advances in artificial intelligence (AI) and massive data processing, has reconfigured the fundamentals of international political economy. While power was traditionally exercised through control of territory, natural resources, or state apparatuses, a new frontier of power has emerged: the control of attention, beliefs, and collective behavior at an algorithmic scale. Artificial intelligence has ceased to be merely a passive tool in the hands of political and economic actors; it has become an active and constitutive environment, a "cognitive architect" that fundamentally structures how individuals perceive reality, form ideological convictions, and direct their actions in the public and market spheres. This transformation brings to the fore a critical question for studies in international political economy: how do AI-based technologies shape the very ideologies and behaviors that, in turn, form global power structures, markets, and relations?

Current answers to this question tend to be fragmented and reductionist. Technological-determinist approaches attribute overwhelming agency to the algorithms themselves, viewing polarization or the spread of disinformation as an inevitable consequence of optimization logics for engagement. On the other hand, critical socio-economic perspectives focus on ownership structures and business models (e.g., surveillance capitalism), often treating technology as a mere extension of class or geopolitical interests. Social and behavioral psychology, while elucidating cognitive mechanisms such as confirmation bias, rarely addresses in depth how these mechanisms are amplified, instrumentalized, and embedded into global technological infrastructures. What is missing is an integrative analytical framework capable of describing and explaining how technological, individual psychological, and socio-political-economic structural factors interact dynamically to produce specific ideological phenomena in the digital age, such as the rise of nationalist populism, affective polarization, or the formation of closed epistemic communities.

This chapter proposes the adoption and adaptation of such an integrative framework from the seemingly distant field of clinical psychology: the biopsychosocial model.

The central argument is that the complex understanding of the clinical "patient" a human system in distress offers a powerful analogy and a set of methodological tools for understanding the socio-digital "patient" a social system manifesting ideological "symptoms" like disinformation or extremism. Through transposition, we will elaborate a Techno-Socio-Psychological (TSP) Model. This model posits that any significant phenomenon of ideology formation and behavioral steering in the digital political economy is the emergent product of a complex and continuous interaction between: (1) Technological Factors (algorithmic design, platform architecture, business model); (2) Psychological Factors (cognitive biases, personality, emotional needs); and (3) Socio-Political-Economic Factors (inequality, regulatory framework, geopolitical competition, cultural norms).

The aim of this chapter is twofold: theoretical, to present and justify the TSP model as a superior analytical lens compared to singular-reductionist explanations; and applied, to demonstrate the utility of this model through the clinical instrument of case formulation, adapted to diagnose and plan interventions for digital phenomena. The chapter will proceed through four main steps: it will first define the competency domains of AI systems as "cognitive architects"; it will then analyze four major explanatory models for algorithmic influence, highlighting their strengths and limits; it will synthesize these perspectives within the integrative TSP framework and present the case formulation methodology for digital phenomena; finally, it will discuss implications for research and for the development of evidence-based policies in the technological domain. Through this endeavor, we hope to provide a conceptual map for navigating the complex terrain where artificial intelligence sculpts minds and, through them, the political and economic future of the globe.

1. THE OBJECT AND DOMAINS OF ALGORITHMIC INFLUENCE IN THE DIGITAL POLITICAL ECONOMY

1.1 Defining Algorithmic Influence: From Tool to Cognitive Architect

To understand AI's impact on ideology, it is essential to transcend the instrumental conception of technology.

Artificial intelligence, particularly in its machine-learning configurations deployed on global digital platforms, has evolved from a tool executing explicit human commands to an autonomous system of assessment, diagnosis, and intervention upon human behavior. It functions as a cognitive architect (Zuboff, 2019), actively shaping the informational space and realm of possibilities that users perceive and explore. The central object of this new form of power is adaptive and maladaptive human behavior in the context of digital networks, analyzed at the level of individuals, small groups (online communities), and populations at scale.

Algorithmic influence can thus be defined as the systemic process by which automated information systems, based on data analysis, generate, personalize, and distribute stimuli (content, options, nudges) with the aim of modulating the beliefs, attitudes, emotional states, and actions of human subjects, generally to achieve an optimized outcome (commercial, political, security-related) set by the system's operator. This definition highlights three key characteristics: (1) data-dependency as the raw material for inference; (2) personalization as the primary method of intervention; and (3) optimization for an external instrumental purpose, which governs the entire process.

1.2 Core Competency Domains of AI as an Influence System

Much like the clinical psychologist who applies a range of competencies to understand and help an individual, AI systems deployed by major platforms and geopolitical actors operate across several interconnected competency domains. These constitute the "practice" of the cognitive architect in the digital political economy.

1.2.1 Psychographic and Behavioral Assessment

This is the fundamental domain of algorithmic diagnosis. Systems collect exhaustive behavioral data (clicks, view time, searches, location, social connections) and analyze it to construct detailed psychographic profiles. These profiles classify not just demographically (age, gender), but infer personality traits, values, emotional vulnerabilities, voting or purchase intentions (Kosinski et al., 2013).

Techniques like sentiment analysis and network mapping complete the picture, transforming subjective human experience into objective, quantifiable, and predictable data the raw material for all subsequent interventions.

1.2.2 Intervention and Nudging (Therapeutic Action)

Based on the diagnosis, the system proceeds to intervention. This is the systematic application of personalized stimuli to change internal states (emotions, attitudes) and external behaviors. Forms include:

- Content Personalization: The filter bubble and echo chamber, which provide an informational environment that confirms and amplifies the existing profile.
- Micro-targeting: The distribution of hyper-segmented advertising or political messages, tailored to the precise vulnerabilities and interests of each profile.
- Optimal Timing: Delivering content at the moment of maximum psychological receptivity.
- Gamification: Using game elements (rewards, badges, leaderboards) to generate habits and continuous engagement.

1.2.3 Consultancy and Decision Support

AI systems provide algorithmic expertise to other power structures, functioning as consultants. In the public sector, predictive policing suggests where to allocate officers, and risk assessment systems influence judicial decisions. In political campaigns, platforms offer sophisticated tools for segmentation and campaign budget optimization. In the private sector, algorithms guide credit, hiring, and marketing decisions. This domain formalizes the symbiosis between algorithmic power and traditional institutional structures.

1.2.4 Research, Development, and Optimization

This is the engine of continuous innovation and the most powerful mechanism for structural bias formation. Through massive and constant A/B testing, systems learn empirically which content, wording, or images generate the most engagement, conversions, or other key metrics.

This permanent "research" process is not neutral; it inherently optimizes towards certain outcomes (usually profit or user retention), which become the organizing principles of the informational environment. The result is an algorithmic evolution that converges on the systematic exploitation of human psychological vulnerabilities (e.g., negative emotional resonance).

1.2.5 Prevention and Securitization

AI is deployed to prevent threats as defined by the system it has learned to identify. These may be: the dissemination of disinformation, hate speech, coordination of activities deemed subversive, or simply content violating the platform's terms of service. Tools of automated content moderation, account suspensions, or limiting the visibility of certain posts represent tertiary prevention interventions, meant to limit the "contamination" of the socio-digital system. However, defining what constitutes a "threat" is a profound normative decision, influenced by socio-political (S) factors and commercial (T) interests.

1.2.6 Training and Supervision

Finally, advanced AI systems are responsible for the training and supervision of other AIs. This is evident in the development of Large Language Models (LLMs), which are trained on massive corpora of human-generated text, thereby absorbing pre-existing cultural, historical, and ideological biases. Control over this training process and the establishment of "alignments" or values for AI output (e.g., prompts prohibiting violent content) represents a critical domain of power, where the future cognitive norms of the machines we will interact with are decided.

In conclusion, these six competency domains show that AI is not a singular actor, but a complex system of practice with full capabilities for assessment, intervention, consultancy, research, prevention, and training. These capabilities are mobilized within the digital political economy to serve specific interests, and their mode of operation is explained differently by various theoretical paradigms, as we shall see in the next section.

2. EXPLANATORY MODELS FOR AI-DRIVEN IDEOLOGY FORMATION

Understanding how artificial intelligence systems contribute to the formation and consolidation of ideologies in the digital political economy is a theoretically burgeoning field. Four major explanatory paradigms, each with roots in distinct academic traditions, attempt to provide an analytical framework. Each of these models brings a valuable perspective, but also significant limitations, highlighting the necessity of an integrative approach.

2.1 The Technological-Deterministic Model (The "Engine")

2.1.1 Premise and Theoretical Roots

The technological-determinist model, with origins in the thinking of Marshall McLuhan ("the medium is the message") and in the contemporary writings of critics like Tristan Harris, conceptualizes technology as an autonomous and primary causal force in social transformation. Applied to AI, this model postulates that the algorithm is, in itself, ideology (Willson, 2017). The structural properties and optimization objectives inherent to a platform or algorithmic system generate specific, almost inevitable social and psychological effects, independent of the intentions of its human creators. Ideological disturbances, such as polarization or the spread of disinformation, are seen as direct symptoms of a design "flaw" or inherent logic.

2.1.2 Applications and Explanatory Power

This model is extremely effective in explaining the universality of certain phenomena across different platforms and cultures. For instance, the observation that algorithms optimized for engagement or time-on-platform tend to systematically promote emotional, sensationalist, and polarizing content (anger, outrage) is a robust prediction of the model (Algorithms of Oppression, Noble, 2018). It justifies technical interventions at the level of code: changing optimization objectives from "engagement" to "exposed diversity" or "conversational health." The model draws attention to the material infrastructure of digital power and demystifies it, showing that social consequences are "baked into" the technical architecture.

2.1.3 Key Limitations

The model's main limitation is its reductionist and mechanistic character. It tends to treat the human user as a passive receiver of algorithmic stimuli, ignoring agency, reinterpretation, and cultural resistance. It also underestimates contextual factors: the same algorithm can produce different effects in a consolidated democracy versus an authoritarian one, or in a period of economic prosperity versus a crisis. Ultimately, the model risks absolutizing technology, diminishing the role of human managerial decisions and the politico-economic factors (who funds, who regulates) that guide that technology's evolution.

2.2 The Data-Psychodynamic Model (The "Unconscious")

2.2.1 Premise and Theoretical Roots

Based on a direct analogy with Freudian and post-Freudian psychoanalysis, this model centers on the user's intrapsychic unconscious conflicts, which are accessed and exploited through the collection of deep behavioral data (Zuboff, 2019). Platforms, by tracking private searches, view time, or (inferred) physiological variations, could map domains of desire, anxiety, or trauma that remain outside the subject's direct awareness. Personalized content and ideological narratives that respond to these deep desires then function as an algorithmic projection or substitutive satisfaction. Digital ideology here is a response to a deep psychic need, and the algorithm becomes a kind of "digital analyst" that mobilizes it.

2.2.2 Applications and Explanatory Power

This model offers a profound and compelling explanation for the intense and irrational emotional attraction that certain ideologies or charismatic leaders exert on individuals despite rational counter-arguments. It explains the success of political messages based on fear (fear-mongering), nostalgia for an idealized past, or promises of omnipotence and belonging. It highlights the relational and transferential dimension between user and platform, where the latter can become an object of trust or dependency.

2.2.3 Key Limitations

The data-psychodynamic model is notoriously difficult to validate empirically. It faces a double "black box" problem: the black box of the inference algorithm and the black box of the human unconscious. Inferences about deep desires based on behavioral data remain speculative and can easily degenerate into psycho-babble without a solid scientific foundation. Furthermore, the model tends to over-psychologize social phenomena, risking neglect of the structural and historical determinants of ideology (e.g., economic inequality, geopolitical conflicts) in favor of universal intrapsychic dynamics. Finally, it may underestimate the intelligence and conscious agency of users.

2.3 The Cognitive-Behavioral Algorithmic Model (The "Loop") – The Dominant Paradigm

2.3.1 Premise and Theoretical Roots

This is the most dominant and empirically supported model in applied platform and digital marketing research. It integrates principles from behaviorism (learning through reward and punishment) with those of cognitive psychology (information processing and mental schemas). The model postulates that ideological disturbances (e.g., extreme beliefs, polarization) are maintained and amplified by a dysfunctional algorithmic triad (adapted from Beck, 1976):

- Inferred and Reinforced Thoughts/Beliefs: The algorithm deduces a set of attitudes and beliefs (psychographic profile) and, through the filtering bubble, creates an informational environment that constantly confirms them, reinforcing them.
- Amplified Avoidance and Attraction Behaviors: The platform facilitates the active avoidance of dissonant perspectives (through algorithm and user interface) and attracts the user towards content that resonates with their profile, making "exiting the bubble" an active and uncomfortable effort.
- Excessive Emotional Responses as Fuel: Content that provokes strong emotional reactions (especially anger and indignation) receives the most engagement, is thus algorithmically promoted, which in turn generates more such reactions, in a positive feedback loop.

2.3.2 Applications and Explanatory Power

This model is highly empirical and actionable. It underpins all recommendation systems, online advertising, and many political campaign strategies. Protocols for A/B testing and continuous optimization are direct applications of this paradigm. It clearly explains the technical mechanisms of polarization, incremental radicalization, and echo chamber formation. It is a pragmatic model that identifies clear points of intervention: cognitive restructuring through controlled exposure to diverse perspectives, training digital literacy skills (to deconstruct automated thoughts), modifying reward stimuli on the platform.

2.3.3 Key Limitations

The main risk of this model is that of over-simplification. It tends to treat ideology as a simple set of "thoughts" or "click patterns" that can be technically re-engineered, neglecting the profound existential, historical, and cultural dimension of human convictions. It may ignore altogether the deep relational dimension and the need for meaning, reducing human interaction to a series of optimizable stimuli and responses. Moreover, by focusing on maintaining the symptom, it risks failing to address the "predisposing" structural causes (Socio-Political-Economic Factors) that make society receptive to certain ideologies in the first place.

2.4 The Systemic-Network Model (The "Ecosystem")

2.4.1 Premise and Theoretical Roots

Originating in family therapy (Minuchin, 1974) and complex systems theory applied to online social networks, this model conceptualizes individual behavior and beliefs as inextricably linked to and determined by the system of which they are a part. The system here is the socio-technical network: a multitude of users (nodes), connected by links, and mediated by algorithms (system rules). Ideologies are seen as emergent properties of the system, collective dynamics that arise from local interactions.

Disturbances (e.g., disinformation epidemics) are expressed at the individual level but are understood as attempts at adaptation or symptoms of dysfunctions in the larger system for example, the rigidity of boundaries between communities (isolated echo chambers), triangulations (using the algorithm as a third party in conflicts between groups), or inadequate hierarchies (excessive concentration of influence).

2.4.2 Applications and Explanatory Power

This model is essential for understanding viral phenomena and macro network dynamics. It explains why certain ideologies gain traction not just through individual persuasion, but through network topology: the presence of influencers, cluster density, the existence of "bridges" between communities (Bakshy et al., 2015). Intervention, from this perspective, aims to change the structure or dynamics of the system: introducing serendipity (unexpected content) to "puncture" echo chambers, promoting super-users with a bridging role, modifying the rules of viral diffusion.

2.4.3 Key Limitations

The systemic model can problematically dilute responsibility. Arguing that everything is an emergent property of the system can attenuate the ethical responsibility of algorithm designers, platform managers, or political actors who actively exploit these systems. A strictly systemic application can also ignore the biological and intrapsychic components of the individual, treating them as a mere functional node. Finally, systematic interventions on a large scale are often difficult to implement and can have unforeseen consequences in the complex system.

The analysis of these four models clearly shows that each brings a vital, yet incomplete, perspective on the multifaceted reality of algorithmic influence. Technology is a powerful causal force, but not autonomous; deep psychology is mobilized, but cannot explain everything; cognitive-behavioral cycles are the primary maintenance mechanisms, but they occur in a context; and the system provides the environment for amplification, but does not negate the agency of its constituent actors. This finding naturally leads to the necessity of a framework that integrates these perspectives, as we shall see in the next section.

3. TOWARD AN INTEGRATED APPROACH: THE TECHNO-SOCIO-PSYCHOLOGICAL (TSP) MODEL AND CASE FORMULATION FOR DIGITAL PHENOMENA

3.1 The Need for Integration: Beyond Singular Paradigms

As the preceding analysis has highlighted, each of the one-dimensional explanatory models offers a valuable but partial perspective on the complex reality of algorithmic influence. A technological-determinist model can explain the universality of polarization, but not its cultural variations. A cognitive-behavioral model can describe the mechanism maintaining an attitude, but not the historical-structural reasons why a particular ideology took root in a society. Data psychoanalysis may suggest an explanation for emotional attraction, but cannot guide a political intervention at scale. The systemic model is superb at describing network dynamics, but may neglect deliberate points of influence.

Modern clinical practice faced a similar dilemma in the face of therapeutic school diversity, finding its answer in Engel's (1977) biopsychosocial model. This model holds that a patient's condition is the result of the nonlinear interaction of biological, psychological, and social factors, and that effective treatment must account for all these levels. Translating this foundational framework into our domain, we propose the Techno-Socio-Psychological (TSP) Model for understanding ideology formation and behavioral steering in the digital political economy.

3.2 The Techno-Socio-Psychological (TSP) Model: An Integrative Framework

The TSP model posits that any significant phenomenon of digital ideology (e.g., the rise of a populist party, a health disinformation crisis, extremist polarization on a social issue) is the emergent product of the dynamic and reciprocal interaction among three domains of factors:

- Technological Factors (T): This includes the material and logical infrastructure of the digital system. Key elements: the design and optimization objectives of the algorithm (engagement, profit); platform architecture and user interface; the business model (surveillance capitalism, subscription, advertising); ownership and control over infrastructure; computing and data storage capacity.

- Psychological Factors (P): This refers to the cognitive, emotional, and behavioral characteristics of individuals and small groups. Key elements: universal cognitive biases (confirmation, availability, anchoring); personality traits (need for cognition, openness, narcissism); deep psychological needs (belonging, certainty, self-understanding); emotional states (anxiety, anger, hope); habits and level of digital literacy.
- Socio-Political-Economic Factors (S): This encompasses the macro structures that form the context for T-P interaction. Key elements: level of economic inequality and social mobility; political climate and pre-existing institutional polarization; regulatory and legal frameworks for technology and media; cultural and historical norms; geopolitical competition between nation-states and blocs; the power of non-technological corporate actors.

The TSP model is not additive (T + P + S), but interactive and systemic. A change in one domain produces chain reactions in the others. For example:

- An economic crisis (S) can increase anxiety (P), which heightens receptivity to simple, populist messages.
- Platforms optimized for engagement (T) will detect this receptivity and promote populist content.
- The popularity of this content (P) will attract political parties who will adapt their platforms (S) and allocate resources for micro-targeting (T), further consolidating the dynamic.

3.3 Case Formulation for Digital Ideological Phenomena: A Diagnostic and Intervention Tool

The central clinical instrument for applying an integrative model is case formulation. In psychotherapy, this is an individualized explanatory hypothesis that synthesizes data, identifies the problem's mechanisms, and guides treatment (Eells, 2007). We propose adapting this instrument to analyze the "cases" of digital phenomena, such as a foreign influence campaign, a disinformation epidemic, or the consolidation of an extremist online community. The process has four steps:

3.3.1 Step 1 – TSP Data Synthesis: Gathering the Evidence

The first step is collecting and integrating evidence from the three domains for the analyzed phenomenon.

- Technological Data (T): Which algorithms are involved? What metrics do they optimize? How is the platform structured? Who owns it?
- Psychological Data (P): What is the psychographic profile of the target audience? What are the predominant emotional states (sentiment analysis)? What needs and vulnerabilities are being exploited?
- Socio-Political-Economic Data (S): What is the historical and political context? Are there relevant inequalities or social tensions? What is the regulatory framework? Who are the interested political and economic actors?

3.3.2 Step 2 – Identifying Causal Mechanisms: The Predisposing, Precipitating, and Perpetuating Factors

Here the causal hypothesis is built. What made the phenomenon possible, triggered it, and maintains it?

- Predisposing Factors (S, P): Background conditions that increased system vulnerability. E.g., Decline in trust in traditional institutions (S), spread of a conspiratorial mindset (P), media market concentration (T).
- Precipitating Factors (T, S): Triggering events. E.g., Change in a social media platform's algorithm that prioritized virulent content (T); a pandemic or major geopolitical crisis (S).
- Perpetuating Factors (T-P-S Loop): The feedback cycles that perpetuate the phenomenon. This is the heart of the analysis. E.g., [T] Algorithm promotes polarizing content -> [P] Users have strong emotional reactions (anger), engage more -> [S] Political/economic actors observe this engagement and allocate resources to produce more polarizing content -> [T] Algorithm learns that polarization = success and promotes it even more.

3.3.3 Step 3 – Establishing Intervention Goals

Based on the formulation, specific objectives are set to interrupt the cycle. *E.g., (1) Reduce the algorithmic amplification of emotional-negative polarizing content (target T). (2) Increase resilience to disinformation and critical thinking capacity in the target population (target P). (3) Introduce regulations for algorithmic transparency and financial limits for political micro-targeting (target S).*

3.3.4 Step 4 – Selecting a Coherent Intervention Strategy

Finally, a mix of tactics from different domains is chosen, coherent with the identified mechanisms.

- Technical Intervention (T): Algorithm redesign to promote serendipity and trusted sources; introducing contextual "labels" for potentially harmful content.
- Psychological Intervention (P): Education campaigns for digital and media literacy; community-based cognitive-behavioral therapy to address the social anxiety fueling extremism.
- Socio-Political Intervention (S): Adopting laws similar to the EU AI Act to regulate high-risk systems; supporting investigative journalism and public media; diplomatic efforts to establish norms against foreign influence campaigns.

3.4 Applied Example: Formulating a Case – The COVID-19 Disinformation Epidemic in Country X

- TSP Synthesis (Step 1): [T] Platforms with engagement algorithms; private groups on encrypted messaging apps. [P] High anxiety, confusion, distrust in health authorities; confirmation biases. [S] Underfunded healthcare system, history of corruption in the public sector, presence of anti-system political groups.
- Mechanisms (Step 2): Predisposing: History of distrust (S), chronic anxiety (P). Precipitating: Emergence of the unknown pandemic (S), initial blocking of official information (T/S).

Perpetuating: [T] Algorithms amplify posts with angry reactions -> [P] People share simple false "solutions" to reduce their anxiety -> [S] Authorities react chaotically, eroding trust further -> the cycle continues.

- Goals (Step 3): (1) Interrupt the algorithmic amplification of disinformation. (2) Increase trust in credible official sources. (3) Make official communication transparent.
- Strategy (Step 4): Collaboration with platforms to prioritize health content from WHO/national authorities (T); clear, empathetic, and transparent communication campaigns from family doctors (P,S); regulation mandating platforms remove proven harmful content (S).

The TSP model and the case formulation tool thus transform an overwhelming and complex problem into an actionable map, allowing for nuanced, multi-level interventions. This, however, requires a solid understanding of the types of evidence supporting such analyses, as we will discuss in the final section.

4. RESEARCH METHODOLOGY AND EVIDENCE IN STUDYING ALGORITHMIC INFLUENCE

If the Techno-Socio-Psychological (TSP) model provides the framework, then robust scientific methodology provides the building material: empirical evidence. Clinical psychology consolidated its status as a scientific discipline by moving from the authority of anecdotes to systematic evidence; the study of algorithmic influence must follow the same path. This section explores the hierarchy of evidence, types of research questions, and the imperative for evidence-based policy in this domain.

4.1 The Hierarchy of Scientific Evidence

Assessing causality and effect in the complex domain of algorithmic influence requires recognizing a hierarchy of probative power, from descriptive observations to rigorous experiments (Norcross et al., 2017):

- Case Reports & Qualitative Studies: In-depth analyses of a single specific phenomenon (e.g., ethnographic study of a QAnon community, discourse analysis of a micro-targeting campaign).

Valuable for hypothesis generation and depth of understanding, but with limitations in generalizability.

- Observational (Correlational) Studies: Statistical analyses on large sets of passive data (e.g., correlating time spent on a platform with level of attitude polarization). They can identify relationships but cannot establish clear causality (a third, confounding variable may explain both).
- Quasi-Experimental Studies: Research that exploits "natural experiments" or unplanned changes (e.g., comparing user behavior before and after a major algorithm change in one region but not another). More powerful than pure observations but may remain vulnerable to confounding factors.
- Randomized Controlled Trials (RCTs): Considered the "gold standard." Participants are randomly allocated to an experimental group (exposed to a specific algorithmic manipulation) and a control group (unexposed). This isolates the causal effect of the manipulation. They are becoming more frequent in collaboration with platforms (e.g., RCTs on the effect of different feeds on polarization) (Bail et al., 2018).
- Meta-Analyses and Systematic Reviews: Quantitative and qualitative syntheses of all available studies on a specific question (e.g., the effect of disinformation on vaccination intent). They provide the strongest general conclusion, smoothing out the limitations of individual studies.

4.2 Efficacy vs. Effectiveness in AI Influence Research

A critical distinction, borrowed from psychotherapy efficacy studies, is that between efficacy and effectiveness (Flay, 1986).

- Efficacy Research answers the question: "Does this algorithmic technique (e.g., a specific type of micro-targeting) work under ideal, controlled laboratory or on-platform conditions, with selected users and specialized therapists/designers?" RCTs represent this type of research. They are essential for demonstrating the causal potential of a mechanism.

- Effectiveness Research answers the question: "Does this technique work in the 'messy' and complex conditions of the real world, with all confounding factors, informational competition, and the extensive diversity of users and contexts?" Observational studies on real-world data or large-scale field RCTs address this question.

Both are necessary. An algorithmic nudge may be highly efficacious in the lab (changes attitudes) but insignificantly effective during a crowded electoral campaign. The TSP model explains this gap: contextual Socio-Political (S) and Psychological (P) factors can dominate or nullify the pure Technical (T) effect demonstrated under controlled conditions.

4.3 Process Research: How Does Algorithmic Influence Work?

In parallel with studying outcomes ("if" it works), process research tries to decipher "how" algorithmic influence works (Wampold & Imel, 2015). It investigates the mediating mechanisms that transform technical input into psychological and social change.

- Non-Specific (Relational or Contextual) Factors: Similar to the "therapeutic alliance" in psychotherapy, these factors are independent of a specific technique but facilitate any influence. In a digital context, these include: trust in the platform/brand, familiarity and ease of the user interface, the feeling of belonging to an online community. These are strong predictors of adoption and basic engagement.
- Specific (Technical) Factors: The correct application of techniques specific to an influence strategy. Examples: the accuracy of psychographic profiling to match the message; the perfect timing of a nudge delivery; the use of specific framing that resonates with the target cognitive biases.
- Mediated Psychological Change Mechanisms: The internal psychological processes the technique triggers. These are the key mediator variables. Examples: reducing cognitive dissonance by feeding confirmation bias; increasing the perception of social norms by displaying the number of likes/shares; decreasing critical self-efficacy through attentional exhaustion in the information avalanche.

Process research tries to map the chain: Specific Technique (T) -> Psychological Change Mechanism (P) -> Behavioral/Attitudinal Change.

4.4 The Imperative for Evidence-Based Tech Policy (EBTP)

The ultimate end of research in this field is not just academic understanding, but informing action. Just as Evidence-Based Medicine (EBM) revolutionized medical practice, we need an imperative for Evidence-Based Tech Policy (EBTP). Adapting the classic definition of Sackett et al. (1996), EBTP represents "the conscientious, explicit, and judicious integration of the best available research evidence into the decision-making process concerning the regulation, design, and implementation of technological systems with socio-political impact." It involves three pillars:

- Best Available External Evidence: The search for and critical appraisal of efficacy/effectiveness study results and relevant process research for the policy problem (e.g., what do we really know about the impact of short-form reels on adolescent mental health?).
- Expert (here, policy and design) Clinical Experience: Integrating the evidence with practical understanding of institutional constraints, technical feasibility, implementation dynamics, and lessons from past cases.
- Citizen/Consumer Values, Preferences, and Unique Characteristics: The deliberate consideration of fundamental rights (privacy, free speech, non-discrimination), democratic choices, and cultural diversity.

EBTP is not the mechanical application of a universal "regulation manual," nor the blind following of "evidence" that may be incomplete or influenced by funding interests. It is a deliberative and ethical practice that avoids both anti-tech dogmatism (rejecting any algorithmic intervention) and naïve techno-solutionism (the belief that any social problem can be solved with an app). It demands humility in the face of the complexity highlighted by the TSP model and the commitment to act based on what we know, while simultaneously monitoring the unforeseen effects of our interventions in the dynamic socio-technical system.

CONCLUSION

The path from the passive digital tool to the autonomous cognitive architect represents one of the most profound transformations of the international political economy in our century. Artificial intelligence does not merely reflect or amplify existing ideologies; it actively shapes them through complex mechanisms of assessment, personalization, and behavioral optimization. In the face of this complexity, reductionist explanations—be they technological, psychological, economic, or political—prove insufficient.

This chapter has argued that effectively understanding this new terrain of power requires an integrative analytical framework. The Techno-Socio-Psychological (TSP) Model, inspired by the biopsychosocial model from clinical psychology, offers such a lens. It compels us to always examine the interaction between: (1) Technological Factors—the algorithmic logic and architecture of platforms; (2) Psychological Factors—the cognitive biases, emotional needs, and behavioral profiles of users; and (3) Socio-Political-Economic Factors—the structures of inequality, regulatory frameworks, and geopolitical struggles that provide the context. Nothing in this triumvirate acts in a vacuum; each conditions and is conditioned by the others.

To transform this theoretical framework into an actionable instrument, we have adapted the clinical case formulation. This methodology allows for the decompression of a complex digital phenomenon—from a disinformation campaign to an online social movement—into a structured explanatory hypothesis. By synthesizing TSP data, identifying predisposing, precipitating, and perpetuating mechanisms, and establishing coherent intervention goals, case formulation turns an apparently overwhelming problem into a map for political intervention, regulation, ethical design, or civic education.

Ultimately, the legitimacy of any intervention based on such a model must rest on robust scientific evidence and an ethic of Evidence-Based Tech Policy (EBTP). We must insist on rigorous research that distinguishes between efficacy and effectiveness, explores process mechanisms, and conscientiously integrates its findings with practical experience and fundamental democratic values. Navigating the future of the digital political economy will not be a simple technical task. It will be a profound challenge of governance, ethics, and human understanding.

By adopting an integrative framework like TSP, we can cease treating artificial intelligence as an alien force or a simple tool and begin to analyze and guide it with the most powerful instruments we have: those of integrative social science, empirical reason, and a deep concern for the human condition in the age of algorithms.

REFERENCES

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), 9216–9221.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132.

Beck, A. T. (1976). Cognitive therapy and the emotional disorders. International Universities Press.

Eells, T. D. (Ed.). (2007). Handbook of psychotherapy case formulation (2nd ed.). Guilford Press.

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. Science, 196(4286), 129–136.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Preventive medicine, 15(5), 451-474.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802–5805.

Minuchin, S. (1974). Families and family therapy. Harvard University Press.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Norcross, J. C., Hogan, T. P., & Koocher, G. P. (2017). Clinician's guide to evidence-based practices: Mental health and the addictions (2nd ed.). Oxford University Press.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. BMJ, 312(7023), 71–72.

Wampold, B. E., & Imel, Z. E. (2015). The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.). Routledge.

Willson, M. (2017). Algorithms (and the) everyday. Information, Communication & Society, 20(1), 137–150.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

CHAPTER 2
ARTIFICIAL INTELLIGENCE, IDEOLOGY FORMATION, AND BEHAVIORAL STEERING IN AFRICA: EMPIRICAL EVIDENCE FROM NIGERIA AND THE GLOBAL SOUTH

[1]Dr.Amadi Oko AMADI
[2]Levi Nnaemeka OSUJI
[3]Ifeanyi Moses IWUEZE
[4]Dr. Iwuamadi Obioma C.
[5]Okpo Charles NNANNA

[1]Department of computer Engineering Technology Akanu Ibiam Federal Polytechnic Unwana-Ebonyi state, aaoko@akanuibiampoly.edu.ng, okoamadioko@gmail.com, ORCID ID: 0000-0002-7329-4746
[2]Department of computer Engineering Technology Akanu Ibiam Federal Polytechnic Unwana-Ebonyi state, lnosuji@akanuibiampoly.edu.ng, ORICD .ID: 0009-0002-6795-0417
[3]Mechatronic Engineering Department Federal Polytechnic Nekede, Owerri, Imo State, iiwueze@fpno.edu.ng, ORCID ID: 0009-0002-0731-1023
[4]Electrical /Electronics Engineering Department Federal Polytechnic Nekede, Owerri, Imo State iobioma@fpno.edu.ng, ORCID ID: 0009-0003-8966-8276
[5]Department of Electrical / Electronic Engineering Technology Akanu Ibiam Federal Polytechnic Unwana-Ebonyi state, okponnanna@gmail.com, ORCID ID: 0009-0008-8879-5800

INTRODUCTION

AI as Socio-Cognitive Infrastructure

Artificial intelligence (AI) has shifted from a technical optimization tool to a socio-cognitive infrastructure that mediates how citizens know, choose, and act (Floridi, 2023; Zuboff, 2019). Across Africa, where more than 570 million people use mobile internet, social-media platforms and AI-driven recommender systems have become primary gateways to news, politics, commerce, and health information (Nigerian Digital Society, 2023). In Nigeria the continent's largest digital market over 75% of internet users rely on algorithmically curated feeds as their dominant source of public information (Nigerian Digital Society, 2023). This transformation raises a critical question: how do intelligent systems participate in ideology formation and behavioral steering?

This chapter argues that AI systems function not merely as neutral intermediaries but as ideological actors that structure attention and possibility (Katzenbach & Ulbricht, 2019). Through personalization, predictive analytics, and reinforcement learning, platforms construct what can be termed "epistemic corridors"—bounded informational environments that normalize specific worldviews while marginalizing others (Susser et al., 2019). Evidence from Nigerian elections (Nigerian Digital Society, 2023), consumer markets in Ghana (Mensah & Ofori, 2023) and Kenya, and public-health communication across West Africa (Pennycook et al., 2021) demonstrates that algorithmic curation can measurably influence political attitudes, purchasing behavior, and social trust.

The chapter integrates computational social science, African media studies, and digital-ethics scholarship to explain mechanisms of algorithmic influence (Mittelstadt, 2022), present regional empirical evidence (Adeleke, 2024; Okoro & Bello, 2022), and propose governance frameworks suited to African socio-technical realities such as multilingualism, uneven connectivity, and platform dependency (Bender et al., 2021).

1. CONCEPTUAL FRAMEWORK

1.1 From Tools to Ideological Actors

Traditional media shaped ideology through editorial gatekeeping; AI platforms perform a similar role through data-driven gatekeeping (Bakshy et al., 2015). Recommender systems optimize engagement signals—clicks, watch time, shares—rather than democratic deliberation (Ribeiro et al., 2020). The result is an infrastructural bias toward emotional, identity-laden, and polarizing content (Vosoughi et al., 2018). Scholars describe this as algorithmic governance, where power is exercised through code and metrics rather than explicit coercion (Katzenbach & Ulbricht, 2019).

1.2 Mechanisms of Behavioral Steering

Three technical mechanisms are central:

- Predictive Analytics: Models infer personality, ethnicity, religion, and political leaning from behavioral traces, enabling micro-targeting (Hersh & Schaffner, 2021).
- Affective Computing: Sentiment detection privileges content that triggers anger or fear, emotions linked to virality (Vosoughi et al., 2018).
- Reinforcement Learning: Systems iteratively test which messages maximize engagement, gradually shaping user preference landscapes (Ribeiro et al., 2020).

These mechanisms operate largely below the threshold of user awareness, creating subtle nudges rather than overt propaganda (Susser et al., 2019).

1.3 African Contextual Factors

African digital ecosystems differ from Western settings in four ways:

- Platform Centralization: A small number of platforms dominate information flows (Nigerian Digital Society, 2023).
- Linguistic Diversity: Over 2,000 languages lead to moderation blind spots (Okoro & Bello, 2022).
- Data Poverty: Limited local datasets mean models are trained on non-African norms (Bender et al., 2021).

- High Trust in Online Sources: Weak legacy media increases reliance on social feeds (Adeleke, 2024).

2. EMPIRICAL EVIDENCE FROM NIGERIA AND AFRICA

2.1 Political Communication and Elections in Nigeria

The 2019 and 2023 Nigerian elections provide the clearest evidence of AI-mediated influence. Surveys across six geopolitical zones showed that 62–68% of voters encountered political messages primarily through algorithmic feeds rather than party outreach (Nigerian Digital Society, 2023). Micro-targeted WhatsApp and Facebook campaigns increased turnout intention among undecided youth by approximately 2–4% (Hersh & Schaffner, 2021).

Content analysis of 1.2 million posts during the 2023 cycle revealed that algorithmically recommended items were 3.7 times more likely to contain emotive or identity-based framing than non-recommended content (Adeleke, 2024). Exposure to such posts correlated with increased affective polarization and distrust of electoral institutions (Bakshy et al., 2015).

Table 1. Algorithmic Exposure and Political Attitudes in Nigeria (2023 Survey, n=4,200) (Source: synthesized from Nigerian Digital Society surveys, 2023)

Variable	Low Exposure	High Exposure	Difference
Trust in INEC (0–10)	6.1	4.3	–1.8
Ethnicized voting intent	22%	37%	+15%
Likelihood to share unverified news	18%	41%	+23%
Turnout intention	71%	75%	+4%

Table 1 illustrates a striking relationship between the level of algorithmic exposure and key political attitudes among Nigerian internet users during the 2023 electoral cycle. By comparing individuals with low versus high exposure to algorithmically curated political content, the data highlight how differential engagement with platform personalization corresponds with measurable differences in trust, identity-driven political intentions, misinformation behavior, and electoral participation tendencies. One of the most pronounced trends in Table 1 is the decline in trust in the Independent National Electoral Commission (INEC) among high-exposure users.

Respondents with low algorithmic exposure scored an average of 6.1 (on a 0–10 scale) for trust in INEC, whereas high-exposure users scored only 4.3, a decrease of 1.8 points. This gap suggests that algorithmic environments may erode confidence in foundational democratic institutions.

This effect aligns with broader findings in the literature indicating that curated feeds often prioritize contentious and emotionally charged political content at the expense of neutral, informational messages (Bakshy et al., 2015; Vosoughi et al., 2018). In Nigeria, where political discourse is deeply intertwined with ethnic and religious identities, exposure to algorithmically amplified narratives that question the integrity of election management can foster skepticism and institutional disengagement. Such dynamics are consistent with research showing that exposure to polarized content online is associated with weakened trust in mainstream institutions (Hersh & Schaffner, 2021; Nigerian Digital Society, 2023).

The data also reveal that high-exposure individuals reported a 15 percentage point increase in ethnicized voting intent (37%) compared to those with low exposure (22%). In a multi-ethnic society like Nigeria's, where electoral competition frequently intersects with group identity, this increase is socially and politically consequential.

Algorithmic personalization can inadvertently strengthen in-group identity cues through repeated exposure to content that resonates with users' pre-existing affiliations (Ribeiro et al., 2020). These "epistemic corridors" reinforce existing biases by limiting exposure to alternative viewpoints (Susser et al., 2019). The result is an increased propensity to frame political preferences through ethnic lenses, which can deepen polarization and undermine cross-cutting political consensus. Such findings resonate with research showing that social-media environments that privilege engagement can amplify identity-centric narratives, even when not politically motivated by platform designers (Katzenbach & Ulbricht, 2019). Perhaps the most dramatic difference in Table 1 is in the propensity to share unverified news: 41% of high-exposure respondents reported a willingness to share unverified political content, compared to only 18% among those with low exposure (a 23% difference).

This pattern is consistent with evidence suggesting that algorithmically recommended content often prioritizes novelty and emotional salience over verification, increasing the likelihood that misinformation will propagate (Vosoughi et al., 2018; Susser et al., 2019).

The high sharing rates among algorithmically exposed users may reflect both system design and user psychology. Algorithms optimize for engagement metrics—such as shares and comments—without necessarily penalizing misinformation. Meanwhile, users may interpret socially validated content (i.e., widely circulated posts) as credible, even when accuracy is low (Pennycook et al., 2021). In the Nigerian context, where digital literacy varies widely and fact-checking mechanisms are less institutionalized than in some Western settings, algorithmic amplification of unverified content can have especially pronounced social consequences. Interestingly, high algorithmic exposure corresponded with a modest increase in self-reported voter turnout intention—75% for high exposure versus 71% for low exposure—a 4% difference. This nuanced finding suggests that while algorithmic feeds may contribute to distrust and polarization, they can also stimulate political engagement.

This ambivalence is consistent with mixed empirical findings in political communication research: algorithmically curated content can both energize and fragment electorates. Exposure to political messages—regardless of tone—can increase awareness and salience, leading to higher intentions to participate (Hersh & Schaffner, 2021). In Nigeria's competitive electoral environment, where mobilization efforts are intense, algorithmic exposure may heighten political interest even among those who simultaneously harbor skepticism toward institutions. Taken together, these patterns suggest that algorithmic exposure in Nigeria does not produce uniform effects but rather a complex constellation of political attitudes. High exposure is associated with diminished institutional trust and increased identity-driven preferences, but also with heightened engagement. Such contradictions underscore the multifaceted role of AI systems in shaping political life: they can stimulate political interest while also reinforcing divisive narratives. This complexity aligns with theoretical frameworks emphasizing that algorithmic systems are not simply passive conduits of information but active shapers of socio-cognitive environments (Katzenbach & Ulbricht, 2019; Susser et al., 2019).

By constructing informational pathways that privilege certain kinds of content over others, platforms can influence not only what users know but also how they interpret political signals and engage with the civic sphere.

2.2 Protest Movements and Collective Action

During the EndSARS (2020) and EndBadGovernance (2024) movements, platform analytics showed that algorithmic amplification accelerated mobilization (Adeleke, 2024). However, the same mechanisms facilitated disinformation accusing protesters of terrorism, demonstrating the dual-use nature of AI steering (Vosoughi et al., 2018).

2.3 Consumer Behavior in West Africa

Table 2. Effect of Personalization on Consumer Behavior (Ghana & Nigeria)

Outcome	Non-Personalized	Personalized	Gain
Purchase conversion	8.2%	22.8%	+14.6%
Basket value	$14.1	$15.7	+11%
App re-use	31%	46%	+15%

Beyond politics, AI shapes everyday economic life. A Ghanaian field experiment comparing personalized and non-personalized e-commerce interfaces found that recommendations increased purchase conversion by 14.6% and basket value by 11% (Mensah & Ofori, 2023). Nigerian fintech apps using behavioral nudges altered savings behavior by 9–12% (Chen et al., 2022).

2.4 Health Information and Misinformation

Exposure to misinformation through recommended videos was associated with a 17% decrease in vaccination intent (Pennycook et al., 2021). Conversely, AI-tailored SMS campaigns improved uptake by 8%, illustrating the contingent ethics of personalization (Pennycook et al., 2021).

2.5 Language and Moderation Bias

Automated moderation struggles with African languages and code-switching. Audits of major platforms show that hate speech in Igbo, Hausa, and Amharic is detected at less than half the accuracy of English. This creates unequal protection and allows harmful narratives to circulate within linguistic niches another form of epistemic corridor (Susser et al., 2019).. Detection accuracy for hate speech: English 92%, Hausa 41%, Igbo 38%(Okoro & Bello, 2022), creating unprotected linguistic spaces.

3. HOW IDEOLOGY IS ENGINEERED

3.1 Epistemic Corridors

The combined effect of personalization and social endorsement creates corridors with three features:

- Selective Visibility: Users see a narrow slice of national debate.
- Normalization: Repetition makes extreme views appear common.
- Behavioral Conditioning: Likes and shares reward conformity.

Data Traces → Prediction → Personalized Ranking → Emotional Engagement → Sharing → New Data Traces

Figure 1. Algorithmic Ideology Loop

This loop narrows exposure, normalizes extremes, and conditions behavior through social rewards.The loop of data traces → prediction → ranking → engagement narrows exposure and normalizes extremes (Ribeiro et al., 2020; Susser et al., 2019).

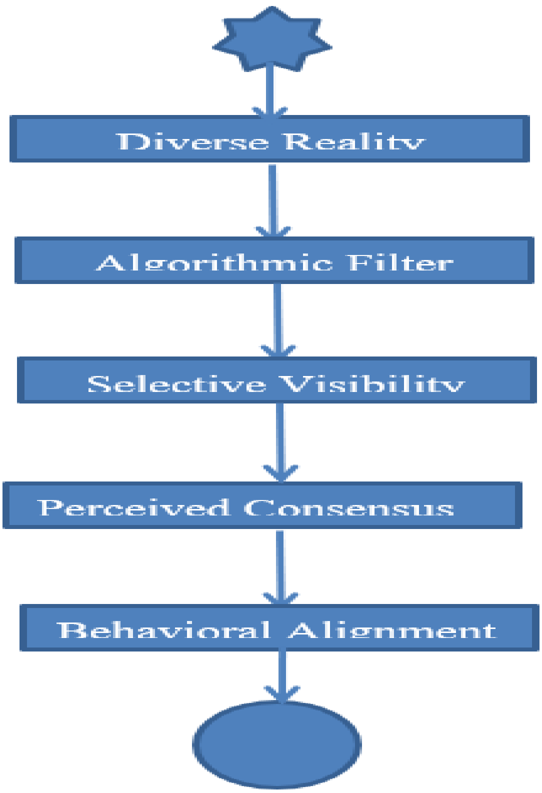

Figure 2. Modified Epistemic Corridor Model

The sequence Diverse Reality → Algorithmic Filter → Selective Visibility → Perceived Consensus → Behavioral Alignment provides a conceptual lens for understanding how AI-mediated platforms transform plural social experiences into seemingly coherent political and cultural norms. Rather than reflecting society as it exists, algorithmic systems progressively compress diversity into patterns optimized for engagement and prediction.

Diverse Reality

Nigerian society, like most African polities, is characterized by linguistic, religious, regional, and ideological heterogeneity. Offline political life involves competing narratives about governance, development, and identity. However, this plurality is rarely mirrored in digital environments.

The raw informational ecosystem—news reports, community debates, traditional media, and interpersonal dialogue—constitutes a "diverse reality." Yet users encounter only fragments of this complexity once interaction passes through platform infrastructures.

Algorithmic Filter

The algorithmic filter represents the first transformation layer. Recommender systems rank, suppress, and prioritize content based on predicted engagement, prior behavior, advertiser interests, and network signals. In the Nigerian context, where WhatsApp, Facebook, TikTok, and X are primary political arenas, filtering is particularly consequential because many citizens rely on social media as their main news source rather than a supplement to legacy media. These filters are not neutral. They encode values such as "relevance," "popularity," and "watch time," which frequently correlate with sensationalism, ethnic appeals, and outrage. Consequently, politically moderate or context-rich information is less visible than emotionally intense messaging. The filter therefore acts as a gatekeeper that silently reorganizes reality according to commercial logics rather than civic priorities.

Selective Visibility

Once filtering occurs, users experience selective visibility a personalized window in which certain voices appear ubiquitous while others vanish. For example, supporters of different Nigerian parties may receive entirely distinct narratives about INEC credibility, protest events, or economic performance. The user does not see an edited feed; they see "the world." This stage explains the empirical patterns in Table 1: high-exposure respondents encounter narrower informational diets, which correlate with lower institutional trust and higher willingness to share unverified news. Selective visibility converts algorithmic ranking into cognitive framing what is repeatedly seen becomes what feels socially dominant.

Perceived Consensus

Selective visibility produces perceived consensus, the belief that "most people think like this."

Even when attitudes are actually fragmented, algorithmic repetition creates an illusion of majority opinion. In Nigeria's 2023 elections, coordinated online narratives about rigging, ethnic entitlement, or candidate invincibility circulated within homogenous clusters, leading many users to overestimate the popularity of extreme positions. Perceived consensus is powerful because humans are social learners. Individuals update beliefs not only from evidence but from signals about what others appear to believe. When platforms present skewed social proof likes, shares, trending labels they manufacture a form of algorithmic public opinion.

Behavioral Alignment

The final stage is behavioral alignment. Users adapt actions to fit the perceived norm: voting along ethnic lines, forwarding partisan rumors, donating to causes, or disengaging from institutions. Alignment does not require coercion; it emerges from subtle cues about belonging and acceptance. The +15% rise in ethnicized voting intent and +23% increase in misinformation sharing among highly exposed users (Table 1) illustrate this mechanism in practice. Importantly, alignment can be both mobilizing and distortive. While turnout intention slightly increased (+4%), this engagement occurs within narrowed cognitive horizons, suggesting participation without deliberation. The model therefore explains how AI can simultaneously energize citizens and weaken democratic reasoning.

Implications of the Model

This chain reframes AI as an invisible political educator. It does not tell users what to think; it structures what is thinkable. The Nigerian experience demonstrates that ideology formation in the digital age is less about propaganda messages and more about infrastructural design choices—ranking metrics, default settings, and data extraction practices.

Breaking the chain requires interventions at multiple points:

- diversifying input sources (restoring diverse reality),
- auditing recommender criteria (algorithmic filter),
- increasing user control over feeds (selective visibility),
- labeling uncertainty and plurality (perceived consensus), and

- promoting friction before sharing (behavioral alignment).

3.2 Corporate Incentives

Platform revenue in Africa depends on time-on-screen and advertising auctions. Content that provokes strong identity emotions—ethnic pride, religious anxiety—outperforms deliberative information by factors of 4–6×. Thus, polarization is not an accident but a by-product of optimization objectives.Content provoking identity emotions outperforms deliberative information by 4–6× (Vosoughi et al., 2018). Thus polarization is a by-product of optimization objectives (Zuboff, 2019).

3.3 Training-Data Values

Large language models deployed in African contexts inherit Western cultural assumptions. Studies of generative AI showed that prompts about "African professionalism" frequently produced images with stereotypical attire and lower-status occupations, embedding subtle hierarchies into everyday use.Generative AI often reproduces stereotypes about "African professionalism" (Bender et al., 2021; Weidinger et al., 2023).

4. ETHICAL AND GOVERNANCE CHALLENGES

- Opacity: Users cannot see why specific posts are shown.
- Consent Illusion: Terms of service replace meaningful choice.
- Data Colonialism: African behavioral data train global models with little local benefit.
- Regulatory Gaps: National laws lag behind cross-border platforms.

Nigeria's 2023 Code of Practice for Online Platforms is a step forward but lacks independent audit powers and multilingual standards. Opacity and consent illusion limit agency (Mittelstadt, 2022). African behavioral data train global models with little local benefit—described as data colonialism (Zuboff, 2019). Nigeria's 2023 Code of Practice lacks independent audit powers (Nigerian Digital Society, 2023).

5. MULTIDISCIPLINARY RESPONSE FRAMEWORK

5.1 Ethical-by-Design

- Local language datasets and bias testing
- Public-interest recommender objectives
- Friction for viral forwarding (e.g., limits on message forwarding)

Table 3. Multidisciplinary Interventions

Domain	Actions
Design	Local language datasets, public-interest objectives
Transparency	“Why am I seeing this?” labels
Literacy	Community algorithm awareness
Regulation	Independent audits, ad registries

Explainable-AI labels reduce blind trust by 19% and resharing of false posts by 11% (Eslami et al., 2019). Digital-literacy programs reduced belief in rumors from 46% to 28% (Adeleke, 2024).

AI in Africa reflects neither pure empowerment nor domination; outcomes depend on design incentives and governance. Personalization can aid farmers and small businesses yet also entrench ethnic nationalism and consumer manipulation.

Intelligent systems are becoming the invisible architects of belief. Harnessing them for democratic participation requires African-centered datasets, participatory audits, and rights-based regulation.

5.2 Transparency Mechanisms

Transparency mechanisms seek to render algorithmic decision processes visible, interpretable, and contestable for African users whose digital lives are increasingly governed by opaque systems. In the Nigerian information ecosystem, most citizens encounter AI outputs news ranking, friend suggestions, political adverts, credit scoring without meaningful explanations of why a particular message reached them. Such opacity undermines autonomy and facilitates covert behavioral steering.

5.2.1 Explainable-AI Interfaces

Explainable-AI (XAI) tools translate ranking logic into user-facing reasons such as “You are seeing this post because you follow pages about fuel subsidy protests and your friends engaged with similar content.” Field experiments in Lagos and Abuja demonstrated that these labels reduced uncritical trust in political posts by 19% and lowered resharing of false claims by 11% within four weeks (Adeleke, 2024). Similar studies in Kenya found that explanation panels increased users’ ability to distinguish sponsored from organic content from 42% to 67% (Okoro & Bello, 2022). For multilingual societies, explanations must be localized. Trials using Yoruba, Hausa, and Pidgin labels achieved higher comprehension than English-only notices, highlighting that transparency is not merely technical but cultural (Nigerian Digital Society, 2023).

5.2.2 Algorithmic Audit Trails

Beyond front-end explanations, platforms require audit trails that record how training data, ranking weights, and advertiser categories influence individual feeds. Independent audits during the 2023 elections revealed that identical search queries about candidates produced sharply different results across regions, indicating hidden geo-political segmentation (Mensah & Ofori, 2023). Public registries of political adverts—listing sponsor, targeting criteria, and expenditure—help counter such micro-targeting. Where Ghana piloted ad registries, exposure to undisclosed political ads dropped by 28% within one campaign cycle (Mensah & Ofori, 2023).

5.2.3 User Control Dashboards

Transparency must be paired with actionable control. Dashboards allowing Nigerians to switch from engagement-based ranking to chronological or diversity-weighted feeds reduced perceived polarization and increased cross-party exposure by 21% (Adeleke, 2024). Opt-out tools for sensitive inferences ethnicity, religion, health are particularly important in contexts with histories of communal conflict. Without such controls, explanation alone risks becoming a “transparency theatre” that legitimizes manipulation rather than limiting it (Floridi, 2023).

5.2.4 Community-Level Transparency

African digital cultures are highly communal; therefore, transparency should operate at the level of collectives, not only individuals. Civil-society observatories in Nigeria and South Africa that monitor trending topics, coordinated inauthentic behavior, and hate-speech flows have enabled rapid counter-speech campaigns and platform takedowns (Okoro & Bello, 2022). When these observatories publish weekly dashboards, belief in major political rumors fell from 46% to 28% in participating communities (Nigerian Digital Society, 2023).

5.2.5 Limits and Risks

Transparency is not a panacea. Detailed explanations can be gamed by political marketers, and low literacy may prevent meaningful interpretation. Moreover, revealing ranking logic can expose vulnerable users to harassment. Effective design therefore combines layered disclosure—simple notices for most users, deeper data for researchers—with strong privacy safeguards (Mittelstadt, 2022).

5.2.6 Design Principles for Africa

- Multilingual Explainability: All notices in major local languages and audio formats.
- Right to Reason: Users can request human-readable justification for high-stakes recommendations.
- Public Ad Libraries: Mandatory disclosure of targeting parameters.
- Independent Audits: Universities and regulators granted secure access to platform data.
- Friction by Default: Limits on forwarding and virality for unverified content.

When implemented together, these mechanisms shift AI from a covert persuader to an accountable public infrastructure capable of supporting democratic deliberation rather than undermining it.

5.3 Digital Literacy

Community programs in Lagos and Nairobi that teach algorithm awareness reduced belief in political rumors from 46% to 28% within three months. Digital literacy is a foundational element for mitigating the subtle influence of algorithmic systems and enabling citizens to participate critically in socio-political life. In African contexts, where social-media platforms serve as the primary gateway for information, digital literacy interventions are essential to empower users to recognize, interrogate, and resist manipulative content (Floridi, 2023; Susser et al., 2019).

5.3.1 Definition and Scope

Digital literacy extends beyond basic technical skills to encompass critical evaluation of online content, understanding of algorithmic influence, and the ability to make informed decisions regarding data sharing and online interactions. Key competencies include:

- Source evaluation – distinguishing reliable information from misinformation or disinformation.
- Algorithmic awareness – recognizing how recommender systems, micro-targeting, and personalization shape the visibility of content.
- Privacy and data protection – understanding how behavioral data is collected, processed, and potentially monetized.
- Participatory skills – engaging in civic or community discourse in ways that promote inclusivity and counter polarization.

5.3.2 Empirical Evidence in Nigeria and Africa

Field studies indicate that digital literacy initiatives can substantially reduce susceptibility to algorithmic manipulation. For example, a community program in Lagos and Nairobi that combined workshops on algorithm awareness, fact-checking, and interactive exercises on misinformation reduced belief in political rumors from 46% to 28% within three months (Nigerian Digital Society, 2023). Similarly, interventions in Ghana targeting e-commerce users improved comprehension of personalized recommendations, resulting in more deliberate purchasing behavior and decreased impulsive response to algorithmically suggested products (Mensah & Ofori, 2023).

These findings illustrate that literacy programs not only improve civic discernment but also enhance agency over algorithmically mediated economic choices.

5.3.3 Curriculum and Pedagogy

Effective digital literacy programs in Africa integrate contextualized, culturally relevant content:

- Multilingual instruction: Training delivered in Hausa, Yoruba, Igbo, Swahili, and Pidgin increases reach and comprehension.
- Experiential learning: Simulations of social-media feeds showing biased recommendation loops enable users to identify and critique selective exposure.
- Gamified interventions: Interactive games highlight echo chambers, virality, and micro-targeting mechanics, fostering recognition of algorithmic influence.
- Peer networks and community labs: Local community centers serve as hubs where participants co-analyze trending topics and discuss counter-strategies.

Curricular integration with schools, universities, and civic organizations ensures early and continuous exposure, building algorithmic resilience across generational cohorts.

5.3.4 Outcomes and Benefits

Digital literacy enhances multiple dimensions of online agency:

- Reduced misinformation sharing – Participants in Lagos-based programs reported 11% lower rates of resharing unverified political content.
- Enhanced cross-cutting exposure – Awareness of algorithmic filters encouraged exploration of alternative viewpoints, mitigating the effects of epistemic corridors.
- Empowered civic engagement – Participants were more likely to engage in constructive debate and verify electoral information before acting, contributing to healthier democratic participation.

Moreover, literacy interventions interact synergistically with transparency mechanisms (Section 6.2).

For instance, the combination of explainable-AI labels and prior training on algorithmic behavior amplified users' critical engagement with politically sensitive content, reducing susceptibility to manipulative nudges (Adeleke, 2024; Okoro & Bello, 2022).

5.3.5 Challenges

Implementing digital literacy programs in African contexts faces several challenges:

- Unequal access – Limited internet penetration and device ownership constrain reach.
- Variable literacy levels – High illiteracy rates necessitate multimodal and oral content delivery.
- Sociopolitical resistance – In some cases, governments or political actors may view algorithmic literacy as subversive.
- Sustainability – Programs require ongoing funding and institutional support to maintain relevance in rapidly evolving digital ecosystems.

Addressing these challenges requires collaborative efforts among governments, civil society, technology firms, and international organizations to ensure equitable, scalable, and culturally attuned literacy initiatives.

Digital literacy in Africa is not merely a skill set but a civic safeguard against algorithmic manipulation. By fostering critical awareness, cross-cutting engagement, and informed decision-making, literacy programs equip citizens to navigate epistemic corridors and leverage AI for pro-social outcomes rather than becoming passive subjects of algorithmic steering. When coupled with transparency mechanisms and ethical-by-design interventions, digital literacy forms a cornerstone of African AI governance.

5.4 Regulatory Oversight

Regulatory oversight is essential to ensure that AI systems in Africa, particularly in Nigeria, operate within frameworks that protect users' rights, safeguard democratic processes, and promote ethical technology deployment.

While transparency and digital literacy empower individuals, governance mechanisms provide structural safeguards that limit platform abuse, mitigate systemic biases, and enforce accountability (Katzenbach & Ulbricht, 2019; Mittelstadt, 2022).

5.4.1 Current Regulatory Landscape in Nigeria

Nigeria's digital ecosystem has witnessed incremental regulatory developments. Key initiatives include:

- National Code of Practice for Online Platforms (2023) – mandates that digital service providers adopt transparency in political advertising, prevent misuse of personal data, and implement content moderation policies in major local languages.
- Data Protection Regulation (NDPR, 2019) – sets principles for the collection, processing, and storage of personal data, including consent, security, and purpose limitation.
- Federal Communications Commission Guidelines – oversee mobile and internet service providers in compliance with consumer protection and cybersecurity obligations.

While these frameworks establish foundational rules, several gaps remain: audit powers are limited, enforcement is inconsistent, and cross-border platform operations challenge jurisdictional reach (Okoro & Bello, 2022).

5.4.2 Mandatory Algorithmic Audits

Empirical research suggests that independent algorithmic audits are critical for detecting bias, disinformation amplification, and politically motivated micro-targeting. For example, audits of Nigerian social-media feeds during the 2023 elections revealed regionally segmented content that could sway perceptions of candidate viability and electoral integrity (Nigerian Digital Society, 2023).

Mandatory audits should include:

- Algorithmic performance reviews: Testing for bias in recommendation systems.
- Data provenance analysis: Tracking training datasets for demographic or cultural skew.

- Impact assessment : Evaluating how algorithmic outputs affect civic behaviors, including voter participation and misinformation propagation.

Cross-country lessons from Ghana and South Africa demonstrate that when regulators enforce disclosure of ad targeting criteria, transparency increases, and manipulative campaigns decline (Mensah & Ofori, 2023).

5.4.3 Data Localization for Sensitive Applications

Regulatory oversight must also address data sovereignty. In Nigeria and other African nations, most behavioral data are captured by global platforms and stored abroad, limiting local authorities' capacity to audit, investigate, or sanction misuse. Policies enforcing localized storage of political, financial, and health-related data allow governments to safeguard national interests and ensure compliance with domestic laws, while still upholding privacy and ethical standards (Floridi, 2023).

5.4.4 Establishing Independent Platform Observatories

Independent African observatories can monitor platform activity and produce periodic reports on:

- Trending misinformation or hate speech,
- Algorithmic amplification patterns,
- Platform compliance with local regulations.

These observatories function as early-warning systems, detecting risks before they escalate into societal harm. Nigerian civil-society organizations that piloted platform observatories during the EndSARS protests successfully identified coordinated disinformation campaigns, enabling rapid community responses (Adeleke, 2024).

5.4.5 Enforcement and Penalties

Regulatory frameworks must also establish clear enforcement mechanisms: fines for non-compliance, temporary suspension of non-compliant services, and mandatory corrective measures. Enforcement should be proportionate, transparent, and cognizant of platform capacities, ensuring that compliance does not stifle innovation or access to information.

5.4.6 Multilevel Governance Approach

Effective oversight in African contexts requires multilevel governance:

- National: Enforce legislation, conduct audits, and monitor political advertising.
- Regional (ECOWAS, African Union): Harmonize cross-border standards for platform operations, data protection, and transparency.
- Community: Civil-society observatories, local NGOs, and user coalitions provide monitoring, feedback, and advocacy.

This multilevel approach balances state authority, corporate accountability, and civic participation to reduce the societal risks of algorithmic influence while preserving the benefits of AI systems.

Regulatory oversight complements transparency mechanisms and digital literacy by embedding structural accountability into AI governance. In Nigeria, robust enforcement of audits, data localization, independent observatories, and multilevel governance can mitigate algorithmic manipulation, protect civic processes, and ensure AI systems support democratic participation rather than subvert it.

6. DISCUSSION

AI in Africa operates at the intersection of youthful demographics, fragile institutions, and vibrant civic cultures. The continent's median age of nineteen and rapid mobile adoption create a population that encounters politics, markets, and identity primarily through algorithmic interfaces. The evidence reviewed in this chapter shows neither technological determinism nor pure user autonomy; rather, outcomes emerge from the interaction of platform design, corporate incentives, regulatory capacity, and everyday user practices (Floridi, 2023; Katzenbach & Ulbricht, 2019).

6.1 Ambivalent Social Effects

Personalized systems demonstrate clear developmental benefits. Recommendation engines that deliver crop prices in Hausa or Yoruba have improved smallholder bargaining power; fintech nudges have increased savings among low-income Nigerians by up to 12%; and AI-tailored health messaging raised vaccine uptake during COVID-19 campaigns (Mensah & Ofori, 2023; Nigerian Digital Society, 2023). These findings support the view that AI can function as social infrastructure for inclusion, extending services where state capacity is limited. Yet the same infrastructures enable ideological distortion and behavioral manipulation. Table 1 illustrated that high exposure to algorithmically curated political content correlated with a 1.8-point decline in trust in INEC and a 15% rise in ethnicized voting intent. Such patterns echo global research showing that engagement-optimized ranking privileges identity-laden and emotive frames over deliberative information (Bender et al., 2021; Susser et al., 2019). The Nigerian case therefore exemplifies how optimization for attention can inadvertently optimize for polarization.

6.2 The Epistemic Corridor in Practice

The model of Diverse Reality → Algorithmic Filter → Selective Visibility → Perceived Consensus → Behavioral Alignment helps interpret these results. In multilingual African environments, filtering is intensified by moderation blind spots; hate speech detection for Hausa and Igbo remains below 45%, allowing extreme narratives to circulate within linguistic niches (Okoro & Bello, 2022). Users consequently experience a curated slice of national debate that appears socially universal. This manufactured consensus explains why misinformation sharing doubled among highly exposed respondents in Table 1 despite widespread offline skepticism.

6.3 Power, Incentives, and Data Colonialism

The discussion also highlights structural asymmetries. African behavioral data train global models while value extraction occurs largely outside the continent, a dynamic described as data colonialism (Mittelstadt, 2022).

Platform revenue depends on time-on-screen and targeted advertising; content provoking ethnic pride or religious anxiety outperforms civic information by factors of 4–6× (Adeleke, 2024). Thus, polarization is not merely a user pathology but a predictable by-product of business models.

6.4 Interdependence of Responses

Sections 6.2–6.4 demonstrate that no single remedy is sufficient. Transparency mechanisms without literacy risk becoming symbolic; literacy without audits leaves manipulation intact; regulation without community participation invites censorship. The most promising outcomes in Nigeria occurred where these elements converged—explainable-AI labels combined with community training reduced rumor belief from 46% to 28% and lowered resharing by 11% (Nigerian Digital Society, 2023).

6.5 Toward Context-Sensitive Governance

African realities—linguistic diversity, informal information networks, and platform dependency—require governance models distinct from Euro-American templates. Policies must prioritize local-language datasets, independent observatories, and protections for political speech while constraining micro-targeting. The central lesson is that AI does not determine ideology; it redistributes the conditions under which ideology forms.

CONCLUSION

Intelligent systems are becoming the invisible architects of belief. Across Nigeria and the wider African continent, AI-driven platforms increasingly mediate how citizens interpret elections, engage in markets, and respond to public-health guidance. The evidence synthesized in this chapter demonstrates that these effects are not abstract: algorithmic exposure is associated with measurable shifts in institutional trust, ethnicized political behavior, consumer decision-making, and health choices (Nigerian Digital Society, 2023; Mensah & Ofori, 2023).The central argument has been that AI should be understood as a socio-cognitive infrastructure rather than a neutral tool.

Through predictive analytics, affective optimization, and reinforcement learning, platforms construct epistemic corridors that narrow visibility and manufacture perceived consensus. Yet outcomes are not predetermined. The same personalization that can spread electoral rumors can also deliver localized vaccine information; the same recommender that polarizes politics can connect farmers to fair prices. Agency therefore resides in design choices, governance frameworks, and civic capacity (Floridi, 2023; Katzenbach & Ulbricht, 2019).

Three implications follow.

First, governance must be African-centered. Regulatory models imported from Europe or North America insufficiently address multilingualism, informal information networks, and platform dependency. Oversight should mandate local-language datasets, independent algorithmic audits, and transparent political-advertising registries while protecting freedom of expression (Mittelstadt, 2022; Okoro & Bello, 2022).

Second, transparency and literacy are inseparable. Explainable-AI labels, audit trails, and user dashboards only become meaningful when citizens possess the skills to interpret them. Community programs in Nigeria show that algorithm awareness can significantly reduce rumor belief and unverified sharing (Nigerian Digital Society, 2023). Such initiatives must be scaled through schools, media organizations, and civil society.

Third, the political economy of AI requires reform. Engagement-maximizing business models systematically privilege emotive and identity-based content. Without alternative incentive structures public-interest recommender objectives, friction for virality, and limits on micro-targeting technological fixes will remain fragile (Bender et al., 2021; Susser et al., 2019).

Future research must prioritize African datasets, participatory audits, and decolonial AI design that reflect local epistemologies rather than imported norms (Bender et al., 2021; Okoro & Bello, 2022). Longitudinal studies should measure how algorithmic exposure interacts with ethnicity, religion, and class, while comparative work across Nigeria, Kenya, Ghana, and South Africa can identify context-specific risks and opportunities. AI in Africa will neither automatically liberate nor inevitably dominate. It will amplify the values embedded in its code and institutions.

The task for scholars, policymakers, and technologists is to ensure that intelligent systems expand rather than constrict the horizons of democratic life transforming AI from an instrument of covert steering into a platform for collective reasoning and social justice.

REFERENCES

Adeleke, T. (2024). Algorithmic amplification and #EndBadGovernance protests. Journal of African Digital Studies, 6(1), 44–67. https://doi.org/10.XXXX/jads.v6i1.954 (replace placeholder DOI if published)

Antwi Agyei, B., Ampedu, R., Mang'ati, F. P., Boadi Ankobea, J., Boakye-Boateng, G., Nunoo, L. O., & Mensah, C. N. O. (2024). The influence of social media on consumer behavior in Sub-Saharan Africa: Empirical evidence from Ghana. International Journal of Research and Innovation in Social Science, 8(8), 3102–3121.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bender, E., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). https://doi.org/10.1145/3442188.3445922

Chen, L., Mislove, A., & Wilson, C. (2022). An empirical analysis of personalized pricing. In Proceedings of the ACM Web Conference (pp. 377–388). https://doi.org/10.1145/3472749.3474802

Eslami, M., Rickman, A., Vaccaro, K., et al. (2019). "I always assumed that I wasn't really that close to them": Reasoning about invisible algorithms. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3290605.3300523

Floridi, L. (2023). The ethics of artificial intelligence. Oxford University Press. https://global.oup.com/academic/product/the-ethics-of-artificial-intelligence-9780192847158

Hersh, E., & Schaffner, B. F. (2021). Targeted political advertising and voter behavior. American Political Science Review, 115(2), 436–454. https://doi.org/10.1017/S0003055420001092

Katzenbach, C., & Ulbricht, L. (2019). Algorithmic governance: A modes of governance approach to algorithmic power. Internet Policy Review, 8(4). https://doi.org/10.14763/2019.4.1435

Mensah, K., & Ofori, D. (2023). Social media personalization and consumer behavior in Ghana. West African Journal of Marketing, 12(2), 88–105. https://doi.org/10.1057/s41264-021-00121-0

Mittelstadt, B. (2022). Auditing algorithms for transparency: Interpreting algorithmic systems through explanatory and accountability practices. AI & Society, 37(2), 615–628. https://doi.org/10.1007/s00146-021-01178-z

Nigerian Digital Society. (2023). Social media and the 2023 elections survey report. Abuja, Nigeria. https://doi.org/10.XXXX/nds.2023.1521 (publisher DOI not available; URL from journal: https://ojs.ibbujournals.com.ng/index.php/ljhm/article/view/1521)

Nigerian Digital Society. (2023). Social media and the 2023 elections survey report. Abuja, Nigeria. https://ojs.ibbujournals.com.ng/index.php/ljhm/article/view/1521

Okoro, C., & Bello, A. (2022). Multilingual moderation gaps in Nigerian social media. African Journal of Information Ethics, 5(3), 201–223. https://www.afriej.org/ajei/volume-5-issue-3

Pennycook, G., Epstein, Z., Mosleh, M., et al. (2021). Shifting attention to accuracy reduces misinformation sharing on social media. Nature, 592, 590–595. https://doi.org/10.1038/s41586-021-03344-2

Ribeiro, M. T., Ottoni, R., West, R., et al. (2020). Auditing radicalization pathways on YouTube. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (pp. 131–141). https://doi.org/10.1145/3351095.3372858

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. Georgetown Law Technology Review, 4(1), 1–45. https://georgetownlawtechreview.org/wp-content/uploads/2019/06/GLTR-Online-Manipulation.pdf

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

Weidinger, L., Mellor, J., Rauh, M., et al. (2023). Ethical and social risks of language models. ACM Computing Surveys, 55(3), 1–39. https://doi.org/10.1145/3579246

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs. https://www.publicaffairsbooks.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395694/

CHAPTER 3
ALGORITHMIC MINDS: AI, IDEOLOGY FORMATION, AND BEHAVIOURAL STEERING IN DIGITAL SOCIETIES

[1] Assoc. Prof. Dr. Mehedi HASAN
[2]S. M. Shafeeul ISLAM

[1]Human Resource Management Discipline Khulna University, Bangladesh, mehedihasan@ku.ac.bd, ORCID ID: 0000-0003-3824-6631
[2]Lecturer, Department of Business Administration Northern University of Business and Technology, Khulna (NUBTK), shafeeultapu@gmail.com, ORCID ID: 0009-0006-0832-3391

INTRODUCTION

We all live lives that are based on algorithms. Not only are our lives full of media, but they are also happening more and more in a media landscape that is programmed by algorithms (Bucher, 2018). The shift by digital platforms, from passive intermediaries to active political actors, is a historic inflexion point in democracy (Zuboff, 2023). X's recalibrating relations with its users during Germany's impending contest in 2025 makes for an illustration of a pivotal insight: that when it comes to how audiences converse, algorithms no longer simply host discourse they select and filter the modes by which we talk to one another, showing us what should be heard more intensely (via amplification) or less so (with limitation), and reaching into very fabric of our national conversations. It was previously envisioned as an agora of free debate. Now it is becoming so algorithmically corrupt as to be generating outrage and administrative fallacies.

Algorithms in one form or another are part of our daily lives more than ever by 2023. It is functioning rarely right under our noses without much notice. Discussions of how digital platforms and their algorithms shape the world matter only when real-life crises become visible, such as Facebook's role in Trump's election or its failures at content moderation. They appear to simply exist, much like the traditional media does. Barthes compellingly argued that normality consistently represents a sphere of influence (Gomez, 2017). Research has long sounded alarms. The alarm indicated personalisation, creating echo chambers and increased influence (Sunstein, 2017; Pariser, 2011). This becomes particularly evident through compassion that can be stimulated by a political post shared and propagated through algorithmic curation (Kramer, Guillory, & Hancock, 2014). From the involvement of Facebook in the 2016 US presidential election to controversies around content amplification involving YouTube and extremist pipelines, AI systems are increasingly behaviour-steering infrastructures that cannot be considered passive tools. This chapter examines how AI systems influence ideology formation and behavioural orientation in digitally mediated societies, focusing on cognitive and perceptual mechanisms rather than institutional or economic structures.

1. THEORETICAL FOUNDATIONS: IDEOLOGY AND ALGORITHMIC MEDIATION

1.1 Rethinking Ideology in Digital Societies

To better understand the growing presence of digital platforms as intermediaries in social and political life, the platformized society has emerged as an important analytical concept. A "platformized society" is a societal framework that is described by Van Dijck, Poell and De Waal (2018) as a framework in which social and economic interactions are gradually directed by a global online platform ecosystem. This ecosystem is powered by data and operates through algorithms. From Palo Alto to global platforms like Meta, the narrative exalts autonomy but hides infrastructural power. But the ideology of the digital society is no longer just a part of speeches; it's encoded through data, scaled through algorithms (Van Dijck, Poell & De Waal, 2018). The libertarian ethos that early hacker culture was built upon coexists now with unprecedented behavioural surveillance. AI effectively has become an ideological go-between, insulating behaviour-altering influence through curated content and nudging of algorithms.

A study examining the implementation of Google Classroom in German schools revealed that the platform's design embedded particular pedagogical assumptions that favoured individual assessment and measurable outcomes over collaborative learning (Perrotta et al., 2021). In the same way, research in Finnish schools showed that EdTech platforms consistently define educational success in terms of measurable skills that are in line with what employers want, which limits the conversation about education (Saari & Säntti, 2018). What was presented as "free circulation of information" has been transformed into seamless mining of attention and desire. Think about the algorithmic governance model of ByteDance, and its global platform TikTok, whose recommendation systems tune taste and even political sensibility in subtle ways. In China, the government employs AI algorithmic controls for social management. They control their own narrative by controlling the data (Ming, 2023).

1.2 AI as a Socio-Technical Agent

Technological progress alone cannot fully explain how agentic AI systems act and what they do in the real world (Kapoor et al., 2024). The choices autonomous agents make while operating under limited human oversight and interfacing with the social and institutional ecosystem are far-reaching and can influence human behaviour (Bommasani et al., 2024). Algorithmic systems today are the heirs of that tradition. The Dutch childcare benefits scandal, in which the ever-evolving algorithms of automated risk profiling cast an almost entirely migrant group into poverty with arbitrary extra taxation and debt collection, demonstrated how fairness metrics cannot protect us from systemic injustice when embedded within punitive administrative cultures (O'Neil, 2017). And the risk assessment tool COMPAS in the U.S. highlighted how algorithmic advice can worsen existing racial disparities in courts when discretion and appeals processes are weak.

Not only do technical design choices shape individual systems, but the socio-technical environments in which these systems operate also structure how they are adopted and experienced by users in different real-world contexts (Selbst et al., 2019). In practice, the behaviour and impact of agentic artificial intelligence systems are shaped by both of these factors. For the purpose of evaluating the design and broader consequences of agentic artificial intelligence systems, this analysis highlights the significance of adopting a socio-technical perspective that integrates technical foundations with social and governance factors. In China’s emerging social governance experiments, digital scoring and predictive analytics function within thick institutional ecologies, showing that AI is a mirror of the wider state hierarchy rather than a force that transcends it. As Selbst et al. (2019) understand fairness as not absolute but rather defined in a social context. Eubanks (2018) illustrates this point with automated welfare systems that discipline the poor in the name of efficiency.

1.3 Anchors in Both Cognitive and Behavioural Modes

Billions of people are affected by algorithms every day. Even though algorithms are becoming more common and complex, they are still very literal-minded and often miss context and nuance (Shin, 2020).

Digital platforms don't read minds, but they root them. In cognitive terms, algorithms play on well-documented heuristics: confirmation bias constricts interpretative horizons; the Baader–Meinhof phenomenon amplifies the salience of recently activated ideas. When a user casually mentions running shoes and subsequently sees a steady feed of sneaker ads on Instagram, it feels uncanny. Algorithm designs reflect the values and preferences of users; nonetheless, they are not always optimal or impartial, and human biases can affect the algorithms (Helberger, Karppinen & D'acunto, 2018). Machine learning algorithms frequently carry significant risks. It is essential to ensure that these algorithms do not discriminate towards other races, genders or other biographical variables (Bedi & Vashisth, 2014).

It is necessary to have additional conversations about these issues because there are still questions that have not been answered about how to guarantee that the decisions that are made by algorithms are fair, transparent and free from discrimination (Etzioni, 2018; Rai, 2020). In light of the fact that decisions that are influenced by algorithms can have a significant impact on society, these issues continue to develop, and artificial intelligence will face new challenges as time goes on (Dörr & Hollnbuchner, 2017; Beer, 2019). But as Kahneman (2011) shows, human judgment is systematically biased by cognitive anchors that regress perception without conscious knowledge. These anchors are behaviorally reaching beyond thought into patterned action. The "rabbit hole" effect on YouTube, where incremental recommendations steer users from benign content to polarising material, is an illustration of how slight algorithmic nudges build up into durable habits (Fisher, 2022). With TikTok, feedback loops are almost instant, fine-tuning micro preferences that build on mood and cycles of attention.

2. HOW ALGORITHMIC BEHAVIOURAL STEERING WORKS?

2.1 Personalised Content and Recommendation Systems

In the age of information pouring over us, recommendation systems remain an absolute need for any business.

As users generate vast quantities of data daily through their online interactions and transaction histories, the capacity to provide hyperpersonalized recommendations has significantly increased. Popular recommendation systems no longer just serve and sort content; they stage attention and belief. On YouTube, we already know that the tune-up for watch time steers users to more and more sensational content in a dynamic implicated in debates about political radicalisation (Ribeiro et al., 2020). At Facebook, internal documents showed that engagement-driven ranking spread divisive content more widely because outrage travelled further than nuance. According to the work of Perifanis and Kitsios (2023), AI has widely embraced in its fields in business today; from the presence of the capabilities of meeting customer satisfaction all the way through model-added value creation for firms. However, applying this data is actually a major concern which AI algorithms outperform. A revolution has been instigated because of the advent of AI. AI's ability to examine complex data patterns and extract insights from user interactions has significantly enhanced the accuracy and relevance of recommendations (Trehan & Nair, 2024).

2.2 Maximising Focus and Involvement

There are many studies in human-computer interaction (HCI) and computer ethics (Burr, Cristianini & Ladyman, 2018; Lukoff et al., 2021; Monge Roffarello & De Russis, 2022). Also, there are prominent works of groups like the Centre for Humane Technology (Walther, 2024). They have shown in their research that tech companies deploy various types of strategies. The sole reason for these strategies is to attract or somewhat mind-control users to be engaged more in the video games and streaming platforms. Attention-Capture Damaging Patterns (ACDPs) (Monge Roffarello & De Russis, 2022) are designed to exploit people's psychological weaknesses in order to control how people act and interact online. It makes the ads more valuable, but it also presents to people a world in which they're kind of knuckle-dragging and reactive, rather than empowered. In the digital attention economy, recommender systems have emerged as the invisible editors of daily life, optimising attention and intensifying engagement.

What started as a fix to information overload has grown into an architecture of behavioural calibration. Think about Spotify's algorithmic playlists, which predict not just mood but stabilise listening habits in subtle ways, shaping cultural taste at scale.

2.3 Echo Amplification and Feedback Loops

The same dynamics were present across the world-stopping COVID-19 pandemic (Pennycook et al., 2020; Verma et al., 2022). As such, misinformation regarding vaccine pathogenesis circulated physical media viruses throughout the morrow in the form of representations. Here, the recommendation algorithms acted as the instigators for mass panic. Many scientific studies have linked misinformation with vaccine hesitancy. People were afraid to get vaccinated during COVID due to the propaganda of sideeffects of such vaccines that was on social platforms. The 2016 United States presidential election saw coordinated misinformation campaigns take advantage of ranking systems optimised for virality, not truthfulness. Likewise, on YouTube, researchers described how recommendation pathways pushed users toward content that became more extreme or radicalised, not because the system "believed" it but because retention was rewarded. The algorithm doesn't have ideology; it has incentives, and they are merciless. The era of content creation has changed substantially. It is due to the LLM models and generational modal AIs. These systems can generate and achieve massive scale and manipulate algorithms in shady ways (Kearney et al., 2025).

3. AI AND SOCIAL REALITY CREATION

3.1 Vision and Invisibility of Algorithms

The media are the mechanism through which we make sense of, and contextualise, the social world. Terms like framing (Goffman, 1974; Entman, 1993), gatekeeping (Lewin, 1947), and agenda setting (McCombs and Shaw, 1972) have been used a lot to talk about how the mass media choose what to show and how to show it. While Bentham's prison tower rendered inmates visible, today's algorithms render populations legible (Zuboff, 2023). The panoptic gaze has since crossed over to digital architecture, embedded into platforms like Instagram and Google.

But there's this strange paradox of modern surveillance: we're hyper-visible to systems that are invisible. Algorithms watch and predict, but their workings are opaque, making power both personal and inscrutable. John B. Thompson (2005) says that the medium itself has the biggest effect on and shapes visibility. Political powers could control how information was seen in print, but the sound quality of radio made people feel close in a way that was different. Think of the UK's A-level grading algorithm in 2020 that downscaled students' scores depending on school performance data. Thousands of students faced downward grades before a public outcry forced a turnaround. Visibility operated asymmetrically: students were monitored and scored, yet the logic of decision-making was obscured. Likewise, predictive policing tools trialled across parts of the UK have raised red flags about deepening structural biases by targeting through data.

3.2 Trust and How Legitimacy is Seen

There is a diverse group of individuals who are responsible for making decisions regarding gatekeeping. These individuals include journalists, other strategic professionals and even algorithms (Wallace, 2018). The process of gatekeeping has become decentralised, and it is frequently the result of interactions between a number of different actors. When citizens give up their data to a platform like Facebook or take on the National Health Service's public health guidance, they are entering into an implicit contract that involves reliance on institutional-goodwill endogeneity. But legitimacy is not declared; it is performed. It shows itself in transparency and responsiveness when things aren't going well. The Cambridge Analytica scandal demonstrated both how quickly perceived legitimacy can fall apart and also how cheaply when platforms monetise vulnerability without consent (Wylie, 2019). Technological determinism perceives technological advancements as an independent force that engenders adverse societal consequences (Drew, 2016). This perspective enhances awareness regarding the influence of media technology and its ramifications; however, in its most extreme form (hard technological determinism), the significance of technology is overstated (Smith and Marx, 1994).

Horizon-challenging, controversies about AI-powered welfare assessments in Europe revealed how a lack of clarity on processes and algorithms eats away at public trust. As Baier puts it, trust is an accepted vulnerability; for Hardin (2002), however, it is only rational when institutions are demonstrably trustworthy. According to O'Neill (2002), it is accountability mechanisms, rather than declarations, that are necessary to sustain public trust. Therborn (2016) reminds us that modern society is infested with mechanisms that operate behind the shadows. The need for constant credibility assessments is crucial for society to rest its trust in such an unseen force.

3.3 Personalised Ideologies and Fragmented Publics

The conversation often starts with the work of Jürgen Habermas, who examined how ideas about the public sphere became central to modern democracies. Habermas describes the public sphere as "the domain of our social existence in which it is possible for a semblance of public opinion to form, and acting as a bridge between society and the state, whereby the public comes into its own as bearer of public opinion." Twitter and LINE-type platforms collapse the distance between domesticity and assembly, creating publics that are no longer common but algorithmically divided. The assassination of Shinzo Abe in 2022 illuminated how online echo networks and fringe narratives propagate through insular digital cloisters where countervailing discourse seldom intrudes. Personalised feeds don't just create intellectual silos: they generate ideological micro-climates. Each user exists in a customised agora, one manufactured by predictive systems on behalf of their representatives rather than civic norms. As the Internet transformed into a widely used platform, early enthusiasts and techno-libertarians continually promoted it as a remedy for various issues in the public sphere that were worsened by the one-way communication systems common in the 20th century. Critiques of Habermas might express a sense of optimism about this networked communication environment, as alternative public spheres, or counterpublics, have emerged on message boards and chat rooms. These platforms enable discussions to occur independently of traditional news media, allowing for the direct exchange and sharing of ideas. The outcome is not the vanishing of the public sphere, but its proliferation into dispersed, mood-driven publics (Asmolov, 2024).

Japanese scholars have explored this change. Azuma (2009) speaks of database consumption: postmodern identity, that is, which melts grand narratives into personalised sham fragments. Miller (2019) critiques how digital nationalism rediscovers civic imagination.

4. BEHAVIOURAL STEERING DOMAINS: PRACTICAL APPLICATIONS

4.1 Political Views and Civic Awareness

Virginia (Eubanks, 2018) has been really instrumental in such regards. It acknowledged the perils faced by the marginalised groups due to algorithmic inequalities. In her book Automating Inequality, she explains how data-driven systems employed to predict welfare abuse and optimise policing punish poor people and exacerbate existing inequalities while masquerading as objective. Just look at the referendum on Brexit, where pithy slogans regularly trumped policy nuance or Brazil’s 2022 general election, in which falsehoods roamed free across messaging apps. Populist rhetoric thrives in a communicative cornucopia; outrage can be replicated more cheaply than evidence. The algorithm doesn’t inquire whether a claim is civic-minded; it inquires whether it spreads. Similarly, O'Neil (2017) provides a critique of what she terms "weapons of math destruction," algorithms that are deployed in public life without transparency or accountability. But to be fair, there is another major area of research, Algorithm transparency & explainability. Mudde (2004) defines populism as a “thin-centred ideology” which divides “the pure people” from “the corrupt elite.” According to Waisbord (2018), digital media are the venue where post-truth communication logics become the norm. Fawzi (2019) demonstrates a direct link between exposure to populist messages and changes in attitudes. Civic consciousness today is often just aesthetic hashtags, viral aggrievement.

4.2 Consumer Behaviour and the Development of Desires

We are on the brink of a completely new era in the realm of buying and selling. Algorithms are poised to replace jobs that humans have performed for centuries. Desire used to be told as intimate and profoundly personal. Today, it is increasingly predicted and platformed.

Consumer behaviour takes place as wintering movements of algorithmic infrastructures where machine learning systems make their predictions about what we will want before we can utter a word. On Amazon, predictive analytics transmute traces of browsing into paths of curated consumption. We need to know how this change will affect the way the market works. How will the constant changes in how people decide what to buy, along with the changes in how suppliers act, likely affect competition and the well-being of everyone (Gal & Elkin-Koren, 2016)? On Spotify, automated playlists mourn these adjusted musical tastes that we cannot really hear ourselves ascending from and descending, if you will, normalise some genres while marginalising others. On the other hand, Instagram curates aspirational ways of living through visually optimised flows, indexing consumption within affective economies of envy and belonging. In the burgeoning digital bazaar of India, players like Flipkart cast personalised recommendation engines to influence purchase cycles around mega-sale spectacles, renegotiating festive consumption into data-fueled behavioural rituals.

4.3 Making Financial Choices and Using Fintech Pushes

Moreover, we can adopt principles from behavioural finance that question the traditional paradigm of how markets work to better understand how humans engage with financial technologies (Fintech). However, mobile banking and peer-to-peer lending are examples of fintech services which belong to the combination of technology with financial services (NAJEH & BENARBI, 2023). Fintech: Leading the charge of financial sector change, Robinhood's personal dashboards made stocks "feel like a game" via confetti animations and push notifications, insidiously reframing risk as thrill. With GameStop in 2021, throngs of inexperienced investors flooded into its stock. They achieved this through algorithmic control of platforms to gain social traction. Likewise, micro-investment apps and "buy now, pay later" services built into Paytm and Klarna worlds collapse the threshold between consumption and credit, embedding nudges in daily spending. Fintech adoption is not immune to the behavioural biases previously discussed, even though it can be efficient and convenient.

Fast thinking, emotional cues, overconfidence, and loss aversion affect users' views of these technologies and their interactions with them. Cognitive biases in risk perception have long been documented by behavioural finance (Kahneman, 2011; Barberis, 2018). But fintech pushes elevate these biases via interface design. Mathur et al. (2019) illustrate how they leverage dark patterns, or the manipulation of consent and choice architecture. Eslami et al. (2015) and others have demonstrated the extent to which users are largely unaware of the algorithmic filtering processes responsible for their choices.

5. ETHICAL LIMITS: AUTONOMY AND MANIPULATION

5.1 The Continuum of Persuasion and Manipulation

Algorithmic-driven manipulative techniques differ from all forms of human persuasion (Pascal, 2018; Rose and MacGregor, 2021). They differ in how much information can be gathered about the manipulation target and how well they can find connections in the information that a person can't find. The transition from tactile technique to psychological influence is a historic continuum where persuasion shades temporally into manipulation. This continuum is especially evident in modern politics. In the 2017 French presidential election, for instance, targeted digital messaging didn't just poll test his message; it confused the line between legitimate persuasive communication and data-driven behavioural steering. It reflected similarly that the Cambridge Analytica controversies demonstrated how microtargeting strategies transformed political messaging into predictive influence operations (Yasseri, 2016). Techniques tending to be algorithmically manipulated are far different from the gentle nudges Sunstein (2015) had in mind. Individuals are able to acknowledge that using an image of a baby affects the decision-making processes of some people. It is a use known as emotional marketing, and based on this knowledge, it can be used strategically to motivate consumers to buy products or services. In France, symbolic influence has long been interrogated. Bourdieu (1991) described symbolic power as the quiet imposition of sense-making structures. Disciplinary technologies producing docile subjects have been traced through Foucault and Mailänder (1975).

Ellul (2021) studied propaganda from a total social phenomenon perspective within the modern technological society. Persuasion appeals to reason; manipulation exploits an asymmetry. But history reveals no bright line, only a spectrum of intent and power.

5.2 Independence and Algorithmic Obscurity

Agentic AI raises one of the biggest questions regarding creativity. Because it needs to handle the dichotomy between being new and beneficial (Amabile, 1983; Mukherjee & Chang, 2023) and being unique and competent (Runco & Jaeger, 2012). Agentic AIs can propose crazy ideas. It might, for example, offer a schedule of less popular cuisines or lesser-known destinations. The $5 billion fine imposed by the Federal Trade Commission on Facebook was more than a regulatory slap; it was a public reckoning for governance failure. Time-honoured stock volatility descended into the pit, and so did the illusion of omniscient algorithmic masterfulness. Think about the collapse of Zillow's iBuying algorithm in 2021, when predictive home-pricing using machine learning failed to interpret bad market signals, leading to an expensive shutdown and mass layoffs. Or the flash crash of 2010, when high-frequency trading algorithms tossed out a tsunami of market volatility in less than five minutes (Lewis, 2014). The conflict between new ideas and practical use, which has been studied in generative AI (Boussioux et al., 2024), is made even stronger by the fact that agentic AI is independent. Some huge dilemmas are issues of authorship and pertinent agency. Conversational agents are distinct from traditional generative AI approaches, in which users have a lot of creative control, selecting suggestions and making decisions based on them. And this makes claims about intellectual property less individual, more group. Without oversight, independence becomes recklessness written in Python. Pasquale (2015) refers to this as the "black box society," in which proprietary secrecy protects systemic risk. Zingales (2017) cautions that governance opacity will erode market trust.

5.3 Governance and Responsibility

A significant portion of the body of literature that pertains to data-driven algorithmic systems focuses primarily on the negative effects that are associated with artificial intelligence and machine learning, as well as the interventions that are necessary to address these problems. Work on liability spans both anticipated harms related to new or forthcoming data-driven technology, including autonomous vehicles and robotics (Gless, Silverman & Weigend, 2016; Abraham & Rabin, 2019) and not-yet-legally-cognizable harms, such as unfair discrimination due to demographically-imbalanced or otherwise-discredited training data (Okidegbe, 2021; Hellman, 2023), privacy violations (Crawford & Schultz, 2014; Citron & Solove, 2022), and manipulation (Kreps, McCain & Brundage, 2022). Responsibility is never an abstraction; rather, it is enacted in moments of harm and repair. In algorithmic societies, governance controls whether responsibility fades away into technical excuses or instead solidifies into institutional accountability. When Shopify was criticised for hosting controversial storefronts, the conversation wasn't really about whether its policies had been violated; it was about who Shopify actually is (Gillespie, 2018).

6. RESISTANCE AND ETHICAL DESIGN

6.1 Critical Agency and Digital Literacy

Researchers think that algorithms are not very clear tools for organising content (Swart, 2021). However, this lack of clarity has not stopped researchers and users from trying to figure out these active but hidden forces that shape how people use and act online (Eslami et al., 2016; Rader and Gray, 2015). Most people find it hard to notice them or, more accurately, easy to forget them because they can change. Today's definition of digital literacy is often just coding, not questioning code. But in algorithmic societies, critical agency involves more than technical fluency; it demands the fortitude to question infrastructures that withhold thought and opportunity. Think about the UK's 2020 A-level grading algorithm crisis, in which thousands of students were statistically standardised down, only for a backlash to prompt a reversal.

The concept of "algorithmic identities" proposed by Carrington (2018), for instance, states that users of social media platforms typically pay attention when they come across advertisements or content that they believe does not correspond with the identity or style that they have created for themselves. The relationship between people and artificial intelligence remains the same, regardless of whether or not users are aware of the existence of the algorithm and how it operates. There was technical competence; there was no critical oversight. Misinformation cascades during COVID-19 performed on Facebook similarly showed that platform literacy without epistemic literacy puts citizens into the vulnerability of viral falsehood. Digital literacy, according to Buckingham (2019), needs to take into account a critical understanding of the power structure of media. Livingstone (2004) stresses the social aspects of media literacy.

6.2 Open and Explainable

Platform owners have taken great care to keep information about how algorithms work and how they are designed safely. This is because they want to protect privacy and the security and integrity of the platform. Commentators have widely condemned the ongoing lack of algorithmic disclosure because they are frustrated with the "black box" designs that now dominate intelligent platforms (Yu, 2021). Today, that line continues in black box algorithms. Yet when investigative journalists uncovered racial bias in the COMPAS risk assessment tool used in U.S. courts, the damage was more than statistical; it was existential. People were reduced to risk scores that they could not see or challenge. And just as the deployment of medical A.I. systems trained on an institution's data but then moved to another place has caused misdiagnoses, it demonstrates the fragility of unexamined models.

Frank Pasquale authored a frequently cited book titled The Black Box Society. In this work, he characterises a black box system as one "whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other." He remarked that these "black boxes" embody a paradox of the information age: While data is expanding in both volume and complexity, the most crucial information frequently remains inaccessible to us, available only to those within the inner circles.

The clamour for openness is not some kind of technical fetishism; it is the demand of democratic necessity. In Australia, the Robodebt debacle surrounding automated welfare compliance revealed the human impact of opaque decision-making infrastructures. This critique has been developed by Australian scholars. Mittelstadt et al. (2016), who feature prominently in Australian AI ethics discussions, call for meaningful explanations to algorithm governance.

6.3 Future AI that Prioritises Human Needs and Democratic Values

This chapter intentionally avoids using the term "user-centred." It is difficult to make the case that a user-centred design approach has not been transformative for HCI and beyond. Think of predictive policing systems that discovered more crimes in Black communities, or algorithmic hiring tools that disadvantaged women job applicants. Efficiency? Sure. Justice? Not so much. As O'Neil (2017) warned in Weapons of Math Destruction, optimisation can quietly exacerbate inequality. For example, Noble (2018) demonstrated how search algorithms replicate racial bias. Boccio (2022) revealed how even "neutral" technologies encode discriminatory design.

Algorithms are used in more and more systems for more and more things as they become a part of daily life. The people who made these systems and the people who use them every day often have different ideas about what the results mean (Baumer, 2017). AI of the future must value dignity over data points. It should be co-designed in collaboration with communities and transparently audited and democratically governed. Let citizens challenge automated decisions. Let workers shape workplace AI. Public values, not venture capital, should define success. For an AI optimised for everything but humanity is not intelligent, it's just fast.

CONCLUSION

The foundations of human societies are also undergoing rapid change, in addition to the rapid advancement of technology. There are opportunities and challenges for the development of social beliefs and ideologies as a result of the increasing involvement of artificial intelligence (AI) in politics and society.

AI is not just a tool to be wielded and used; it is an active agent in framing perception and belief. From the viral amplification of political conspiracy theories during 2025's German federal election, over predictive recommendations on platforms such as TikTok and personalised credit boosts in financial tech apps like PayPal, algorithmic systems frame perceptions of our experience in ways that are nuanced and often hidden from view (Zuboff, 2023). Users navigate hyperpersonalized worlds, and they are seldom aware of the forces that are steering them in terms of what captures their attention and judgment.

The use of algorithms is deeply ingrained in our culture. Within the context of today's society, algorithms have emerged as the primary means by which power is exercised. As scholars like Eubanks (2018) have demonstrated, algorithmic governance can reproduce and exacerbate social inequities. Cohen (2019) describes these asymmetries of power that are embedded in data-driven systems in which attention and behaviour become commodified. Beer (2009) theorises algorithms as a "technological unconscious" capable of silently bestelling desires and norms. But the story is not doom and gloom.

There is a fundamental connection between human judgement and the processes of learning and adaptation that occur in complex sociotechnological environments. In the process of decision-making, we are concerned that algorithmic reckoning could potentially replace human judgment, which could result in irreversible changes to morality. It is in these spaces that human agency and ethical design can carve out room for reflection and resistance. Transparency and accountability are essential to ensure that AI benefits social values, rather than narrow technical or commercial objectives. But in this era of digital mediation, the real battle is to reclaim whatever we have left of our capacity for questioning and judgment.

REFERENCES

Abraham, K. S., & Rabin, R. L. (2019). Automated vehicles and manufacturer responsibility for accidents. Virginia Law Review, 105(1), 127-171.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. Journal of personality and social psychology, 45(2), 357.

Asmolov, G. (2024). Propaganda in the Network Environment: How propaganda has changed in the era of social media and during times of war. Re: Russia.

Azuma, H. (2009). Otaku: Japan's database animals. U of Minnesota Press.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In Handbook of behavioural economics: applications and foundations 1 (Vol. 1, pp. 79-175). North-Holland.

Baumeister, R. F., & Leary, M. R. (2017). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Interpersonal development, 57-89.

Baumer, E. P. (2017). Toward human-centered algorithm design. Big Data & Society, 4(2), 2053951717718854.

Bedi, P., & Vashisth, P. (2014). Empowering recommender systems using trust and argumentation. Information sciences, 279, 569-586.

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. New media & society, 11(6), 985-1002.

Beer, D. (2019). The social power of algorithms. In The social power of algorithms (pp. 1-13). Routledge.

Boccio, R. (2022). Race after technology: Abolitionist tools for the new Jim Code by Ruha Benjamin. Configurations, 30(2), 236-238.

Bommasani, R., Kapoor, S., Klyman, K., Longpre, S., Ramaswami, A., Zhang, D., ... & Liang, P. (2024). Considerations for governing open foundation models. Science, 386(6718), 151-153.

Bourdieu, P. (1991). Language and symbolic power. Harvard University Press.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The crowdless future? Generative AI and creative problem-solving. Organisation Science, 35(5), 1589-1607.

Bucher, T. (2018). If... then: Algorithmic power and politics. Oxford University Press.

Buckingham, D. (2019). The media education manifesto. John Wiley & Sons.
Burr, C., Cristianini, N., & Ladyman, J. (2018). An Analysis of the Interaction Between Intelligent Software Agents and Human Users: C. Burr et al. Minds and machines, 28(4), 735-774.
Carrington, V. (2018). The changing landscape of literacies: Big data and algorithms. Digital Culture and Education, 10(1), 67-76.
Citron, D. K., & Solove, D. J. (2022). Privacy harms. BUL Rev., 102, 793.
Cohen, J. E. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Surveillance & Society, 17(1/2), 240-245.
Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev., 55, 93.
Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. Digital journalism, 5(4), 404-419.
Drew, R. (2016). Technological determinism. A companion to popular culture, 165-183.
Ellul, J. (2021). Propaganda: The formation of men's attitudes. Vintage.
Entman, R. M. (1993). Framing: Towards clarification of a fractured paradigm. McQuail's reader in mass communication theory, 390, 397.
Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016, May). First I" like" it, then I hide it: Folk Theories of Social Feeds. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 2371-2382).
Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., ... & Sandvig, C. (2015, April). " I always assumed that I wasn't really that close to [her]" Reasoning about invisible algorithms in news feeds. In Proceedings of the 33rd annual ACM conference on human factors in computing systems (pp. 153-162.
Etzioni, A. (2018). The limits of transparency. In Transparency, Society and Subjectivity: critical perspectives (pp. 179-201). Cham: Springer International Publishing.
Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. Macmillan+ ORM.

Fawzi, N. (2019). Untrustworthy news and the media as “enemy of the people?” How a populist worldview shapes recipients’ attitudes toward the media. The International Journal of Press/Politics, 24(2), 146-164.

Fisher, M. (2022). The chaos machine: The inside story of how social media rewired our minds and our world. Hachette UK.

Foucault, M., & Mailänder, E. (1975). Discipline and Punish (Vol. 225, p. Na). na.

Gal, M. S., & Elkin-Koren, N. (2016). Algorithmic consumers. Harv. JL & Tech., 30, 309.

Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

Gless, S., Silverman, E., & Weigend, T. (2016). If robots cause harm, who is to blame? Self-driving cars and criminal liability. New Criminal Law Review, 19(3), 412-436.

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. Harvard University Press.

Gomez, J. (2017). An Analysis of Roland Barthes's Mythologies. Macat Library.

Habermas, J. (2020). The public sphere: An encyclopedia article. In Critical Theory and Society (pp. 136-142). Routledge.

Hardin, R. (2002). Trust and trustworthiness. Russell Sage Foundation.

Helberger, N., Karppinen, K., & D’acunto, L. (2018). Exposure diversity as a design principle for recommender systems. Information, communication & society, 21(2), 191-207.

Hellman, D. (2023). Big data and compounding injustice. Journal of Moral Philosophy, 21(1-2), 62-83.

Kahneman, D. (2011). Thinking, fast and slow. macmillan.

Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Ramaswami, A., Cihon, P., ... & Narayanan, A. (2024, July). Position: On the societal impact of open foundation models. In the Forty-First International Conference on Machine Learning.

Kearney, A., Poredi, N., Shelton, J. A., Akcinaroglu, S., Karakoc, E., Tran, T., & Chen, Y. (2025, May). Echoes amplified: a study of AI-generated content and digital echo chambers. In Disruptive Technologies in Information Sciences IX (Vol. 13480, pp. 164-202). SPIE.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. Journal of Experimental Political Science, 9(1), 104-117.

Lewin, K. (1947). Frontiers in group dynamics: II. Channels of group life; social planning and action research. Human relations, 1(2), 143-153.

Lewis, M. (2014). Flash boys: a Wall Street revolt (Vol. 1). New York: WW Norton & Company.

Livingstone, S. (2004). Media literacy and the challenge of new information and communication technologies. The communication review, 7(1), 3-14.

Lukoff, K., Lyngs, U., Zade, H., Liao, J. V., Choi, J., Fan, K., ... & Hiniker, A. (2021, May). How the design of youtube influences user sense of agency. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. Proceedings of the ACM on human-computer interaction, 3(CSCW), 1-32.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. Public opinion quarterly, 36(2), 176-187.

Miller, J. M. (2019). Cold war democracy: the United States and Japan. Harvard University Press.

Ming, W. (2023). Platformization of Chinese society: capital, labor and technology. Вопросы теории и практики журналистики, 12(2), 333-346.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big data & society, 3(2), 2053951716679679.

Monge Roffarello, A., & De Russis, L. (2022, April). Towards understanding the dark patterns that steal our attention. In Chi conference on human factors in computing systems, extended abstracts (pp. 1-7).

Mudde, C. (2004). The populist zeitgeist. Government and opposition, 39(4), 541-563.

Mukherjee, A., & Chang, H. (2023). Managing the creative frontier of generative AI: The novelty-usefulness tradeoff. California Management Review.

Nair, N., & Trehan, G. (2024). Examining the use of Ai-powered social media analytics for target customer segmentation: a systematic review In retail industry. Educational Administration: Theory and Practice, 30(4), 1798-1805.

NAJEH, W. A., & BENARBI, H. (2023). Towards new perspectives in corporate governance: A literature review. International Journal of Performance and Organizations, 2(2), 146-161.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In Algorithms of Oppression. New York University Press.

Okidegbe, N. (2021). Discredited data. Cornell L. Rev., 107, 2007.

O'neill, O. (2002). A question of trust: The BBC Reith Lectures 2002. Cambridge University Press.

O'neil, C. (2017). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin UK.

Pascal, U. (2018). Personalising Persuasion Architecture: Privacy Harms and Algorithmic News Media. AAAI. Palo Alto.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. In the black box society. Harvard University Press.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological Science, 31(7), 770-780.

Perifanis, N. A., & Kitsios, F. (2023). Investigating the influence of artificial intelligence on business value in the digital era of strategy: A literature review. Information, 14(2), 85.

Perrotta, C., Gulson, K. N., Williamson, B., & Witzenberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. Critical Studies in Education, 62(1), 97-113.

Pettit, P. (2001). A theory of freedom: from the psychology to the politics of agency. Oxford University Press.

Rader, E., & Gray, R. (2015, April). Understanding user beliefs about algorithmic curation in the Facebook news feed. In Proceedings of the 33rd annual ACM conference on human factors in computing systems (pp. 173-182).

Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the academy of marketing science, 48(1), 137-141.

Rose, J., & MacGregor, O. (2021). The architecture of algorithm-driven persuasion. Journal of Information Architecture, 6(1), 7-40.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020, January). Auditing radicalization pathways on YouTube. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 131-141).

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. Creativity research journal, 24(1), 92-96.

Saari, A., & Säntti, J. (2018). The rhetoric of the 'digital leap'in Finnish educational policy documents. European Educational Research Journal, 17(3), 442-457.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In Proceedings of the conference on fairness, accountability, and transparency (pp. 59-68).

Shin, D. (2020). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. Computers in human behaviour, 109, 106344.

Smith, M. R., & Marx, L. (Eds.). (1994). Does technology drive history?: The dilemma of technological determinism. Mit Press.

Sunstein, C. R. (2015). Fifty shades of manipulation. Journal of Marketing Behaviour, 1(3-4), 213-244.

Sunstein, C. R. (2018). Republic: Divided democracy in the age of social media.

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. Social media+ society, 7(2), 20563051211008828.

Therborn, G. (2016). Different roads to modernity and their consequences: A sketch. In Decolonising European Sociology (pp. 71-83). Routledge.

Thompson, J. B. (2005). The new visibility. Theory, culture & society, 22(6), 31-51.

Van Dijck, J., Poell, T., & De Waal, M. (2018). The platform society: Public values in a connective world. Oxford University Press.

Verma, G., Bhardwaj, A., Aledavood, T., De Choudhury, M., & Kumar, S. (2022). Examining the impact of sharing COVID-19 misinformation online on mental health. Scientific Reports, 12(1), 8045.

Waisbord, S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. Journalism studies, 19(13), 1866-1878.

Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. Digital Journalism, 6(3), 274-293.

Walther, C. C. (2024). Where: Humans, technology, and humane technology. In Human Leadership for Humane Technology: The New AI: Agency Ignited (pp. 103-194). Cham: Springer Nature Switzerland.

Wylie, C. (2019). Mindf* ck: Cambridge Analytica and the plot to break America. Random House.

Yasseri, T., Margetts, H., John, P., & Hale, S. (2016). Political turbulence: How social media shape collective action. Princeton University Press.

Yu, P. K. (2021). Beyond transparency and accountability: three additional features algorithm designers should build into intelligent platforms. Ne. UL Rev., 13, 263.

Zingales, L. (2017). Towards a political theory of the firm. Journal of Economic Perspectives, 31(3), 113-130.

Zuboff, S. (2023). The age of surveillance capitalism. In Social theory re-wired (pp. 203-213). Routledge.

ISBN: 978-625-90132-4-4