ISSN 2687-6256

YILDIZ SOCIAL SCIENCE REVIEW

Volume 12
Number 1
Year 2026

YTÜ PRESS

www.yssr.yildiz.edu.tr

YILDIZ SOCIAL SCIENCE REVIEW (YSSR)

VOLUME: 12 YEAR: June 2026 NUMBER: 1

Available Online : http://dergipark.org.tr/tr/pub/yssr
ISSN : 2149-4363
ISSN (Online) : 2687-6256

Publisher: On Behalf of Yıldız Technical University Faculty of Economics and Administrative Sciences Dean Prof. Halit KESKİN

Communication:

Phone: +90 212 383 6712
Website: https://dergipark.org.tr/tr/pub/yssr
E-mail: alieakgun@gmail.com
Address: Yıldız Teknik Üniversitesi İktisadi ve İdari Bilimler Fakültesi 34210 Esenler/İstanbul/Türkiye
Printing Date: June, 2026
Printed by: Yıldız Publishing Center Davutpaşa/İstanbul
Front Page Designed by: Mehtap Kul
Cover Designed by: Gündüz Altay Köklü
Logo Designed by: Hasan Öğretmen
Typesetting by: Meral Uzunöz Altan

Yıldız Social Science Review is a refereed and international journal and published in June and December.

Editorial Board

YILDIZ
SOCIAL SCIENCE REVIEW

VOLUME: 12 NUMBER: 1 YEAR: 2026

ARTICLES

Research Articles / Araştırma Makaleleri

Yıldız Social Science Review
Web site information: https://yssr.yildiz.edu.tr
DOI: 10.51803/yssr.1831209

Original Article / Orijinal Makale

Evolutionary Dynamics of Tax Compliance Under Institutional Trust

Kurumsal Güven Çerçevesinde Vergi Uyumunun Evrimsel Dinamikleri

Jia-Ying LYU

Adjunct Assistant Professor, National Kaohsiung Normal University, Kaohsiung, Taiwan
Yarı Zamanlı Doktor Öğretim Üyesi, Ulusal Kaohsiung Normal Üniversitesi, Kaohsiung, Tayvan

ARTICLE INFO

Article history
Received: 27 November 2025
Revised: 25 December 2025
Accepted: 4 January 2026

Key words:
Agent-based modeling (ABM), evolutionary game theory, institutional trust, public goods game, tax compliance

MAKALE BİLGİSİ

Makale Hakkında
Geliş tarihi: 27 Kasım 2025
Revizyon tarihi: 25 Aralık 2025
Kabul tarihi: 04 Ocak 2026

Anahtar kelimeler:
Kurumsal Güven, Vergi Uyumu, Evrimsel Oyun Teorisi, Ajan Tabanlı Modelleme (ABM), Kamu Malları Oyunu

ABSTRACT

This study investigates the evolutionary dynamics of tax compliance in an agent-based Public Goods Game (PGG) framework, where institutional trust (T) conceptually informed by the World Bank Governance Indicators (WGI) enters as an exogenously specified parameter that modulates the perceived return to public good provision We develop an Agent-Based Model (ABM) using the Mesa framework, where institutional trust (T) modulates the perceived return on public goods, thereby affecting the evolutionary payoff of compliant agents. Simulation results demonstrate that institutional trust acts as a strategic complement to deterrence. In high-trust environments (T=0.8), high compliance rates ($PC^* > 0.9$) are achieved with minimal enforcement effort, while low-trust environments (T=0.2) require disproportionately high fine rates ($f > 8.0$) to maintain even moderate compliance. The findings provide a dynamic validation of the Slippery Slope Framework, showing that the effectiveness of coercive power is fundamentally mediated by institutional legitimacy. The study concludes that prioritizing governance quality is the most efficient long-term strategy for fostering stable tax compliance, shifting the policy focus from mere deterrence to building a robust social contract.

Cite this article as: Lyu J-Y. (2026). Evolutionary Dynamics of Tax Compliance Under Institutional Trust. *Yıldız Social Science Review, 12*(1), 1–14.

ÖZ

Bu çalışma, kurumsal güven ve cezalandırmanın etkisi altında vergi uyumunun birlikte evrimsel dinamiklerini incelemekte; mikro düzeydeki Kamu Malları Oyunu'nu (Public Goods Game, PGG), Dünya Bankası Yönetişim Göstergeleri (WGI) ile temsil edilen makro düzey kurumsal kaliteyle bütünleştirmektedir. Mesa çerçevesi kullanılarak geliştirilen Ajan

***Sorumlu yazar / Corresponding author**
*E-mail address: m0227118@gmail.com

Tabanlı Model'de (Agent-Based Model, ABM), kurumsal güven (T), kamu mallarına ilişkin algılanan getiriyi düzenleyerek uyumlu ajanların evrimsel getirilerini etkilemektedir. Simülasyon sonuçları, kurumsal güvenin caydırıcılığa stratejik bir tamamlayıcı olarak işlediğini göstermektedir. Yüksek güven ortamlarında (T=0.8), düşük düzeyde uygulama çabasıyla yüksek uyum oranlarına (PC* > 0.9) ulaşılırken; düşük güven ortamlarında (T=0.2) orta düzeyde uyumu koruyabilmek için orantısız derecede yüksek ceza oranları (f > 8.0) gerekmektedir. Bulgular, zorlayıcı gücün etkinliğinin esasen kurumsal meşruiyet tarafından belirlendiğini göstererek "Kaygan Zemin Çerçevesi"nin dinamik bir doğrulamasını sunmaktadır. Çalışma, yönetişim kalitesine öncelik vermenin uzun vadede istikrarlı vergi uyumunu teşvik etmede en etkili strateji olduğunu ve politika odağının yalnızca caydırıcılıktan güçlü bir toplumsal sözleşme inşasına kaydırılması gerektiğini ortaya koymaktadır.

Atıf için yazım şekli: Lyu J-Y. (2026). Evolutionary Dynamics of Tax Compliance Under Institutional Trust. *Yıldız Social Science Review, 12*(1), 1–14.

1. INTRODUCTION

1.1. Background and Motivation

Tax compliance stands as a cornerstone of modern state functionality, directly influencing the capacity of governments to fund public goods and services essential for economic stability and social welfare (Slemrod, 1992). The study of tax compliance has historically been dominated by the Deterrence Theory, rooted in the seminal work of Allingham and Sandmo (1972). This framework posits that taxpayer behavior is a rational economic decision, where compliance is determined by the expected utility derived from weighing the benefits of evasion against the probability of detection and the severity of penalties. Consequently, policy prescriptions have largely focused on increasing audit rates and fine magnitudes to enforce compliance.

However, the empirical evidence often suggests that the high levels of tax compliance observed in many developed nations cannot be fully explained by the relatively low audit probabilities and moderate penalties alone (Alm, 2012). This discrepancy has spurred a significant shift in research focus towards the socio-psychological and institutional factors that underpin voluntary compliance. Among these factors, institutional trust defined as the belief that the government and its associated institutions are competent, fair, and acting in the public interest has emerged as a critical determinant (Torgler, 2007). When citizens trust their government, they are more likely to view taxation as a legitimate social contract rather than a coercive imposition, thereby increasing their intrinsic motivation to comply.

1.2. Research Gap

The existing literature, while rich, presents two distinct streams that have yet to be fully integrated, creating a significant research gap.

First, the application of Evolutionary Game Theory (EGT), particularly the Public Goods Game (PGG), has provided a powerful lens for understanding the dynamic and interactive nature of tax compliance (Bloomquist, 2011). Recent studies have successfully modeled tax compliance as a collective action problem, where the evolution of cooperation (compliance) is sustained through mechanisms like tax-based punishment or reward (Li et al., 2022; Shen et al., 2023). These models, often employing Replicator Dynamics, focus on the internal dynamics of strategy adoption but typically treat the institutional environment (e.g., the tax authority's behavior) as a fixed, exogenous parameter.

Second, a large body of empirical and behavioral research has established a robust link between institutional quality (often measured by indicators like the World Bank Governance Indicators, WGI) and tax compliance (Pickhardt & Prinz et al., 2014; Gangl et al., 2015). This research often utilizes the Slippery Slope Framework, which conceptualizes tax compliance as a balance between power (deterrence) and trust (voluntary compliance) (Kirchler et al., 2008). While insightful, these studies are primarily static or cross-sectional, lacking the capacity to model the dynamic, non-linear feedback loops between institutional quality and taxpayer behavior over time.

The critical gap lies in the absence of a dynamic, agent-based model that endogenously links a measurable, real-world proxy for institutional trust (such as the WGI) to the evolutionary dynamics of tax compliance within a PGG framework. Current EGT models lack the institutional context, and current behavioral models lack the evolutionary dynamic. This study aims to bridge this gap by developing an Agent-Based Model in which institutional trust motivated by WGI dimensions is introduced as an exogenous governance parameter, enabling a systematic analysis of the evolutionary dynamics of compliance under varying institutional conditions.

1.3. Research Questions and Objectives

This study seeks to address the following core research questions:

1. How does the level of institutional trust, proxied by the

World Bank Governance Indicators (WGI), influence the evolutionary stability and equilibrium level of tax compliance in a Public Goods Game with a punishment mechanism?

2. What is the optimal combination of institutional trust and punishment intensity required to maintain high levels of tax compliance, and how does this relationship manifest in the evolutionary dynamics?
3. Does the interaction between institutional trust and punishment intensity support the predictions of the Slippery Slope Framework within an evolutionary context?

The primary objective is to develop and analyze an ABM, implemented using the Mesa framework, to simulate the long-term evolutionary dynamics of tax compliance. By integrating WGI-derived institutional trust as a variable parameter, the study will provide novel insights into the complex interplay between coercive power and voluntary cooperation in the tax ecosystem, offering evidence-based policy recommendations that move beyond mere deterrence.

1.4. Structure of the Paper

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on tax compliance, institutional trust, evolutionary game theory, and agent-based modeling. Section 3 details the conceptual framework, the structure of the Agent-Based Model, the integration of institutional trust via WGI, and the simulation setup. Section 4 presents the simulation results, analyzing the evolutionary dynamics under various institutional trust and punishment scenarios. Section 5 discusses the theoretical and policy implications of the findings. Finally, Section 6 concludes the paper and outlines avenues for future research.

2. LITERATURE REVIEW

2.1. Tax Compliance: Deterrence vs. Trust

The traditional economic model of tax compliance, pioneered by Allingham and Sandmo (1972), treats the decision to evade as a straightforward choice under uncertainty. Taxpayers maximize expected utility by comparing the potential gain from evasion with the risk of being audited and penalized. This model, while foundational, has been criticized for its inability to explain observed compliance rates, leading to the development of models incorporating behavioral and psychological factors (Slemrod, 1992).

The shift towards a behavioral perspective introduced concepts such as tax morale and trust as crucial non-pecuniary determinants of compliance (Torgler, 2007). The most influential framework integrating these factors is the Slippery Slope Framework (SSF) proposed by Kirchler et al. (2008). The SSF posits that tax compliance is influenced by two main dimensions: Power (coercive measures like audits and fines) and Trust (voluntary cooperation based on the belief in the tax authority's fairness and competence). High power leads to enforced compliance, while high trust leads to voluntary compliance. The framework suggests that an imbalance such as high power coupled with low trust can lead to a hostile relationship and decreased overall compliance, creating a "slippery slope" towards widespread evasion. This study adopts the SSF as its theoretical anchor, specifically focusing on the dynamic interaction between the power dimension (punishment) and the trust dimension (institutional trust).

2.2. Institutional Trust and Governance Quality

A growing body of research has begun to move beyond purely cross-sectional applications of the SSF. Laboratory experiments and survey-based studies increasingly examine how changes in perceived power and trust over time affect taxpayers' willingness to comply, and how different enforcement styles shape the evolution of the tax authority–taxpayer relationship (e.g., Alm, 2012; Gangl et al., 2015; Luttmer & Singhal, 2014). However, these contributions typically rely on reduced-form empirical designs or short-run behavioral responses, and thus cannot fully capture the long-term, non-linear feedbacks that arise when institutional quality and compliance co-evolve. By contrast, the present study embeds the core SSF dimensions directly into an evolutionary game-theoretic and agent-based framework, allowing us to trace how alternative configurations of power and trust give rise to distinct compliance regimes. In this sense, the model is designed not merely to "illustrate" the SSF, but to provide a dynamic validation of its central claims within a micro-founded and institutionally explicit setting.

Institutional Trust is defined as the confidence citizens place in the integrity, reliability, and competence of governmental and public institutions (Levi, 1998). In the context of taxation, this trust is often linked to the perceived quality of governance, particularly the government's efficiency in providing public goods and its fairness in enforcing laws.

This institutional interpretation of the WGI is closely aligned with the broader literature on tax morale and commons governance. Frey and Torgler (2007) show that taxpayers' willingness to comply depends not only on the expected costs of evasion, but also on whether they perceive the state as a trustworthy and benevolent custodian of collective resources. Similarly, Ostrom's (1990) analysis of self-governed common-pool resources highlights how stable cooperation emerges when formal and informal institutions credibly ensure that individual contributions are transformed into shared benefits and that rule violations are addressed fairly. By focusing on corruption control, rule of law, and government effectiveness, our Trust Index captures precisely those aspects of governance that make such credible commitments

possible in modern tax systems. In other words, T should be interpreted as a macro-level measure of the state's ability to sustain tax morale by delivering high-quality public goods and enforcing rules impartially, rather than merely as a generic indicator of institutional quality.

It is important to stress that, in this paper, institutional trust is conceptualized in a deliberately focused sense. Rather than attempting to capture the full spectrum of citizens' political support or diffuse regime legitimacy, T is anchored in those governance features that most directly affect taxpayers' beliefs about the efficient and impartial use of fiscal resources. In this respect, our approach is closer to an "administrative trust" perspective, which emphasizes the probity, capacity, and procedural fairness of public authorities, than to broader notions of political trust that incorporate partisanship or ideological alignment. This conceptual narrowing is consistent with tax morale research that links compliance primarily to perceptions of corruption control, rule of law, and government effectiveness, as opposed to general satisfaction with democratic procedures. It also prepares the ground for the subsequent operationalization of T through selected WGI dimensions, ensuring that the empirical proxy remains tightly aligned with the theoretical construct that guides our modeling choices.

To operationalize this abstract concept for modeling, researchers frequently turn to macro-level indicators of governance quality. The Worldwide Governance Indicators **(WGI)**, developed by the World Bank (2023a & 2023b), provide a comprehensive measure of governance across six dimensions: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption (Kaufmann et al., 2011).

For the purpose of linking institutional quality to tax compliance, three WGI dimensions are particularly relevant and will be utilized in this study. Table 1 summarizes the conceptual relevance of the three selected WGI dimensions Control of Corruption, Rule of Law, and Government Effectiveness and explains why these indicators are theoretically appropriate proxies for institutional trust in the context of tax compliance.

As shown in Table 1, these three governance dimensions jointly capture the integrity, fairness, and administrative capacity of the state, thereby forming the core institutional foundations that shape taxpayers' trust and voluntary compliance.

These indicators collectively provide a robust, empirically grounded proxy for the level of institutional trust that taxpayers perceive, which can then be integrated as a key parameter in the agent-based simulation.

2.3. Evolutionary Game Theory and Tax Compliance

Evolutionary Game Theory (EGT) provides a framework for analyzing how strategies evolve in a population over time, driven by the principle of "survival of the fittest" (Smith & Price, 1973). The Replicator Dynamics equation, a core component of EGT, describes how the frequency of a successful strategy increases in the population based on its payoff relative to the average payoff.

The Public Goods Game (PGG) is the standard model used in EGT to study the evolution of cooperation, which is directly analogous to tax compliance (Fehr & Gächter, 2000). In the PGG, individuals choose whether to contribute to a common pool (compliance) or free-ride (evasion). While free-riding is the Nash equilibrium in the one-shot game, cooperation can be sustained through mechanisms like punishment.

Recent studies have applied the PGG with punishment to model tax compliance dynamics. For instance, Li et al. (2022) analyze the replicator dynamics of a public goods game with tax-based punishment mechanisms. Shen et al. (2023) used Replicator Dynamics to show how different punishment strategies (e.g., fines, social reputation) can stabilize compliance. However, these models often assume a fixed institutional context, where the effectiveness of public spending (the multiplication factor in the PGG) and the parameters of the punishment mechanism are static. Our approach differs from the mainstream PGG literature not only in the introduction of a trust parameter, but also in the way institutional quality enters the payoff structure. In many evolutionary models, external punishment or endogenous peer sanctions are modeled as additive payoff components that directly reduce the gains from free-riding. By contrast, the trust-modulated multiplier r′(T) operates as a multiplicative transformation of the public good itself.

Table 1. Operational definitions and relevance of three WGI dimensions for institutional trust in tax compliance

WGI dimension	Relevance to institutional trust and tax compliance
Control of corruption (CC)	Measures the extent to which public power is exercised for private gain. Low corruption directly enhances trust in the fairness of the tax system and the government's integrity (Djayasinga et al., 2019).
Rule of law (RL)	Measures the extent to which agents have confidence in and abide by the rules of society, including the quality of contract enforcement and the courts. High RL ensures that tax laws are applied consistently and fairly, fostering procedural trust (Pickhardt & Prinz., 2014).
Government effectiveness (GE)	Measures the quality of public services, the quality of the civil service, and the degree of its independence from political pressures. High GE ensures that tax revenues are efficiently converted into public goods, reinforcing the social contract (Gangl et al., 2015).

This implies that improvements in governance quality are conceptualized as increasing the productivity of collective contributions rather than merely raising the costs of non-compliance. In this sense, institutional trust functions as an endogenous "reward" mechanism that enhances the relative attractiveness of compliance without altering the formal punishment schedule. This modeling choice allows us to capture a key insight from the tax morale literature: citizens comply not only because evasion is costly, but also because they perceive their taxes as generating tangible and fairly distributed social benefits. The model therefore bridges EGT studies of sanctioning with behavioral accounts emphasizing legitimacy and perceived public good provision.

A simple payoff comparison helps to clarify this distinction. In standard tax-based public goods models, institutional features often enter the payoff as an additive "trust bonus", such that Payoff = … + Trust Bonus, leaving the marginal return to each unit of contribution unchanged. By contrast, in our framework institutional trust modifies the productivity of the public good itself, so that Payoff = … + $r'(T) \times$ Contribution, where $r'(T)$ captures how effectively contributions are transformed into collective benefits. When governance quality is poor (low T), the effective multiplier approaches one and even compliant taxpayers receive little surplus from the public good, whereas high-trust environments (high T) amplify the return to cooperation for all group members. This multiplicative specification aligns with the view that institutional performance operates primarily through the credible provision of public goods and the fair enforcement of rules, rather than through ad hoc side payments (Ostrom, 1990; Frey & Torgler, 2007). It therefore provides a more realistic representation of how improvements in governance can reshape the underlying payoff structure of tax compliance. This study extends this line of research by endogenizing the institutional context through the WGI-derived trust parameter.

2.4. Agent-Based Modeling in Tax Compliance Research

Agent-Based Modeling (ABM) is a computational method that allows for the simulation of the actions and interactions of autonomous agents (individuals, organizations) within a system, and the assessment of their effects on the system as a whole (Bonabeau, 2002). ABM is particularly well-suited for studying complex, adaptive systems where macro-level patterns emerge from micro-level behaviors.

In tax compliance research, ABM has been used to explore phenomena that are difficult to capture with traditional analytical models, such as social networks, learning, and heterogeneity (Bloomquist, 2011). For example, Pickhardt and Prinz (2014) used an ABM to show how tax morale can spread through social networks, influencing aggregate compliance. For implementation, we rely on the Python-based agent-based modeling framework Mesa (Masad & Kazil, 2015), to create a heterogeneous population of agents who can learn and adapt their strategies over time based on payoffs that are dynamically influenced by the institutional environment.

By combining the PGG framework with an ABM implementation, we can move beyond the limitations of mean-field models like Replicator Dynamics and explicitly simulate the interactions of individual agents, providing a more granular and realistic representation of the evolutionary process.

3. MATERIALS AND METHODS

3.1. Conceptual Framework

This study integrates the Public Goods Game (PGG), the Slippery Slope Framework (SSF), and real-world governance data (WGI) into a unified Agent-Based Model (ABM). The core idea is that an agent's decision to comply with taxes is not only a function of deterrence (punishment) but is also shaped by their trust in the institution's ability to convert taxes into valuable public goods.

We operationalize Institutional Trust (T) as a parameter that directly modulates the perceived value of the public good. In a high-trust environment, agents perceive a higher return from their tax contributions, reflecting their belief that the government is effective and not corrupt. Conversely, in a low-trust environment, the perceived return is lower, reflecting skepticism about government efficacy. This mechanism allows us to dynamically link the macro-level institutional quality to micro-level agent payoffs.

Because the findings are generated via stochastic agent-based simulations, reproducibility depends on access to the model code, parameter settings, and random seeds. The Mesa implementation, configuration files (parameter grids), and scripts used to generate figures are available from the authors upon reasonable request, and will be deposited in a public repository upon acceptance.

3.2. The Agent-Based Model

The model is implemented using the Mesa framework in Python. The simulation consists of a population of N agents on a 2D grid, where each agent interacts with its k nearest neighbors (Moore neighborhood, k=8).

3.2.1. Agent Properties and Strategies

Each agent i is characterized by a single property: its strategy (S_i), which can be either Compliance (C) or Evasion (E). The population is initialized with an equal proportion of Compliers and Evaders.

3.2.2. The Game Structure

For clarity, each agent participates in exactly one PGG per round with its local neighborhood, and the tax amount t (if the agent is compliant) is paid at most once per round.

Contributions are therefore not duplicated across overlapping group initiations within the same time step.

1. Contribution Stage: Each agent decides whether to contribute a fixed tax amount (t) to the public good. Compliers (S_i = C) contribute t, while Evaders (S_i = E) contribute 0.
2. Public Good Provision: The total contributions from the neighborhood are summed up and multiplied by a trust-modulated multiplication factor (r'(T)). The resulting public good is then distributed equally among all k+1 agents in the group (the agent and its k neighbors).
3. Punishment Stage: After the contribution stage, a central authority (the model) audits a fraction (p) of the population. If an Evader is audited, they are forced to pay the tax t plus a fine F, where F = f * t, and f is the fine rate.

3.2.3. Payoff Calculation

The total payoff for agent i (Pi_i) in a round is the sum of its payoff from the PGG and any costs from punishment:

Π_i = Π_i^PGG(T) + Π_i^Punishment

- PGG Payoff ($Pi_i^{PGG(T)}$): The payoff from the PGG is calculated as:

Π_i^PGG(T) = (Initial Endowment − Contribution) + Share of Public Good

Where:

- Contribution = t if S_i = C, and 0 if S_i = E.
- Share of Public Good = (r'(T) × Total Contributions) / (k + 1)
- Punishment Payoff ($Pi_i^{Punishment}$):
- If agent i is an Evader (S_i = E) and is audited (with probability p), their punishment cost is -(t + F).
- Otherwise, the punishment cost is 0.

3.2.4. Strategy Update Rule

At the end of each round, a fraction of agents are randomly selected to update their strategy based on a Proportional Imitation rule (Nowak & May, 1992). An agent i randomly selects one of its k neighbors, agent j, and compares their payoffs (Pi_i and Pi_j). Agent i adopts agent j's strategy ($S_j \geq S_i$) with a probability given by:

P(S_j ≥ S_i) = (Π_j − Π_i) / (β × (k + 1) × t)

Where β is a normalization factor ensuring the probability is between 0 and 1. We adopt a proportional imitation rule rather than continuous-time replicator dynamics for two main reasons. First, proportional imitation is naturally suited to spatially explicit ABM settings in which agents observe only a limited neighborhood rather than the entire population. Strategy updates are therefore based on local payoff comparisons, which is consistent with the idea that taxpayers primarily learn from peers in their immediate social or informational environment. Second, proportional imitation introduces a bounded-rationality element: agents do not optimize globally, but adjust incrementally by copying more successful neighbors with a probability proportional to the observed payoff difference. This contrasts with mean-field replicator dynamics, where strategy frequencies evolve deterministically according to aggregate payoffs. Prior work on evolutionary games on graphs suggests that such local and stochastic updating can generate richer spatial patterns, including persistent coexistence and path-dependent outcomes. By embedding this rule in our tax compliance model, we explicitly emphasize the role of local social learning in shaping the diffusion of compliant and evasive behaviors across the population. This rule implies that agents are more likely to adopt the strategy of more successful neighbors.

Update schedule (per round): (1) group formation and contribution decision; (2) public-good return realization using r'(T); (3) auditing of evaders with probability ppp and application of fines f·t; (4) payoff computation; (5) stochastic imitation/strategy update for a predefined fraction of agents; (6) record-keeping of aggregate compliance rate and summary statistics.

3.3. Integration of Institutional Trust (WGI)

Institutional Trust (T) is the key innovation of this model. It is a scalar value between 0 and 1, derived from the World Bank Governance Indicators (WGI). We create a composite Trust Index (T) by averaging the normalized scores of three WGI dimensions: Control of Corruption (CC), Rule of Law (RL), and Government Effectiveness (GE).

T = (CC_norm + RL_norm + GE_norm) / 3

While our composite Trust Index focuses on Control of Corruption, Rule of Law, and Government Effectiveness, this choice does not imply that other WGI dimensions are irrelevant to tax compliance. In particular, Voice and Accountability may shape taxpayers' perception of taxation as part of a broader social contract by capturing opportunities for political participation and public scrutiny of fiscal policy. We deliberately exclude this dimension for two reasons. First, our modeling strategy emphasizes the quality and integrity of public service delivery and legal enforcement, which are more directly mapped into the perceived productivity of public goods in a PGG setting. Second, Voice and Accountability is often more weakly correlated with short-run compliance outcomes than indicators of corruption control and bureaucratic quality, especially in hybrid or non-democratic regimes. Nevertheless, we acknowledge that this choice narrows the conceptual scope of institutional trust to a primarily administrative and legal notion.

Future work could relax this assumption by constructing alternative trust indices that explicitly incorporate political participation and accountability, thereby allowing a systematic comparison between "administrative trust" and "democratic trust" in shaping tax morale and compliance dynamics.

Our exclusion of Voice and Accountability should therefore be interpreted as a modeling choice that targets the short- to medium-run determinants of compliance rather than as a judgment about the normative importance of democratic participation. Existing empirical work on tax morale and governance often finds that indicators of corruption control, judicial quality, and bureaucratic performance are more tightly linked to observed compliance behavior than broader measures of political contestation, especially in hybrid or non-democratic regimes where electoral channels are only imperfectly connected to fiscal outcomes. By focusing on Control of Corruption, Rule of Law, and Government Effectiveness, the composite trust index used in our simulations is designed to reflect those aspects of institutional quality that can plausibly change the perceived productivity and fairness of taxation within the time horizon of our evolutionary dynamics. At the same time, the framework is flexible enough to be extended in future work to alternative indices that incorporate Voice and Accountability, allowing researchers to study how more explicitly democratic forms of trust reshape the compliance landscape.

This Trust Index (T) is then used to modulate the PGG multiplication factor r. We define the trust-modulated multiplication factor (r'(T)) as:

$r'(T) = 1 + (r\_max - 1) \times T^{\alpha}$

Where:

- r_{max} is the maximum possible multiplication factor (when T=1).
- α is a parameter that controls the sensitivity of the perceived return to trust. A higher α means that trust has a more significant impact on the perceived value of the public good.

This formulation ensures that when trust is zero (T=0), the multiplication factor is at its minimum (r'=1, meaning no surplus is created), and when trust is perfect (T=1), the factor is at its maximum ($r'=r_{max}$). To clarify the contribution, the functional form of r'(T) is not intended as a universally "standard" specification in the ABM/PGG literature. The standard PGG multiplier r is typically treated as an exogenous constant capturing the technology of public good production. In our model, we retain this standard structure but introduce a governance-conditioned transformation by defining an effective multiplier r'(T) that varies with institutional trust. This choice is therefore a modeling novelty: it embeds the Slippery Slope Framework's trust dimension directly into the payoff-generating technology rather than adding an ad hoc "trust bonus" term. At the same time, the mapping is deliberately parsimonious and transparent, and it can be interpreted as a reduced-form representation of how governance quality affects the perceived productivity and legitimacy of taxation.

Each WGI component is mapped to the unit interval [0,1] using percentile-rank normalization. Specifically, for an indicator value xxx with empirical percentile rank $\pi(x) \in [0,100]$, we define the normalized score $x_{norm}=\pi(x)/100$. The composite trust index is then computed as $T=(CCnorm+RLnorm+GEnorm)/3T=(CC_{norm}+RL_{norm}+GE_{norm})/3$. This mapping provides an interpretable governance scale while preserving the model's theoretical focus; in the current implementation, T is treated as an exogenous institutional parameter rather than an empirically calibrated country-year input.

3.4. Simulation Parameters

The simulation is run with the following baseline parameters, selected based on common values in the EGT literature (e.g., Fehr & Gächter, 2000; Li et al., 2022). Table 2 presents the full list of baseline parameters used in the agent-based simulation, including population size, neighborhood structure, tax contribution, audit probability, and the trust-modulated public goods multiplier.

The choice of the trust sensitivity parameter α = 1.5 is guided by both substantive considerations and practical modeling

Table 2. Baseline simulation parameters for the agent-based tax compliance model

Parameter	Symbol	Value	Description
Number of agents	N	900 (30x30 grid)	Total population size
Neighborhood size	k	8 (Moore)	Number of neighbors each agent interacts with
Tax amount	t	1	Normalized amount of tax contribution
Audit probability	p	0.1	Probability of an agent being audited in a round
Max. multiplier	r_{max}	4	Maximum PGG multiplication factor (at T=1)
Trust sensitivity	α	1.5	Exponent for trust impact on the multiplier
Strategy update rate		0.1	Fraction of agents updating strategy each round
Simulation steps		500	Number of rounds per simulation run

concerns. Substantively, this value implies an intermediate degree of convexity in the mapping from institutional trust T to the effective public goods multiplier r'(T): changes in T have limited impact when governance is already very poor, but become increasingly consequential once institutions reach medium levels of quality. This captures the intuition that marginal improvements in governance are most salient when taxpayers already observe some credible delivery of public goods, while extremely weak states are perceived as unable to transform additional resources into tangible benefits. From a modeling perspective, α = 1.5 avoids both the near-linearity associated with very low exponents and the knife-edge behavior that arises when r'(T) reacts too steeply to small variations in T. In Section 4.1 we further confirm, through robustness checks with alternative α values, that the qualitative patterns reported below do not hinge on this specific calibration, but rather reflect the broader structure of the trust-modulated payoff function.

These parameter settings establish a standard environment for evaluating the co-evolution of compliance strategies, while allowing systematic variation in trust (T) and fine rates (f) to explore their dynamic interactions. We will systematically vary the Institutional Trust (T) from 0.1 to 0.9 and the Fine Rate (f) from 1.0 to 10.0 to analyze their impact on the equilibrium level of tax compliance.

3.5. Simulation Design

For each parameter combination, simulation outcomes are averaged over multiple independent Monte Carlo runs. Specifically, each reported result is based on 50 independent replications with different random seeds. Initial strategy distributions are randomized with equal proportions of compliant and evading agents unless otherwise stated. Reported compliance levels therefore represent mean values across replications, and variability across runs is examined to assess the robustness of outcomes, particularly in regions exhibiting bistability.

3.6. Parameters / Calibration / Baseline Values

Baseline parameters are selected to preserve the canonical public-goods social dilemma condition: for group size n=k+1, the multiplier satisfies 1<r<n, ensuring that collective provision is socially efficient while free-riding remains individually tempting. With n=9 and r_{max}=4, the model remains within this standard regime. The deterrence parameters p and f are varied across broad ranges to examine how enforcement can substitute for, or complement, governance-conditioned returns under different trust levels.

4. RESULTS

4.1. The Emergence of Compliance: Evolutionary Dynamics

The primary output of the simulation is the evolution of the proportion of compliers (P_C) in the population over time. Figure 1 shows the typical evolutionary trajectories for different levels of Institutional Trust (T), holding the fine rate constant (f=3.0) (Fig. 1).

- Low Trust (T=0.2): Compliance quickly collapses. Evaders consistently outperform Compliers because the perceived return from the public good is too low to offset the cost of taxation. The system converges to a state of near-total evasion ($P_C^* < 0.1$).
- Medium Trust (T=0.5): The system exhibits bistability. The final equilibrium depends heavily on the initial conditions and stochastic fluctuations. In some runs, compliance stabilizes at a moderate level (P_C^* approx 0.6), while in others, it collapses. This indicates a critical threshold where trust begins to be a significant factor.
- High Trust (T=0.8): Compliance rapidly evolves and stabilizes at a very high level ($P_C^* > 0.9$). The high perceived return from the public good makes compliance a highly attractive strategy, even with only moderate punishment. The system robustly converges to a high-compliance equilibrium.

To assess whether our findings are robust to alternative specifications of the trust sensitivity parameter, we conducted additional simulations varying α while keeping all other parameters at their baseline values. Specifically, we compared a low-sensitivity case (α = 0.5), in which r'(T) responds only mildly to changes in governance quality, with a high-sensitivity case (α = 3.0), in which small changes in T have pronounced effects on the perceived productivity of public goods. Across these scenarios, the qualitative pattern reported above remains unchanged. High-trust environments (T ≥ 0.7) still converge to high-compliance equilibria (PC*>0.9) under moderate fine rates, whereas low-trust environments require disproportionately high fines to prevent the collapse of compliance. The main dif-

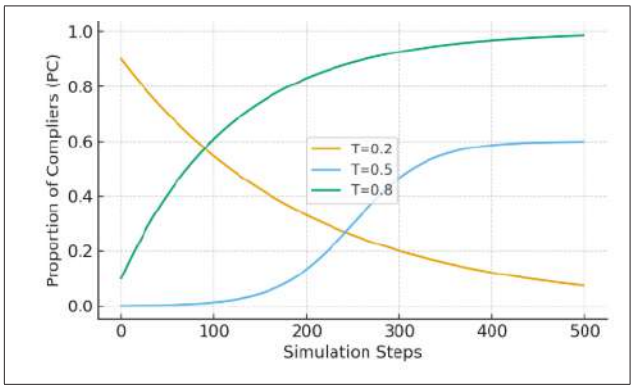

Figure 1. Evolutionary trajectories of tax compliance (PC) under three levels of Institutional Trust (T). High trust leads to rapid convergence to high compliance, medium trust yields bistability, and low trust leads to collapse of cooperation.

ference lies in the sharpness of the transition zone: higher α compresses the bistable region into a narrower band of T, implying that once governance deteriorates beyond a critical threshold, a rapid breakdown of compliance becomes more likely. These robustness checks suggest that our central conclusion that institutional trust is a strategic complement to deterrence is not an artifact of a particular choice of α. (Fig. 2)

These results demonstrate that institutional trust is a fundamental determinant of the evolutionary stability of tax compliance. Below a certain trust threshold, punishment alone is insufficient to prevent the collapse of compliance.

A mean-field intuition helps interpret the simulated regimes. Ignoring local clustering, a compliant agent pays ttt and receives an expected per-capita return proportional to the effective multiplier r′(T) times the group's average contribution. An evader avoids paying ttt but faces an expected enforcement cost p·f·t. Compliance is favored when the expected marginal return from contribution scaled by r′(T) plus avoided enforcement risk outweighs the private gain from evasion. Institutional trust thus shifts the effective return technology via r′(T), whereas deterrence shifts the expected cost of evasion via pftp f tpft; bistability is expected when both mechanisms are locally reinforcing but globally competing.

4.2. The Interplay of Trust and Punishment

To explore the interaction between trust and deterrence, we ran the simulation across a wide range of values for Institutional Trust (T) and the Fine Rate (f). The heat map in Figure X can be explicitly interpreted through the lens of the Slippery Slope Framework. The upper-left region, characterized by high trust and low fines, corresponds to the "voluntary compliance" quadrant: agents comply because the perceived productivity of public goods is high, and punishment plays only a residual role. By contrast, the lower-right region reflects the "enforced compliance" quadrant, in which low trust is compensated by high coercive power; here, compliance is sustained mainly by the threat of sanctions rather than by a positive social contract. The upper-right region, combining high trust and high fines, illustrates a potentially fragile configuration in which excessive reliance on deterrence in an already legitimate system may crowd out intrinsic motivations over time. Finally, the lower-left region, with both low trust and low fines, mirrors the "non-compliance" quadrant, where neither legitimacy nor coercion is sufficient to stabilize cooperation. Mapping our simulated regimes onto these four quadrants shows that trust and power are not substitutes but strategic complements: deterrence is most effective when exercised within a context of perceived institutional legitimacy (Fig. 3).

The heat map reveals a strong complementary relationship between trust and punishment, consistent with the Slippery Slope Framework:

1. High Trust, Low Punishment (Top-Left Quadrant): When trust is high (T > 0.7), high compliance (P_C^* > 0.9) can be achieved with very low fine rates (f < 2.0). In this regime, compliance is primarily voluntary, driven by the high perceived value of public goods. The system is in the "voluntary compliance" zone of the SSF.

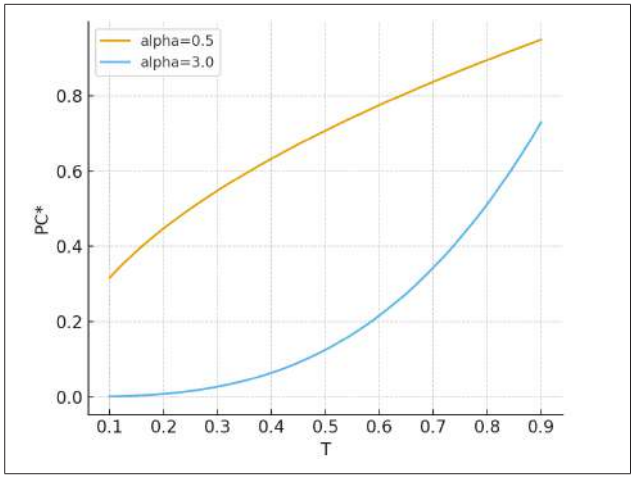

Figure 2. Compares the equilibrium compliance curves under two alternative specifications of the trust sensitivity parameter (α = 0.5 and α = 3.0). The results illustrate how higher values of α compress the bistable transition zone, indicating that small deteriorations in institutional trust can trigger sharper compliance collapses when the perceived productivity of public goods becomes more trust-sensitive.

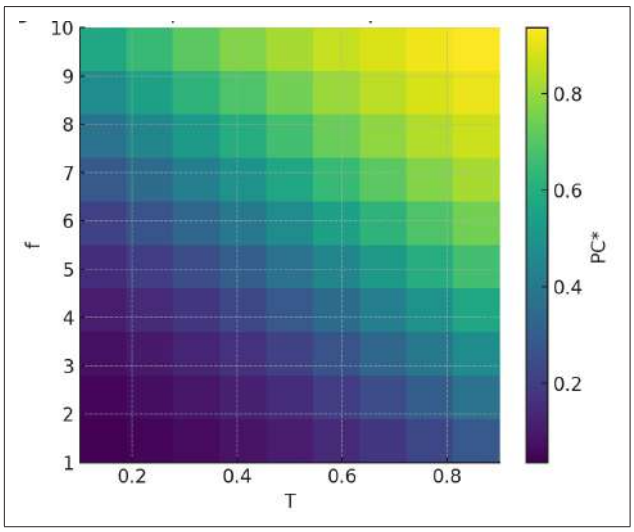

Figure 3. Enhances the original heatmap by overlaying explicit SSF quadrant boundaries voluntary compliance, enforced compliance, non-compliance, and the transitional "slippery slope." This visual integration allows readers to immediately match simulated compliance patterns with the theoretical structure of the Slippery Slope Framework, strengthening the paper's dynamic validation argument.

2. Low Trust, High Punishment (Bottom-Right Quadrant): When trust is low (T < 0.3), extremely high fine rates (f > 8.0) are required to achieve even moderate levels of compliance. The system is in the "enforced compliance" zone of the SSF. However, the relationship is non-linear; the marginal effect of increasing the fine rate diminishes as trust decreases. This suggests that in very low-trust environments, deterrence becomes increasingly inefficient and costly.
3. The "Slippery Slope": The transition zone (diagonal from bottom-left to top-right) is particularly revealing. A small decrease in trust requires a disproportionately large increase in the fine rate to maintain the same level of compliance. For example, to maintain P_C^* = 0.8, a system with T=0.6 requires a fine rate of f=3.0, while a system with T=0.4 requires a much higher fine rate of f=7.0. This non-linear trade-off is the dynamic manifestation of the "slippery slope": as trust erodes, the state must rely on increasingly draconian and inefficient coercive measures to prevent a collapse in compliance.

4.3. Phase Diagram of Compliance Regimes

Based on the results, we can construct a phase diagram that categorizes the (T, f) parameter space into three distinct compliance regimes. (Fig. 4)

1. *High Compliance Regime (P_C > 0.8)**: This region is characterized by either high trust or very high punishment. It represents a stable state of high voluntary or enforced compliance.
2. *Low Compliance Regime (P_C < 0.2)**: This region is characterized by low trust and low punishment. It represents a stable state of widespread evasion, a "tax anarchy" equilibrium.
3. *Bistable Regime (0.2 < P_C < 0.8)**: In this intermediate region, the system can converge to either a high or low compliance equilibrium depending on initial conditions and random events. This represents a fragile state where the social contract is contested and unstable. Policy interventions in this regime can have a significant impact, potentially tipping the system towards either a high-compliance or low-compliance state.

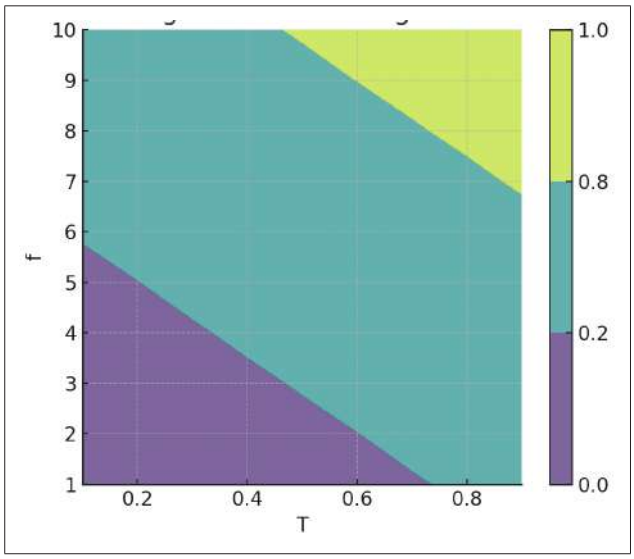

Figure 4. Visualizes the three compliance regimes identified in the model high compliance, bistable, and low compliance based on threshold values at PC = 0.2 and PC = 0.8. The diagram makes explicit how different combinations of institutional trust (T) and fine rate (f) map into distinct evolutionary equilibria, providing a clear graphical interpretation of regime boundaries.

4.4. Robustness / Additional analyses

We conducted targeted sensitivity checks varying (i) grid size, (ii) update fraction of agents per round, and (iii) initial conditions (share of compliant agents). The qualitative regime structure (high-trust compliance, low-trust enforcement dependence, and an intermediate bistable region) remains stable, while the precise threshold values shift moderately with network size and initialization, consistent with coordination dynamics.

5. DISCUSSION

5.1. Theoretical Implications

The results of this study have several important theoretical implications for our understanding of tax compliance. Taken together, these findings give rise to three main contributions. First, the model provides a dynamic operationalization of the Slippery Slope Framework by embedding power and trust in an evolutionary setting, thereby showing how different SSF compliance regimes can emerge endogenously from micro-level interactions. Second, by translating widely used WGI indicators into an agent-level payoff parameter, we offer a novel mechanism for linking macro-level governance quality to individual tax compliance decisions in a manner that is both empirically grounded and behaviorally interpretable. Third, the identification of critical thresholds, bistable regions, and abrupt regime shifts highlights the fundamentally non-linear nature of the tax compliance ecosystem, challenging linear policy intuitions that treat enforcement and governance improvements as separable, additive levers. These contributions situate our work at the intersection of behavioral tax research, institutional economics, and evolutionary game theory, and open up new avenues for studying how institutional reforms propagate through complex adaptive taxpayer populations.

At the heart of these contributions lies a simple but powerful theoretical claim: institutional trust and deterrence are strategic complements in the evolution of tax compliance.

In our model, increases in T do not merely shift the level of voluntary cooperation; they also enhance the effectiveness of a given fine rate by raising the productivity of compliant behavior relative to evasion. Conversely, when trust is eroded, even large increases in f yield rapidly diminishing returns, as taxpayers no longer perceive the state as a credible provider of public goods. This complementarity mirrors the core intuition of the SSF, but our results go further by quantifying how the marginal impact of coercive power depends on the underlying governance environment. Framing trust and power as strategic complements rather than substitutes therefore provides a unifying lens for interpreting the simulated compliance regimes and suggests a reorientation of tax compliance theory towards the joint design of institutional quality and enforcement policy.

First, by endogenizing institutional trust within an evolutionary game theory framework, our model provides a dynamic validation of the Slippery Slope Framework (SSF). We have shown that trust and power are not independent but are strategic complements. The effectiveness of deterrence (power) is fundamentally mediated by the level of institutional trust. This finding extends the SSF from a static, conceptual model to a dynamic, predictive one, demonstrating how the "slippery slope" can emerge from the micro-level interactions of adaptive agents.

Second, the model highlights the existence of critical thresholds and phase transitions in the tax compliance ecosystem. The shift from a high-compliance to a low-compliance regime is not always gradual but can occur abruptly when trust falls below a certain level. The bistable regime observed at intermediate levels of trust can be traced back to the spatial and local interaction structure embedded in the model. Under Moore-neighborhood interactions and proportional imitation, clusters of compliant agents can emerge and persist if their local density is sufficiently high to generate above-average payoffs from the public good. In this region of the parameter space, small differences in initial conditions or early stochastic shocks determine whether such cooperative clusters reach the critical size needed to resist invasion by evaders. When clusters remain below this threshold, evasion spreads and the system converges to a low-compliance equilibrium; when clusters surpass it, cooperation becomes self-reinforcing and the population gravitates towards a high-compliance state. The resulting path dependence is therefore not an artifact of global parameter choices, but a direct consequence of how local social environments shape agents' imitation decisions. This mechanism is consistent with earlier work on spatial public goods games, which emphasizes the role of cooperation "islands" in sustaining collective action under noisy evolutionary dynamics.

Our findings therefore resonate with a broader body of work on spatial evolutionary games. In their seminal study, Nowak and May (1992) showed that even simple imitation rules on regular lattices can generate persistent pockets of cooperation surrounded by defectors, leading to complex spatial chaos rather than uniform convergence. Subsequent contributions, such as Helbing and Yu (2009), further demonstrated that success-driven migration and local adaptation can stabilize cooperation "islands" under noisy and heterogeneous conditions. The bistable regime identified in our tax compliance model can be interpreted as a governance-conditioned analogue of these phenomena: depending on initial local densities of compliers and the sequence of early shocks, compliant clusters either attain the critical mass required for self-sustaining public good provision or are gradually eroded by evasive behavior. By embedding a WGI-based trust parameter into this spatial framework, we extend the classic insights of spatial cooperation to an institutional context, showing how variations in governance quality shift the parameter regions in which such cooperative clusters can survive.

This non-linearity suggests that policies aimed at maintaining compliance must be proactive, as recovering from a collapse in trust and compliance can be significantly more difficult than preventing it in the first place.

Third, our findings challenge the purely rational-choice perspective of the Deterrence Theory. The fact that high compliance can be sustained with minimal punishment in high-trust environments indicates that intrinsic motivations, operationalized here as the perceived value of public goods, are a powerful driver of behavior. The model suggests that a more holistic understanding of tax compliance must integrate both economic incentives and socio-psychological factors like trust.

5.2. Policy Implications

At the same time, the scope of this challenge to the pure deterrence view should be interpreted with care. The model abstracts from many institutional and macroeconomic details that shape real-world tax systems: it considers a single, stylized tax instrument, assumes a relatively simple public goods technology, and does not incorporate business cycle fluctuations, tax competition, or administrative capacity constraints beyond those captured by the trust parameter. As a result, our conclusions are best understood as applying to settings in which the main margin of policy variation lies in the perceived quality of core fiscal and legal institutions, rather than in the fine-grained design of tax schedules or multi-level revenue-sharing arrangements. Recognizing these boundaries does not weaken the central message that trust matters; instead, it clarifies that the model provides a conceptual and exploratory tool for thinking about the joint evolution of trust and deterrence, which should be complemented by more context-specific empirical and institutional analyses.

The findings of this study offer several actionable policy recommendations:

1. Prioritize Governance Quality: The most significant finding is that investing in institutional quality is a highly efficient strategy for fostering tax compliance. Policies aimed at increasing government effectiveness, strengthening the rule of law, and controlling corruption can have a direct and substantial positive impact on voluntary compliance. This shifts the policy focus from a narrow emphasis on enforcement to a broader strategy of building a trustworthy and legitimate state.
2. The notion of "efficiency" can be illustrated more concretely by comparing parameter combinations that yield similar compliance outcomes but rely on different mixes of trust and punishment. For example, our simulations indicate that maintaining a long-run compliance level of approximately $PC^* = 0.8$ is possible under two contrasting regimes: a medium-trust, moderate-fine regime with $T = 0.6$ and $f = 3.0$, and a low-trust, high-fine regime with $T = 0.4$ and $f = 7.0$. While both regimes achieve comparable compliance, the latter requires substantially higher sanction intensity and, in practice, would entail larger administrative and political costs associated with frequent audits, legal disputes, and potential resistance to harsh enforcement. By contrast, improvements in governance quality that raise T from 0.4 to 0.6 reduce the required fine rate by more than half, suggesting a strong leverage effect of institutional reforms. From a policy perspective, this comparison supports the view that investments in governance and institutional trust are not merely normatively desirable, but also cost-effective substitutes for permanently escalating deterrence.
3. This comparison can be expressed more generally in terms of a simple policy cost function. Let the total cost of maintaining a target compliance level be approximated by $\text{Cost} = C_T \times \Delta T + C_f \times \Delta f$, where ΔT and Δf denote the deviations of institutional trust and the fine rate from some baseline configuration, and C_T and C_f represent the unit costs of improving governance quality and tightening enforcement, respectively. While our agent-based model does not explicitly calibrate these cost coefficients, the simulated regimes in Figure 3 suggest that, for a wide range of parameter values, relatively modest increases in T can substitute for very large increases in f in sustaining a given PC^*. Under plausible assumptions that institutional reforms entail fixed but amortizable investments, whereas aggressive enforcement requires recurrent administrative and political expenditures, it becomes reasonable to expect $C_T \times \Delta T < C_f \times \Delta f$ for many relevant scenarios. The conceptual cost function thus provides a compact way of summarizing why, in the long run, governance improvements are likely to be a more efficient lever than permanently escalating fines.
4. Tailor Enforcement Strategies to the Institutional Context: A one-size-fits-all approach to tax enforcement is likely to be ineffective. In high-trust countries, a light-touch regulatory approach, focusing on service and facilitation, may be sufficient. In contrast, low-trust countries may require a stronger emphasis on deterrence, but policymakers should be aware that this is a costly and potentially unsustainable strategy in the long run. The goal should be to use enforcement as a bridge while simultaneously working to build trust.
5. Beware the Bistable Regime: Policymakers in countries within the bistable regime must be particularly cautious. In this fragile state, small policy changes or external shocks can have a dramatic impact on compliance. It is crucial to monitor levels of trust and compliance closely and to implement policies that can "tip" the system towards the high-compliance equilibrium, such as targeted transparency initiatives or public information campaigns highlighting the effective use of tax revenues.

5.3. Limitations and Future Research

This study has several limitations that open up avenues for future research. First, our model uses a simplified, composite index for institutional trust. Future research could disaggregate the components of the WGI to explore how different dimensions of governance (e.g., corruption vs. government effectiveness) differentially impact compliance. Second, the model assumes a homogeneous population in terms of risk aversion and social preferences. Introducing agent heterogeneity could provide a more nuanced understanding of compliance dynamics.

Rather than viewing these limitations as mere caveats, they point to a broader research agenda on the co-evolution of institutional trust and tax morale. Disaggregating the WGI dimensions would, for instance, allow future models to differentiate between the effects of corruption control, administrative effectiveness, and political accountability on compliance dynamics. Introducing heterogeneous agents with varying degrees of risk aversion, social preferences, and prior trust could help explain why similar institutional reforms generate uneven behavioral responses across taxpayer groups. Finally, embedding the model in explicit social networks would make it possible to study how local social norms and institutional trust interact whether strong pro-compliance norms can compensate for weak formal institutions, or conversely, whether institutional reforms accelerate the diffusion of cooperative norms through peer effects. By systematically extending the present framework along these dimensions, future work can move towards a multi-layered theory of tax compliance that integrates

institutional quality, social structure, and individual heterogeneity. Finally, the model does not include social networks, which are known to play a significant role in shaping tax morale. Integrating network structures into the ABM would be a valuable next step.

6. CONCLUSION

Linking these extensions more explicitly to policy design can further enhance the relevance of the research agenda. Heterogeneous agents with different risk attitudes, income levels, and baseline trust could be used to explore how differentiated audit schemes or targeted communication campaigns affect distinct taxpayer segments, moving beyond "one-size-fits-all" enforcement. Similarly, embedding the model in empirically grounded social networks would allow analysts to study when strengthening pro-compliance norms in key communities can compensate for weak formal institutions, and when institutional reforms are needed to unlock the full potential of peer effects. In combination, these extensions would pave the way for a micro-founded theory of differentiated tax policy, in which the optimal mix of deterrence and trust-building instruments is contingent on the joint distribution of individual characteristics and social ties. Such a framework would bring the model closer to the practical challenges faced by revenue authorities seeking to design context-sensitive, yet analytically well-founded, compliance strategies.

This study has developed an Agent-Based Model that integrates the Public Goods Game with a punishment mechanism and a real-world measure of institutional trust derived from the World Bank Governance Indicators. Our simulations demonstrate that institutional trust is a critical factor in the evolutionary dynamics of tax compliance, acting as a strategic complement to traditional deterrence measures. We have shown that high trust can sustain high compliance with minimal enforcement, while low trust renders even high levels of punishment inefficient. The findings provide a dynamic validation of the Slippery Slope Framework and highlight the non-linear, path-dependent nature of tax compliance.

The key takeaway for policymakers is that building and maintaining institutional trust is not a "soft" policy goal but a fundamental prerequisite for a stable and efficient tax system. A long-term strategy focused on improving governance quality by enhancing transparency, controlling corruption, and delivering high-quality public services is likely to be far more effective and sustainable than a strategy based solely on coercive power.

Ethics: There are no ethical issues with the publication of this manuscript.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare that there is no conflict of interest.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics, 1*(3–4), 323–338. [CrossRef]

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance, 19*(1), 54–77. [CrossRef]

Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature, 36*(2), 818–860.

Bloomquist, K. M. (2011). Tax compliance as an evolutionary coordination game: An agent-based approach. *Public Finance Review, 39*(1), 25–49. [CrossRef]

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences, 99*(Suppl 3), 7280–7287. [CrossRef]

Djayasinga, N., & Ariyanto, D. (2019). The effect of corruption on tax compliance: A meta-analysis. *Jurnal Akuntansi, 12*(1), 1–12.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994. [CrossRef]

Frey, B. S., & Torgler, B. (2007). Tax morale and conditional cooperation. *Journal of Comparative Economics, 35*(1), 136–159. [CrossRef]

Gangl, K., Hofmann, E., & Kirchler, E. (2015). Tax authorities' interaction with taxpayers: A conception of compliance in social dilemmas by power and trust. *New Ideas in Psychology, 37*, 13–23. [CrossRef]

Hardin, G. (1968). The tragedy of the commons. *Science, 162*(3859), 1243–1248. [CrossRef]

Helbing, D., & Yu, W. (2009). The outbreak of cooperation among success-driven individuals under noisy conditions. *Proceedings of the National Academy of Sciences, 106*(10), 3680–3685. [CrossRef]

International Monetary Fund. (2023). *World economic outlook database.* International Monetary Fund.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law, 3*(2), 220–246. [CrossRef]

Kirchler, E. (2007). *The economic psychology of tax behaviour.* Cambridge University Press. [CrossRef]

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology, 29*(2), 210–225. [CrossRef]

Levi, M. (1997). *Consent, dissent, and patriotism.* Cambridge University Press. [CrossRef]

Li, X., Li, K., & Chen, X. (2022). Replicator dynamics of public goods game with tax-based punishment. *Chaos, Solitons & Fractals, 164*, 112747. [CrossRef]

Luttmer, E. F. P., & Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives, 28*(4), 149–168. [CrossRef]

Masad, D., & Kazil, J. (2015). Mesa: An agent-based modeling framework. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 14th Python in Science Conference (SciPy 2015)* (pp. 53–60). [CrossRef]

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature, 359*(6398), 826–829. [CrossRef]

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge University Press. [CrossRef]

Pickhardt, M., & Prinz, A. (2014). Behavioral dynamics of tax evasion – A survey. *Journal of Economic Psychology, 40*, 1–19. [CrossRef]

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science, 325*(5945), 1272–1275. [CrossRef]

Shen, Y., Lei, W., Kang, H., Li, M., Sun, X., & Chen, Q. (2023). Evolutionary dynamics of public goods game with tax-based rewarding cooperators. *Chaos, Solitons & Fractals, 175*, 114030. [CrossRef]

Slemrod, J. (Ed.). (1992). *Why people pay taxes: Tax compliance and enforcement.* University of Michigan Press.

World Bank. (2023a). *World development indicators.* World Bank. https://data.worldbank.org/ Accessed on Jun 26, 2026.

World Bank. (2023b). *Worldwide governance indicators.* World Bank. https://www.worldbank.org/en/publication/worldwide-governance-indicators Accessed on Jun 26, 2026.

Yıldız Social Science Review
Web site information: https://yssr.yildiz.edu.tr
DOI: 10.51803/yssr.1830824

Original Article / Orijinal Makale

Green Supply Chain Management and Artificial Intelligence: Bibliometric Analysis Based on Web of Science (WoS) Platform*

Yeşil Tedarik Zinciri Yönetimi ve Yapay Zekâ: Web of Science (WoS) Platformuna Dayalı Bibliyometrik Analiz

Şermin ÖNEM

Department of E-Commerce and Marketing, İstanbul Atlas University, İstanbul, Türkiye

İstanbul Atlas Üniversitesi, E-Ticaret ve Pazarlama Programı, İstanbul, Türkiye

ARTICLE INFO

Article history
Received: 27 November 2025
Revised: 06 April 2026
Accepted: 07 April 2026

Key words:
Artificial intelligence, bibliometric analysis, green supply chain management, supply chain management, sustainability

MAKALE BİLGİSİ

Makale Hakkında
Geliş tarihi: 27 Kasım 2025
Revizyon tarihi: 06 Nisan 2025
Kabul tarihi: 07 Nisan 2026

Anahtar kelimeler:
Yapay zekâ, bibliyometrik analiz, yeşil tedarik zinciri yönetimi, tedarik zinciri yönetimi, sürdürülebilirlik

ABSTRACT

This study aims to provide a comprehensive bibliometric analysis of the intersection between green supply chain management (GSCM) and artificial intelligence (AI). As environmental sustainability and digital transformation have become critical priorities for businesses, the integration of AI into GSCM practices has attracted increasing scholarly attention. In this context, the study analyzes the intellectual structure, development trends, and research patterns in this emerging field. Bibliometric data were collected from the Web of Science (WoS) database using relevant keywords, and the dataset was analyzed using the Biblioshiny application based on RStudio. The findings reveal a significant increase in the number of publications, particularly after 2017, indicating growing academic interest driven by sustainability concerns and technological advancements. The analysis also identifies the most influential authors, keywords, and thematic clusters, highlighting the evolution of research from conceptual discussions to more application-oriented studies focusing on performance, sustainability, and digital transformation. The results suggest that AI plays a transformative role in enhancing the efficiency, transparency, and environmental performance of supply chains. This study contributes to the literature by offering a structured overview of the research landscape and providing insights for future academic and practical developments in the field of AI-driven sustainable supply chain management.

Cite this article as: Önem Ş. (2026). Green Supply Chain Management and Artificial Intelligence: Bibliometric Analysis Based on Web of Science (WoS) Platform. *Yıldız Social Science Review, 12*(1), 15–27.

*The abstract of this study was presented as an Abstract Paper at the 11th International EMI Entrepreneurship & Social Sciences Congress in Istanbul, Turkey, from August 31 to September 3, 2025.

***Sorumlu yazar / Corresponding author**
*E-mail address: sermin.onem@atlas.edu.tr

ÖZ

Bu çalışma, yeşil tedarik zinciri yönetimi (GSCM) ile yapay zekâ (AI) arasındaki kesişimi bibliyometrik analiz yöntemiyle kapsamlı bir şekilde incelemeyi amaçlamaktadır. Günümüzde çevresel sürdürülebilirlik ve dijital dönüşüm, işletmeler açısından kritik öncelikler haline gelmiş olup, yapay zekânın GSCM uygulamalarına entegrasyonu akademik literatürde giderek artan bir ilgi görmektedir. Bu bağlamda çalışmada, söz konusu araştırma alanının entelektüel yapısı, gelişim eğilimleri ve araştırma desenleri analiz edilmiştir. Bibliyometrik veriler, ilgili anahtar kelimeler kullanılarak Web of Science (WoS) veri tabanından elde edilmiş ve veriler RStudio tabanlı Biblioshiny uygulaması aracılığıyla analiz edilmiştir. Elde edilen bulgular, özellikle 2017 yılı sonrasında yayın sayısında belirgin bir artış olduğunu ve bu artışın sürdürülebilirlik odaklı yaklaşımlar ile teknolojik gelişmelerden kaynaklandığını göstermektedir. Ayrıca analiz sonuçları, en etkili yazarları, anahtar kelimeleri ve tematik kümeleri ortaya koyarak, araştırmaların kavramsal tartışmalardan performans, sürdürülebilirlik ve dijital dönüşüm odaklı uygulama alanlarına doğru evrildiğini göstermektedir. Bulgular, yapay zekânın tedarik zincirlerinde verimlilik, şeffaflık ve çevresel performansı artırmada dönüştürücü bir rol oynadığını ortaya koymaktadır. Bu çalışma, ilgili literatüre sistematik bir genel çerçeve sunarak hem akademik araştırmalar hem de uygulayıcılar için yol gösterici nitelikte katkılar sağlamaktadır

Atıf için yazım şekli: Önem Ş. (2026). Green Supply Chain Management and Artificial Intelligence: Bibliometric Analysis Based on Web of Science (WoS) Platform. *Yıldız Social Science Review, 12*(1), 15–27.

1. INTRODUCTION

In today's global business environment, sustainability has become a central concern for organizations seeking to maintain competitiveness while minimizing their environmental impact. The increasing pressure from stakeholders, regulatory bodies, and consumers has compelled firms to adopt environmentally responsible practices across their operations. In this context, green supply chain management (GSCM) has emerged as a strategic approach that integrates environmental thinking into supply chain activities, including product design, sourcing, production, and logistics.

At the same time, rapid technological advancements have accelerated the digital transformation of supply chains. Among these technologies, artificial intelligence (AI) has gained significant attention due to its potential to enhance decision-making processes, improve operational efficiency, and enable real-time monitoring of complex supply chain networks. AI-driven tools such as machine learning, big data analytics, and predictive modeling offer new opportunities to optimize resource usage, reduce emissions, and support sustainable supply chain practices.

Despite the growing body of research on both GSCM and AI, studies examining the intersection of these two domains remain fragmented and lack a comprehensive analytical perspective. Existing research has primarily focused on conceptual discussions or empirical applications, while a systematic mapping of the intellectual structure and evolution of this emerging field is still limited. This gap makes it difficult for researchers and practitioners to understand the current state of knowledge, identify dominant research themes, and recognize future research directions.

To address this gap, this study aims to provide a comprehensive bibliometric analysis of the intersection between GSCM and AI. By analyzing data obtained from the Web of Science (WoS) database, the study reveals the development trends, key contributors, thematic structures, and research patterns in this field.

This study makes several important contributions to literature. First, it offers a structured and systematic overview of the research landscape at the intersection of GSCM and AI through bibliometric techniques. Second, it identifies the most influential authors, keywords, and thematic clusters, thereby clarifying the intellectual structure of the field. Third, it highlights the evolution of research trends, showing a shift from conceptual discussions toward application-oriented and performance-driven studies. Finally, the study provides valuable insights for both researchers and practitioners by outlining potential future research directions and emphasizing the strategic role of AI in enhancing sustainable supply chain management.

2. CONCEPTUAL FRAMEWORK

GSCM, which originated from the supply chain method, began to emphasize management activities that highlight awareness of management processes, environmental protection, and a sense of responsibility towards nature, particularly in the 1990s, due to increasing competitive pressures on businesses (Chin et al., 2015). GSCM is essen-

tially supply chain management integrated with a green focus. Therefore, GSCM is an important point for businesses that want to create a new generation of business perception that emphasizes management processes that can minimize their impact on the environment and continues its activities with this perspective. Because creating the perception that businesses that emphasize and implement this are "reliable" and "conscious" is extremely important in the global world (Zeeshan & Hajra, 2024: 53).

GSCM can be described as a new generation approach that should be emphasized in order to stand out from increasing competitive pressure, increase efficiency, strengthen the company's brand image, and gain consumers' positive perception of the company (Sharma et al., 2017). In the current scenario, where our planet is suffering from the dramatic consequences caused by pollution, a commitment to sustainability, and especially GSCM, has become imperative for organizations looking to improve their corporate environmental performance (Wiredu et al., 2023). Today, businesses are adding the words "green" and/or "sustainable" to the beginning of terms at every stage of their operations to enhance awareness and brand image throughout the supply chain (Tietze, 2023: 11).

Managers need to develop collaborative relationships based on trust with stakeholders to fill resource and technical knowledge gaps and ensure shared problem-solving that promotes environmental sustainability (Dzhengiz & Niesten, 2020: 888). Sustainable supply chain management improves long-term performance by integrating the social, economic, and environmental objectives of the supply chain. Therefore, it helps monitor performance over the long term by evaluating business performance based on social, environmental, and economic dimensions (Peesapati, 2024: 30). Advanced companies that use green logistics and environmental improvements in supply chains as a strategic weapon aim to achieve environmental goals through green management innovations such as green product and green process innovation by following sustainable competitive strategies (Jo & Kwon, 2022: 3).

GSCM evaluates all production and value-added steps of goods and services from an ecological, social and economic perspective. This type of supply chain is characterized by approaching all stages of the supply chain, from raw material extraction and production to the direct supplier, with a holistic perspective. The goal is to minimize or even eliminate negative effects and prevent risks in a timely manner. The aim is to promote responsible and sustainable corporate strategy and management (Hass, 2022: 15). Combining environmental philosophy with supply chain management and organizational practices, supply chain management is a multitude of synchronized efforts carried out by companies to intensify their impact on the regular environment throughout the entire life cycle of a product, including raw material sourcing and selection, product design, manufacturing, finished product distribution, and end-of-life product management processes (Perotti et al., 2012).

Based on their definitions in the literature, Liu & Chang (2017: 4) show GSCM in Figure 1. The solid and dashed lines represent the greening of traditional supply chain and reverse logistics activities, respectively.

A green supply chain can be defined as the use of materials needed by the industry in a way that is environmentally friendly and can be used or reprocessed according to corporate needs (Nguyen & Le, 2020). GSCM refers to the way of managing the supply chain in order to reduce Co2 emissions and waste and protect biodiversity (Tseng et al., 2019). According to another definition, GSCM refers to the consideration of environmental impacts at all stages of the chain, starting from the purchasing process, which is the first

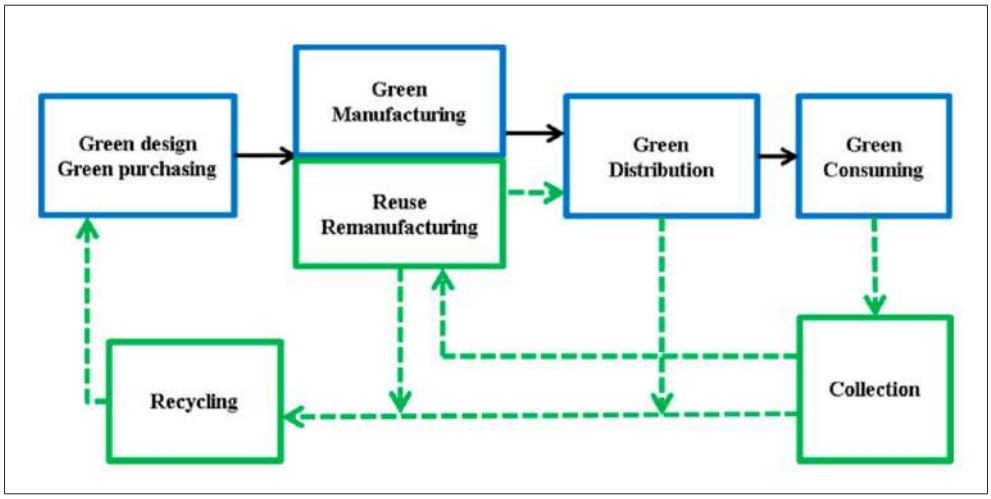

Figure 1. Green supply chain management.
Source: Liu & Chang (2017, s.4).

stage of production, in the process of producing the goods and services needed by the customer, to the recycling of the product after the consumer uses the product (Duran & Şen, 2024: 3). According to Zhu and Sarkis (2004), GSCM; purchasing, operations, marketing and logistics, recycling, reuse, remanufacturing, reverse logistics and monitoring the environmental management process including innovation.

GSCM, integrates environmental factors into supply chain management by covering topics such as "international environmental management," "green procurement," collaboration between customers and the environment, "green logistics," and "ecological design," when compared to traditional supply chain management (Feng et al., 2022). GSCM is defined as an important issue that affects the environmental problems of any business engaged in supply chain activities, which subsequently leads to the improvement of environmental performance (Chin et al., 2015). GSCM aims not only to reveal businesses' environmental sensitivities but also to carefully incorporate competing businesses into their processes. This situation provides businesses with both cost and efficiency advantages in environmental terms (Chu et al., 2017). The development of GSCM should be considered as the inclusion of all elements in the supply chain flow in business processes, approached from an environmental perspective from start to finish (Liu & Ma, 2022; Srivastava, 2007).

The primary objective of GSCM is to conserve resources, reduce pollution, and ensure sustainable environmental development. In addition, both developed and developing countries should actively promote "zero carbon" policies and establish relevant laws and regulations. This will clearly accelerate cooperation and development between the environment and the supply chain, encouraging the implementation of GSCM in different sectors. Therefore, GSCM is a new generation approach applied to strengthen the environmental image of today's businesses, meet public expectations regarding environmental protection, and improve supply chain operations (Lei et al., 2024: 3).

Within the scope of the project funded by the World Bank, the "Türkiye Green Industry Project", which consists of three components: KOSGEB, TUBITAK and the Ministry of Industry and Trade, aims to raise awareness among organizations, companies and individuals related to green transformation, especially businesses in the manufacturing sector, to facilitate networking with various stakeholders and to increase the national visibility of green transformation efforts (Ministry of Industry and Technology, 2024).

Previous research has identified and validated five main components of GSCM practices: Green purchasing, eco-design or design for the environment, investment recovery, customer collaboration for environmental concerns, and internal environmental management (Mastos Gotzamani, 2024; Zahran, 2024: 4). In addition, topics such as carbon border adjustments, green finance, green and circular economy, clean, economical, and secure energy supply, sustainable smart transportation, combating climate change, diplomacy, and awareness-raising activities can also be included in the GSCM components. Within the scope of this plan, our country aims to raise awareness and motivate consumers and producers and to adapt to the green transformation process. In addition, countries are going one step further in the fight against climate change and setting net zero emission targets (Tubitak, 2024). To ensure the sustainability of green supply chain management, today's businesses aim to improve management processes not only by emphasizing green practices across all supply chain factors but also by spearheading the implementation of necessary environmental policies (Agrawal et al., 2022).

GSCM aims to ensure the most efficient logistics aspects of the production process in businesses with a green emphasis. This will also affect the parties involved due to the multifaceted perspective resulting from the inclusion of suppliers, manufacturers, customers, and disposal companies in supply chain activities. For this reason, in the context of GSCM, an interactive link is established between traditional supply chain management and green supply chain factors (Hunke & Prause, 2014: 16).

In today's global world, not only developed countries but also developing countries are prioritizing sustainable approaches, and it is possible to say that they are not only expressing this perspective but also putting it into practice throughout their activities. In addition, it is possible to say that they are taking an extremely sensitive approach to environmental issues around the world, and that both competitors and stakeholders are united on this point. For this reason, the public sector also requires the implementation of environmental management systems and certification, which are mandatory for business activities to address environmental issues. This situation enables the formation of certain standards not only for businesses but for all businesses around the world (Jo & Kwon, 2022: 3).

A literature review reveals that green supply chain management positively impacts business performance by promoting sustainability (Awan et al., 2017; Khan et al., 2019; Yıldız Çankaya & Sezen, 2019). According to Liu et al. (2018), the fundamental way for businesses to adopt green supply chain management is through rapid adaptation to innovations. Furthermore, according to Malviya et al. (2018), implementing green supply chain practices in all operational processes, especially in manufacturing companies, is critical for achieving a long-term sustainable perspective. Kafa et al. (2020) prove that companies need to establish supply chain management processes that take into account adequate criteria to achieve corporate sustainability. Table 1 summarizes the role of artificial intelligence in green supply chain applications:

Table 1. Role of AI in green supply chain practices

Green Supply Chain Practice	The Role of AI
Green procurement	Selecting sustainable suppliers based on environmental and social metrics. Automating repetitive procurement tasks, reducing human error and operational costs. Improving transparency and traceability in supply chains. Providing insights into supplier performance, market trends, and demand forecasting.
Green production	Using intelligent robots to perform repetitive production tasks with high precision and efficiency. Monitoring energy consumption and product output in real time. Reducing unnecessary tasks and emissions through real-time analytics.
Green logistics	Forecasting demand and reducing emissions using predictive analytics and machine learning. Automating warehouse systems. Scheduling vehicles and selecting efficient delivery routes to minimize transportation emissions.
Recycling	Using image recognition and machine learning for waste classification. Analyzing product life cycles and supply chain data to identify waste sources and suggest reduction strategies. Supporting circular economy strategies. Simulating scenarios and evaluating environmental impacts to support decision-making regarding circular economy initiatives.

Beniaich & Hmioui, 2025.

Table 1 clearly demonstrates that artificial intelligence plays a critical role across all stages of green supply chain practices, reinforcing its importance as a key driver of sustainable transformation.

The intersection of AI, GSCM, and technology innovation presents a promising avenue for achieving sustainable performance in supply chains. AI can be used to improve the effectiveness and efficiency of GSCM applications, such as optimizing energy consumption in manufacturing processes, predicting demand for eco-friendly products, and designing sustainable packaging solutions. Furthermore, AI can enable the development of innovative technologies that support GSCM, such as autonomous vehicles for low-emission logistics, smart sensors for real-time monitoring of environmental impact, and blockchain-based platforms for transparent and traceable supply chains (Sharma et al. 2022; Taseen et al., 2024).

It provides the following advantages and opportunities for the integration of AI throughout GSCM (Beniaich & Hmioui, 2025; Fowosere et al., 2025: 329):

- AI-powered optimization of logistics and transportation to reduce emissions and energy consumption,
- AI-powered predictive maintenance and asset management to improve efficiency and extend equipment life,
- AI-based supply chain visibility and traceability to enable better monitoring and reporting of environmental impact,
- AI-powered decision support systems to help make more informed and sustainable choices throughout the supply chain,
- With the development of AI applications, they are becoming more comprehensive, useful, and accessible for business applications.

Overall, the integration of artificial intelligence into green supply chain management represents a significant paradigm shift in how organizations address sustainability challenges. While traditional supply chain approaches primarily focus on cost efficiency and operational performance, AI-driven systems enable predictive, adaptive, and data-driven decision-making processes. This transformation allows organizations to move beyond reactive environmental strategies toward proactive and optimized sustainability practices. In this context, the convergence of AI and GSCM not only enhances operational efficiency but also strengthens long-term environmental performance and competitive advantage, positioning AI as a strategic enabler of sustainable supply chain transformation.

3. MATERIALS AND METHODS

This section presents the purpose of the research, the analyses conducted, and the findings.

3.1. The Purpose of the Research

The primary aim of this study is to examine the intellectual structure, development trends, and thematic evolution of research conducted at the intersection of GSCM and AI. In this context, the study seeks to provide a comprehensive overview of the research landscape by identifying prominent concepts, influential authors, collaboration patterns, and citation structures within the existing literature.

3.2. Data and Analysis

This study employs a bibliometric analysis approach to systematically evaluate the existing body of literature. Bibliometric analysis is widely used to map scientific knowledge, identify research trends, and analyze the intellectual structure of a specific field.

The data used in this study were obtained exclusively from the Web of Science (WoS) database, which is recog-

nized as one of the most reliable and comprehensive sources of high-quality academic publications. WoS was preferred due to its rigorous indexing standards and its ability to provide structured and consistent bibliographic data.

To ensure a comprehensive dataset, multiple keywords were used during the data collection process. The search query included combinations of terms such as "green supply chain management," "artificial intelligence," and "sustainable supply chain." The search was limited to articles published in English within the Business and Management categories.

Following the initial search, the dataset was refined by applying inclusion criteria such as document type (articles), language (English), and subject area. The final dataset consisted of 151 articles covering the period between 2008 and 2025. The data collection process was completed on November 20, 2025, to ensure consistency and replicability.

For data analysis, the Biblioshiny application based on RStudio was used. This tool enables advanced bibliometric analysis, including data visualization, citation analysis, keyword co-occurrence, and thematic mapping. Through this analysis, the study identifies key research trends, influential contributors, and the evolution of the field over time (Table 2).

4. FINDINGS

This section presents the findings derived from the bibliometric analysis of 151 scientific studies conducted at the intersection of green supply chain management (GSCM) and artificial intelligence (AI).

Figure 2 illustrates the annual scientific production in this research field. The findings indicate that although the first studies appeared around 2008, a significant increase in publication volume has been observed, particularly after 2017. This upward trend suggests a growing academic interest driven by increasing global concerns about sustainability, climate change, and the need for digital transformation in supply chain processes. The acceleration in recent years also reflects the expanding role of artificial intelligence technologies in addressing complex environmental and operational challenges.

Figure 3 presents the average number of citations per year. The results show a notable increase in citations until around 2020, followed by a slight decline in subsequent years. This pattern can be interpreted as a typical citation lifecycle, where earlier foundational studies accumulate more citations over time, while more recent studies have not yet reached their full citation potential. It also indicates that the field has reached a certain level of maturity, with well-established core studies shaping subsequent research.

The word cloud presented in Figure 4 highlights the most frequently used terms in the studies analyzed. Keywords such as "artificial intelligence," "performance," "framework," "management," and "supply chain management" emerge as dominant concepts. This finding suggests that literature has evolved beyond purely conceptual discussions and increasingly focuses on performance measurement, strategic frameworks, and practical implementation issues.

Figure 5, which presents the tree map analysis, provides a more detailed and structured view of keyword frequency. The prominence of terms such as artificial intelligence and performance indicates that researchers are increasingly interested in examining the tangible outcomes of AI integration within GSCM practices. This shift reflects a transition from theoretical exploration to application-oriented research, emphasizing efficiency, sustainability, and operational performance.

Figure 6 presents the three-field plot analysis, which links authors, keywords, and sources. The findings reveal a concentration of research around specific authors and thematic areas,

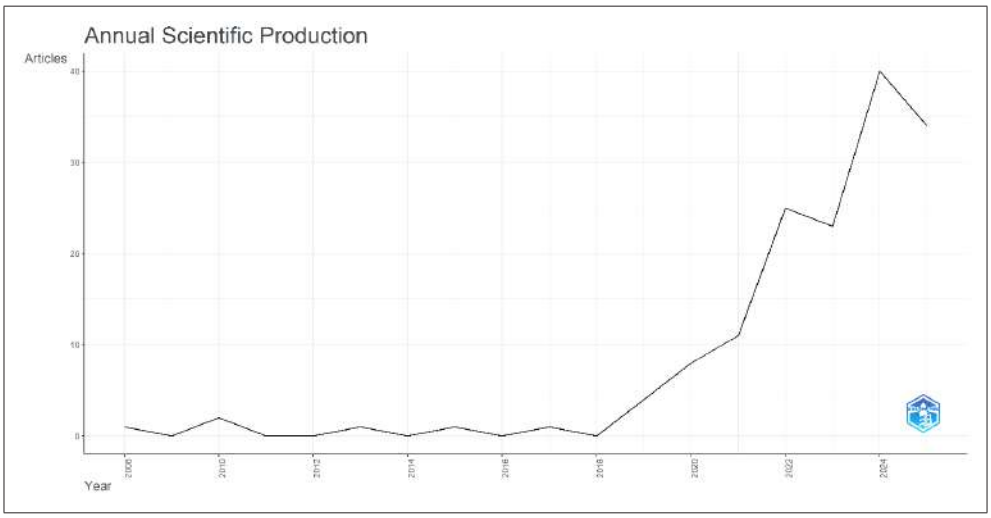

Figure 2. Annual scientific production.

Table 2. Descriptive statistics

Publication year range of articles	**2008-2025**	**Keywords plus (ID)**	**334**
Documents	151	Author's keywords (DE)	515
Annual growth rate %	%23.05	Authors	504
Document average age	2.29	Authors of single-authored docs	18
Average citations per doc	47.45	Co-Authors per Doc	3.75
References	10231	International co-authorships %	%42.38
Web of Sciences			

indicating the presence of influential contributors shaping the intellectual structure of the field. This clustering also suggests the emergence of specialized research streams focusing on AI-driven sustainability and supply chain innovation.

Finally, Figure 7 illustrates the country collaboration network. The results indicate a high level of international collaboration, particularly among developed and emerging economies. This suggests that sustainability and AI-driven supply chain management are global concerns requiring cross-border knowledge exchange and cooperation. The presence of international co-authorship further highlights the interdisciplinary and collaborative nature of this research domain.

Overall, the findings demonstrate that the intersection of GSCM and AI is a rapidly growing and evolving research area. The increasing number of publications, diversification of research themes, and strong international collaboration patterns indicate that this field will continue to expand in the coming years.

In addition to the general trends, further analysis was conducted to identify the most influential studies and sources within the field. Figure 8 presents the most globally cited documents. The findings indicate that a limited number of studies have accumulated a significantly higher number of citations compared to others, highlighting their foundational role in shaping the intellectual structure of the research domain. These highly cited works predominantly focus on the integration of artificial intelligence into supply chain processes and sustainability practices, providing both conceptual frameworks and empirical insights that guide subsequent studies.

In addition, the top 10 most globally cited documents are presented in Table 3. These studies represent the core knowledge base of the field and have significantly influ-

Figure 4. Word cloud.

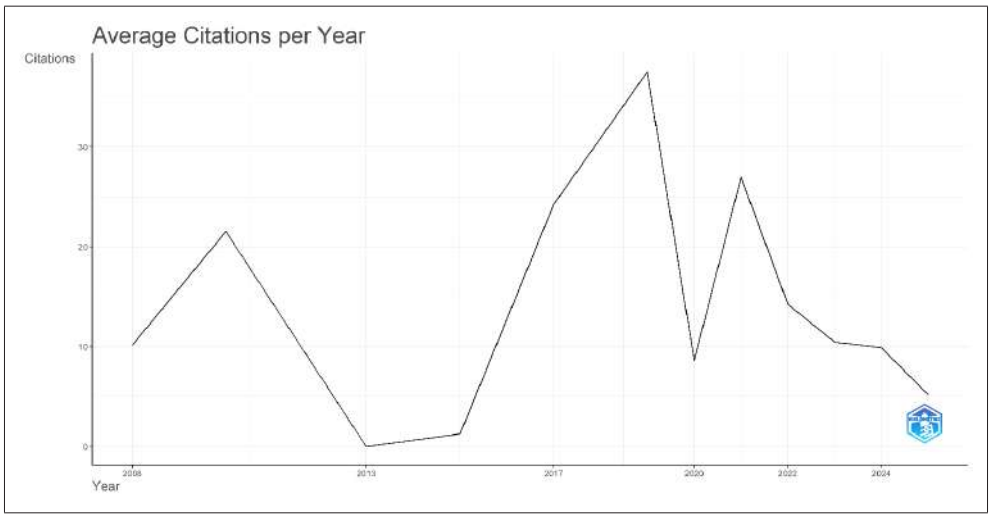

Figure 3. Avarage citations per year.

Table 3. Top 10 most globally cited documents

Rank	Authors	Year	Source	Global Citations
1	Benzidia S.	2021	Technological Forecasting and Social Change	308
2	Dhamija P.	2020	Total Quality Management Journal	177
3	Feng Y.	2022	International Journal of Production Economics	123
4	Sharma R.	2022	International Journal of Production Research	103
5	Bag S.	2021	Journal of Enterprise Information Management	88
6	Ninlawan C.	2010	International Conference (Engineering & Computer)	73
7	Long Q.	2021	IEEE Transactions on Evolutionary Computation	68
8	Zhang Q.	2022	Technological Forecasting and Social Change	48
9	Wang J.	2020	Journal of Cleaner Production	43
10	Wang J.	2020	Computers & Industrial Engineering	42

Web of Sciences

enced subsequent research. The concentration of citations among a limited number of publications indicates that the field is shaped by a set of seminal works, which provide both theoretical foundations and methodological guidance for future studies.

Furthermore, Figure 9 presents the most relevant sources contributing to this research area. The results reveal that journals such as *Sustainability*, *Environmental Science and Pollution Research*, and *International Journal of Production Economics* stand out as the most productive outlets. This concentration suggests that knowledge production in this field is clustered within a limited number of high-impact journals. These findings provide important guidance for future researchers in identifying suitable publication outlets and understanding the core dissemination channels of the field.

Overall, these results provide a comprehensive overview of the intellectual, thematic, and publication structure of the field, offering a solid foundation for further discussion on the implications, research gaps, and future directions.

Figure 5. Tree map.

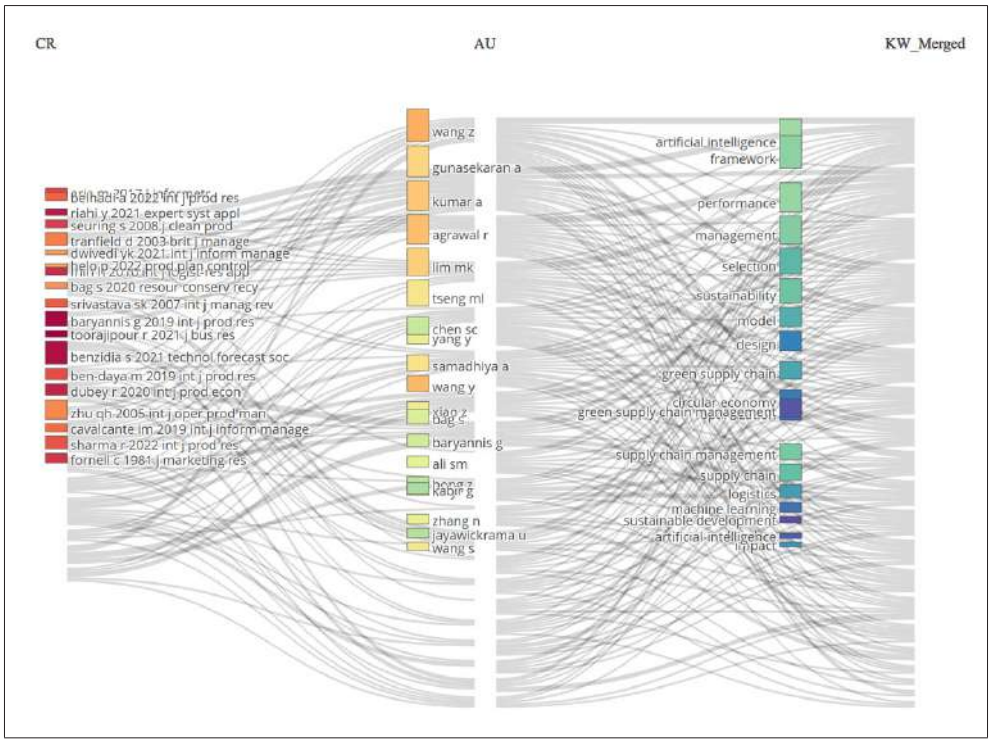

Figure 6. Three-field plot.

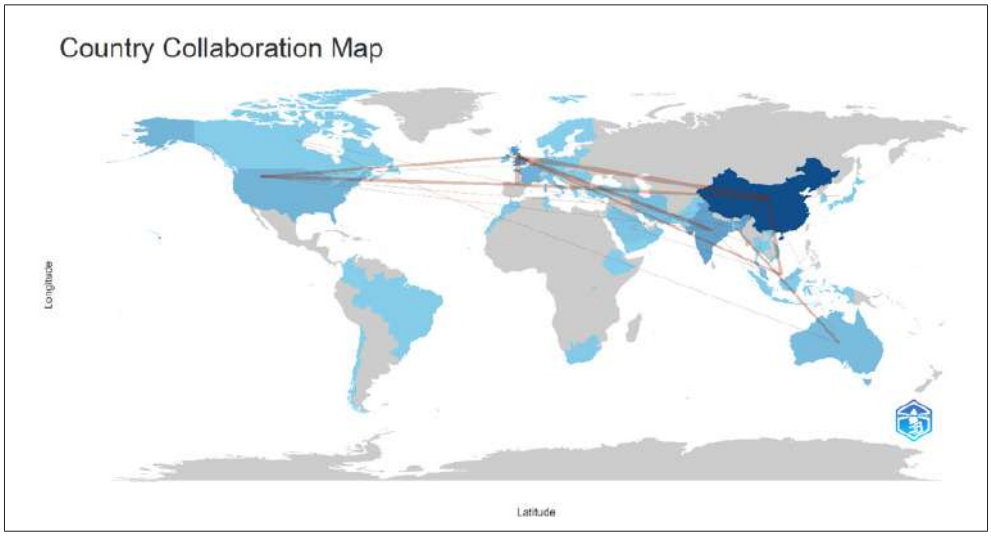

Figure 7. Country collaboration map.

5. RESULTS & DISCUSSION

This study aims to provide a comprehensive bibliometric analysis of the intersection between green supply chain management (GSCM) and artificial intelligence (AI). Considering the environmental challenges faced by today's world, businesses are required to adopt proactive strategies and develop sustainable policies supported by accurate data-driven insights. In this context, GSCM has become a critical approach for achieving both environmental protection and long-term organizational performance.

The findings indicate that the integration of AI into supply chain processes enables organizations to monitor environmental performance in real time, optimize resource allocation, and ensure compliance with sustainability regulations. AI-driven systems enhance decision-making capabilities by processing large volumes of data efficiently, thereby supporting the balance between economic and environmental objectives (Fowosere et al., 2025: 329). In this

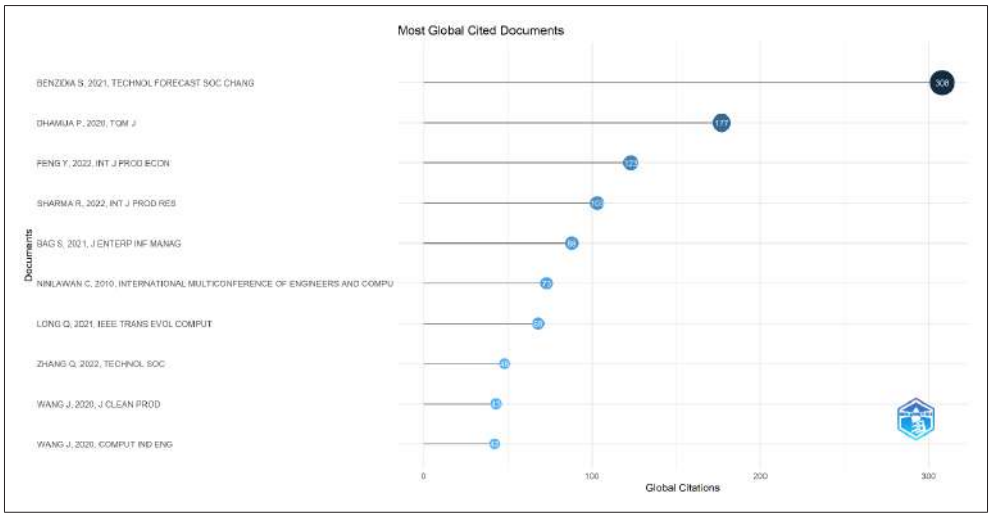

Figure 8. Most global cited documents.

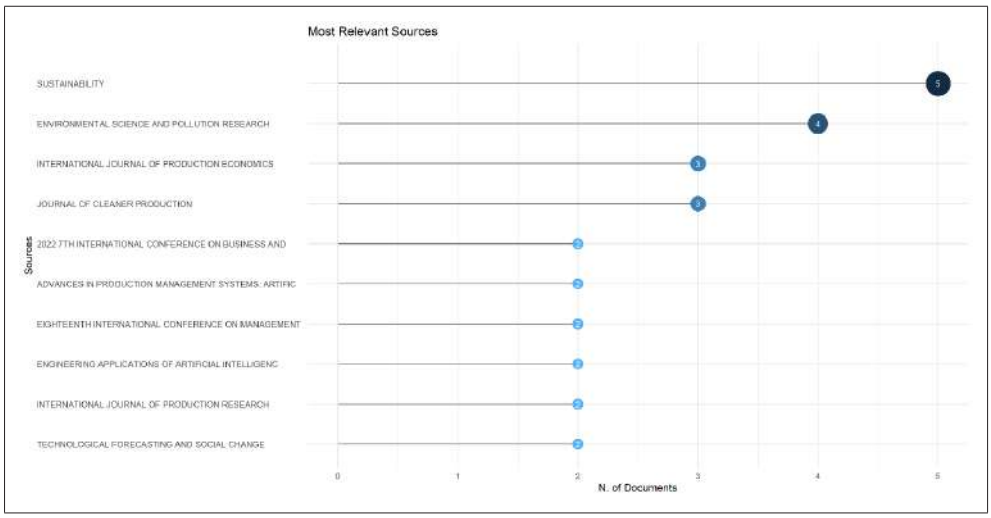

Figure 9. Most relevant sources.

regard, the growing adoption of AI technologies plays a key role in improving sustainability performance and operational efficiency (Chen, 2025).

The bibliometric results reveal that although early studies on AI and GSCM emerged around 2008, there has been a significant increase in publication output in recent years. This rapid growth reflects the rising importance of sustainability concerns and the accelerating impact of digital transformation in supply chain management. Moreover, the findings suggest a transition from theoretical discussions to more application-oriented research areas such as operational efficiency, circular economy practices, and sustainability-driven digital transformation. This shift indicates that the field has reached a certain level of maturity and is increasingly aligned with real-world industrial needs.

Consistent with existing literature, the findings show that traditional performance indicators such as cost and return on investment are gradually being replaced by broader strategic concerns such as environmental sustainability, resilience, and corporate reputation (Fowosere et al., 2025: 329). This transformation is supported by the findings of Gallo et al. (2023), who argue that AI capabilities enhance supply chain visibility and facilitate large-scale decision-making, particularly in addressing environmental challenges. Similarly, Singh et al. (2024) emphasize that AI-driven big data analytics significantly contributes to optimizing supply chain processes and maintaining environmentally sustainable operations. Supporting this perspective, Nozari (2024) and Makhdoom et al. (2025) demonstrate that AI adoption positively influences sustainable performance, highlighting the strategic importance of technological investments.

Furthermore, the results suggest that the implementation of AI-driven GSCM practices is highly influenced by contextual factors such as regulatory environments and

firm-specific characteristics. As highlighted by Handoyo (2023; 2024), variations in national regulations, industry dynamics, and organizational structures play a crucial role in shaping sustainability practices. Therefore, a deeper understanding of these contextual factors is essential for both researchers and practitioners aiming to enhance sustainability outcomes.

Despite the significant advantages of integrating AI into GSCM, several challenges remain. These include high implementation costs, data security concerns, lack of technical expertise, and difficulties in integrating AI technologies into existing supply chain infrastructures. Additionally, the environmental impact of AI systems themselves—particularly in terms of energy consumption—should not be overlooked. These challenges highlight the need for a balanced and strategic approach to AI implementation, ensuring that technological advancements do not undermine broader sustainability goals.

This study also makes important theoretical contributions by providing a structured overview of the intellectual landscape of AI-driven GSCM. By identifying influential studies, dominant journals, and emerging research themes, the study offers a valuable roadmap for future research. In particular, the identification of highly cited works provides a strong foundation for scholars seeking to build upon established knowledge structures.

However, this study is not without limitations. Bibliometric analyses are inherently descriptive and may not fully capture the qualitative depth of individual studies. As emphasized in the literature, such analyses reflect the dataset used and may not represent the most recent developments in the field. Additionally, the reliance on a single database may limit the comprehensiveness of the analysis (Handoyo, 2024). Future studies could address these limitations by incorporating multiple databases and complementary research methods such as systematic literature reviews or empirical analyses.

6. CONCLUSION

In conclusion, the findings of this study demonstrate that the integration of artificial intelligence into green supply chain management is a rapidly evolving and strategically important research area. The increasing number of publications, the emergence of influential studies, and the growing emphasis on practical applications indicate that this field will continue to expand in the coming years. Future research is encouraged to build upon the highly cited works identified in this study and to explore emerging themes such as AI ethics, transparency, and sustainability-oriented innovation at both organizational and national levels.

Ethics: There are no ethical issues with the publication of this manuscript.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare that there is no conflict of interest.

Authorship Contributions: The authors contributed to the study equally.

Financial Disclosure: The authors declared that this study has received no financial support

REFERENCES

Agrawal V, Mohanty R. P., Agarwal S., Dixit J. K. & Agrawal A.M. (2022). Analyzing critical success factors for sustainable green supply chain management. *Environment, Development and Sustainability, 25,* 8233–8258. [CrossRef]

Alsheibani, S., Messom, D., Cheung, Y., & Alhosni, M. (2020). *Reimagining the strategic management of artificial intelligence: Five recommendations for business leaders*. Americas Conference on Information Systems. [CrossRef]

Awan, U., Kraslawski, A., & Huiskonen, J. (2017). Understanding the relationship between stakeholder pressure and sustainability performance in manufacturing firms in Pakistan. *Procedia Manufacturing*, *11*, 768–777. [CrossRef]

Beniaich, R. & Hmioui, A. (2025). Artificial Intelligence in Green Supply Chain Management: a Literature Review. *International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA 2025)*. [CrossRef]

Bytniewski, A., Matouk, K., Chojnacka-Komorowska, A., Hernes, M., Zawadzki, A., & Kozina, A. (2020). *The functionalities of cognitive technology in management control system.* Asian Conference on Intelligent Information and Database Systems. [CrossRef]

Chen, F. (2025). Optimization of Thermal Energy and Logistics Innovation in Flexible Manufacturing Systems Based on Artificial Intelligence: Green Supply Chain Management. *International Journal of System Assurance Engineering and Management*. Preprint. https://doi.org/10.1007/s13198-025-02940-z. [CrossRef]

Chin, T.A., Tat, H.H. & Sulaiman, Z. (2015). Green Supply Chain Management, Environmental Collaboration and Sustainability Performance. *Procedia CIRP, 26*(Supplement C), 695–699. [CrossRef]

Chu, S., Yang, H., Lee, M. & Park, S. (2017). The Impact of Institutional Pressures on Green Supply Chain Management and Firm Performance: Top Management Roles and Social Capital. S*ustainability*, *9*, 764. [CrossRef]

Dzhengiz T. & Niesten E. (2020). Competences for environmental sustainability: A systematic review on

the impact of absorptive capacity and capabilities. *Journal of Business Ethics*, *162,* 881–906. [CrossRef]

Duran Ö. & Şen Y. (2024). A field study on green supply chain management. *International Journal of Social Sciences Troyacademy*, *9*(1), 1–23. [Turkish] [CrossRef]

Enholm, I. M., Papagiannidis E., Mikalef P. & Krogstie J. (2022). Artificial intelligence and business value: a literature review. *Information Systems Frontiers, 24*, 1709–1734. [CrossRef]

Feng, Y., Lai, K.H. & Zhu, Q. (2022). Green supply chain innovation: emergence, adoption and challenges. *International Journal of Production Economics*, *248*(1), 108497. [CrossRef]

Fowosere, S., Esechie O. C., Namboozo, S. & Anwansedo F. (2025). The Role of Artificial Intelligence in Green Supply Chain Management. *International Journal of Latest Technology in Engineering, Management & Applied Science, 14*(2), 325–331. [CrossRef]

Gallo H., Khadem A. & Alzubi A. (2023). The relationship between big data analytic-artificial intelligence and environmental performance: A moderated mediated model of green supply chain collaboration (GSCC) and top management commitment (TMC). *Discrete Dynamics in Nature and Society*, *2023,* 4980895. [CrossRef]

Handoyo, S. (2024) Green supply chain management: A bibliometric analysis of global research trends and future directions. *Production & Manufacturing Research*, *12*(1), 2422614. [CrossRef]

Hass, L. (2022). *Nachhaltiges Lieferkettenmanagement multinationaler Unternehmen in der Textilindustrie*. Heft 49, März 2022, https://www.wiso.uni-hamburg.de/fachbereich-sozoek/professuren/ koerner/fiwa/publikationsreihe.html. Accessed on Oct 07, 2024.

Hunke, K. & Prause, G. (2014). Sustainable supply chain management in german automotive industry: Experiences and success factors. *Journal of Security and Sustainability*, *3*(3), 15–22. [CrossRef]

Jo, D. & Kwon, C. (2022). Structure of green supply chain management for sustainability of small and medium enterprises. *Sustainability*, *14,* 50. [CrossRef]

Kafa, N., Jaegler, A. & Sarkis, J. (2020). Harnessing corporate sustainability decision-making complexity: A field study of complementary approaches. *Sustainability*, 12(24), 1–23. [CrossRef]

Khan, S. A. R., Jian, C., Yu, Z., Golpîra, H. & Kumar, A. (2019). Impact of green practices on Pakistani manufacturing firm performance: A path analysis using structural equation modeling Computational intelligence and sustainable systems. In H. Anandakumar, R. Arulmurugan, & C. C. Onn (Eds.), *Computational intelligence and sustainable systems* (pp. 87–97). Springer. [CrossRef]

Lei, C., Zhang, H., Yan, X. & Miao, Q. (2024). Green supply chain optimization based on two-stage heuristic algorithm. *Processes*, *12*, 1127. [CrossRef]

Liu C. & Ma T. (2022). Green logistics management and supply chain system construction based on internet of things technology. *Sustainable Computing Informatics and Systems, 35*(1), 100773. [CrossRef]

Liu S. & Chang Y. T. (2017). Manufacturers' closed-loop orientation for green supply chain management. *Journal of Sustainability, 9*, 222, 1–29. [CrossRef]

Liu, J., Feng, Y., Zhu, Q. & Sarkis, J. (2018). Green supply chain management and the circular economy: reviewing theory for advancement of both fields. *International Journal of Physical Distribution & Logistics Management, 48,* 794–817. [CrossRef]

Makhdoom, Q., Junejo, I., Sohu J.M., Shah, S. M. M., Alwadi B. M., Ejaz F. & Hossain, M. B. (2025). Impact of green supply chain management on sustainable performance: a dual mediated-moderated analysis of green technology innovation and big data analytics capability powered by artificial intelligence. *F1000Research*, *13,* 1140.

Malviya, R.K., Kant, R. & Gupta, A.D. (2018). Evaluation and Selection of Sustainable Strategy for Green Supply Chain Management Implementation. *Business Strategy and the Environment,* 27, 475–502. [CrossRef]

Mastos, T.D. & Gotzamani, K.D. (2024). Sustainable supply chain management in the food industry: a case study of enablers and inhibitors from greek small enterprises. *World Review of Entrepreneurship, Management and Sustainable Development, 20,* 1–20. [CrossRef]

Nguyen, X. & Le, T. (2020). The impact of global green supply chain management practices on performance: The case of Vietnam. *Uncertain Supply Chain Management*, *8*(3), 523–536. [CrossRef]

Nozari, H. (2024). Green supply chain management based on artificial intelligence of everything. *Journal of Economics and Management*, *46*, 171–188. [CrossRef]

Peesapati, S. R. Y., (2024). Sustainability beyond compliance: Exploring technology`s role in driving sustainability across supply chains. *International Journal of Supply Chain Management*, 13 (2), 29–34. [CrossRef]

Perotti S, Zorzini M, Cagno E. & Micheli GJ (2012). Green supply chain practices and company performance: The case of 3pls in Italy. *International Journal of Physical Distribution & Logistics Management, 42*(7), 640–672. [CrossRef]

Sanayi ve Teknoloji Bakanlığı, (2024). Accessed on: July 03, 2024. https://ticaret.gov.tr/data/65dc9d3113b8762768385d66/STB%20Sunum_T%C3%BCrkiye%20Ye%C5%9Fil%20Sanayi%20Projesi_23022024.pdf.

Sharma, V.K., Chandna, P. & Bhardwaj, A. (2017). Green supply chain management related performance indicators in agro industry: A review. *Journal of Cleaner Production*, *141,* 1194-1208. [CrossRef]

Sharma, R., Shishodia, A., Gunasekaran, A., Min, H. & Munim, Z. H. (2022). The role of artificial intelligence in supply chain management: mapping the territory. *International Journal of Production Research*, *60*(24), 7527–7550. [CrossRef]

Singh, R. K., Modgil, S., & Shore, A. (2024). Building artificial intelligence enabled resilient supply chain: A multi-method approach. *Journal of Enterprise Information Management*, *37*(2), 414–436. [CrossRef]

Srivastava, S. K. (2007). Green supply-chain management: A state-of-the-art literature review. *International Journal of Management Reviews*, *9*(1), 53–80. [CrossRef]

Taseen, M., Yongli, W., Sadiq, B., Saleem, M. A. & Noor F. (2024). Study on the Impact of AI-powered Green Supply Chain Management Practices on Organizational Sustainable Performance with the mediation role of Technology Innovation. *2024 International Conference on Emerging Trends in Smart Technologies.* [CrossRef]

Tietze, T. (2023). "*Globale Supply Chains: Analyse in den Nachhaltigkeitsdimensionen Ökologie*", (Bachelorarbeit). Fakultät Informatik und Medien, Studiengang Verpackungstechnologie und Nachhaltigkeit, Germany. [German]

Tseng, M. L., Islam, M. S., Karia, N., Fauzi, F. A. & Afrin, S. (2019). A Litera-ture review on green supply chain management: Trends and futurechallenges. *Resources, Conservation and Recycling*, *141,* 145–162. [CrossRef]

Tubitak, (2024). Dünya Bankası desteğiyle yürütülecek "Türkiye yeşil sanayi projesi": sanayi kuruluşlarının yeşil inovasyon faaliyetlerinin desteklenmesi" bilgi notu. https://tubitak.gov.tr/sites/default/files/1-tubitak_sanayideyesildonusumprojesibilginotu.pdf. Accessed on: June 01, 2024.

Yıldız Çankaya, S. & Sezen, B. (2019). Effects of Green Supply Chain Management Practices on Sustainability Performance. *Journal of Manufacturing Technology Management*, *30*(1), 98–121. [CrossRef]

Wiredu, J., Yang, Q., Sampene, A. K., Gyamfi, B. A. & Asongu, S. A. (2023). The effect of green supply chain management practices on corporate environmental performance: Does supply chain competitive advantage matter? *Business Strategy and the Environment, 33*(3), 2578–2599. [CrossRef]

Zahran, S. (2024). Investigating the nexus between green supply chain practices and sustainable waste management in advancing circular economy. *Sustainability*, *16,* 3566. [CrossRef]

Zeeshan A. & Hajra B. (2024). Waste reduction and recycling: Schweizer-Sklar aggregation operators based on neutrosophic fuzzy rough sets and their application in green supply chain management. *Neutrosophic Systems with Applications*, *19*, 53–66. [CrossRef]

Zhu, Q. & Sarkis, J. (2004). Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. *Journal of Operations Management, 22*(3), 265–289. [CrossRef]

Yıldız Social Science Review
Web site information: https://yssr.yildiz.edu.tr
DOI: 10.51803/yssr.1896510

Original Article / Orijinal Makale

Contextual (De)Securitization of Migration in Turkish Parliamentary Debates

TBMM Tartışmalarında Göçün Bağlamsal Güvenlikleştirilmesi ve Güvenlikdışılaştırılması

Ezgi GÖKMEN, Fulya MEMİŞOĞLU ZAİMOĞLU

Department of Political Science and International Relations, İstanbul, Türkiye

Yıldız Teknik Üniversitesi, Siyaset Bilimi ve Uluslararası İlişkiler Bölümü, İstanbul,Türkiye

ARTICLE INFO

Article history
Received: 24 February 2026
Revised: 14 May 2026
Accepted: 15 May 2026

Key words:
Contextual securitization, discourse analysis, migration, securitization, Türkiye

MAKALE BİLGİSİ

Makale Hakkında
Geliş tarihi: 24 Şubat 2026
Revizyon tarihi: 14 Mayıs 2026
Kabul tarihi: 15 Mayıs 2026

Anahtar kelimeler:
Bağlamsal güvenlikleştirme, söylem analizi, göç, güvenlikleştirme, Türkiye

ABSTRACT

A significant strand of the migration–security literature approaches migration through a largely singular security discourse, an orientation that tends to obscure its articulation across different political and policy arenas. Drawing on parliamentary debates on migration in the Turkish Grand National Assembly (TGNA), this article shows that migration in Türkiye is framed not through a unified security logic but across multiple security registers that vary by context. Based on qualitative discourse analysis supported by MAXQDA-assisted coding, the study traces how migration is articulated across debates on security policy, the economy, public services, social cohesion, demographic change, as well as foreign policy relations. The findings indicate that securitization and desecuritization operate through contextual reconfiguration rather than a fixed trajectory.

Cite this article as: Gökmen, E., Memişoğlu Zaimoğlu, F. (2026). Contextual (De)Securitization of Migration in Turkish Parliamentary Debates. *Yıldız Social Science Review, 12*(1), 28–39.

ÖZ

Göç–güvenlik literatürünün önemli bir bölümü göçü tekil bir güvenlik söylemi üzerinden ele almakta; bu yaklaşım, göçün farklı siyasal ve politika alanlarında nasıl ele alındığını görünmez kılmaktadır. Türkiye Büyük Millet Meclisi'nde (TBMM) göçe ilişkin parlamento tartışmalarına dayanan bu makale, Türkiye bağlamında göçün tek bir güvenlik mantığıyla değil, bağlama göre değişen çoklu güvenlik çerçeveleri üzerinden kurulduğunu göstermektedir. MAXQDA destekli nitel söylem analizi temelinde yürütülen çalışma, göçün güvenlik politikaları, ekonomi, kamu hizmetleri, toplumsal uyum, demografik değişim ile dış politika alanlarında nasıl konumlandırıldığını incelemektedir. Bulgular, güvenlikleştirme ve güvenlikdışılaştırma

***Sorumlu yazar / Corresponding author**
*E-mail address: e.aydeniz1989@gmail.com

süreçlerinin doğrusal bir seyir izlemediğini; aksine bağlamsal yeniden yapılandırmalar yoluyla işlediğini ortaya koymaktadır.

Atıf için yazım şekli: Gökmen, E., Memişoğlu Zaimoğlu, F. (2026). Contextual (De) Securitization of Migration in Turkish Parliamentary Debates. *Yıldız Social Science Review, 12*(1), 28–39.

1. INTRODUCTION

The post-Cold War reorientation of security studies made it clear that security cannot be reduced to military threats alone. As the referent objects and sectors of security expanded, migration became increasingly entangled with security debates, not only as a matter of border control but as a field through which questions of sovereignty, governance, public order, welfare, and social cohesion are negotiated. Yet a significant strand of the migration-security literature treats securitization and desecuritization as relatively continuous processes, often assuming that migration discourse follows a unified logic across arenas. This article challenges that assumption through the Turkish case and argues that migration is articulated through multiple security registers that vary by context rather than a singular threat narrative.

As a prominent refugee hosting country, Türkiye offers a revealing setting for examining these shifts. The country's long experience with population movements, as well as its geopolitical location at the intersection of regional crises and European Union (EU) border governance, has made migration an especially salient political issue in domestic politics and foreign policy. Against this backdrop, the study asks how political actors reframe migration across different policy contexts. Parliamentary debates capture this salience in concentrated form, bringing together governance claims, partisan contestation, and competing moral and security discourses. In this setting, migration emerges as a discursive terrain through which political actors negotiate state capacity, social stability, and international responsibility.

The study argues that migration in Türkiye is not governed through a single security discourse but through transitions between multiple security logics that emerge in response to shifting political and institutional settings. While the argument is developed through Türkiye, it resonates with broader patterns observed in other transit and destination settings where migration governance blends security rationales with humanitarian and diplomatic considerations. Norman (2020), for example, demonstrates that migration policies in Türkiye and Morocco function not only as security instruments but also as tools of diplomacy. Lara **(2022)** similarly shows how Mexico's migration framework shifted toward securitization under external pressure and rising flows. Such cases point to migration as a domain shaped by interlinked domestic and international arenas, in which security meanings are reworked rather than simply applied.

Building on these insights, the study conceptualizes contextual securitization as the process through which security discourses are reconfigured by different actors across policy domains in response to changing historical, institutional and social conditions. This approach makes it possible to capture how securitization and desecuritization unfold not as a linear sequence but as a set of shifting articulations that move across contexts such as security policy, economy, public services, social cohesion, demographic change, and foreign policy. By focusing on parliamentary discourses, the article contributes to public policy scholarship by showing how legislative actors reframe migration through competing logics with implications for agenda-setting and decision-making in a period of prolonged displacement and political polarization.

The article proceeds as follows. The next section develops the analytical framework of contextual securitization and situates it within migration-security debates. This is followed by a section that positions Türkiye as a contextual site of migration securitization. It then outlines the research design and coding strategy. The findings section examines how migration is securitized and desecuritized across six contextual domains and compares dominant framing across political parties. The conclusion summarises the key findings and reflects on their implications for securitization research.

1.1. Analytical Framework: Contextual Securitization of Migration

For much of the Cold War, security studies centred on state survival, territorial integrity and military threat. Security was largely understood through a single referent object – the state – and a narrow threat spectrum. Post-Cold War transformations unsettled this approach, expanding security debates to include political instability, economic vulnerability, environmental crises, and migration. More importantly, security came to be understood not as an objective condition but as a socially and politically constructed category.

The Copenhagen School provided one of the most influential articulations of this shift. Challenging classical approaches centred on military threats and state survival, scholars associated with this tradition conceptualized security as a political and communicative act. They also introduced securitization and desecuritization as key analytical concepts for explaining how issues move between ordinary

politics and security domains. Buzan (1991) broadened security analysis across multiple sectors, while Baldwin (1997) emphasized that security acquires analytical meaning only when questions 'for whom', 'against what', and 'at what cost" are specified. Wæver's (1995) speech act formulation further argued that an issue becomes securitized when it is framed as an existential threat and accepted as such by a relevant audience, thereby legitimizing extraordinary measures. Security thus emerges through political articulation and social recognition rather than material danger alone.

Despite its transformative impact, securitization theory has often been applied as if it followed a universal logic. Even when conceived discursively, analyses frequently presume a stable trajectory through which issues move from normal politics to emergency politics. Such assumptions risk obscuring how security is articulated differently across institutional arenas, and socio-cultural contexts.

Migration provides a critical site for unpacking these limits. Existing scholarship emphasises that migration is securitized across multiple arenas, including national and regional policy frameworks (Huysmans, 2000; Karyotis, 2007), international governance structures (Oğuz-Gök, 2016), and public discourse. Empirical studies further demonstrate that migrants and refugees are frequently associated with illegality, crime, economic burden and social threat in both media and political narratives (Van der Valk, 2003; Fotopoulos & Kaimaklioti, 2016; Canveren & Akgül-Durakçay, 2017; Goodman et al., 2017). Together, this literature highlights that securitization does not operate through a single logic but unfolds across intersecting discursive fields in which sovereignty, welfare, identity and public order become entangled.

Within securitization theory, political elites occupy a central position in defining and legitimizing security problems through discourse (Baysal, 2020). Migration narratives are actively constructed within political arenas, particularly through leadership rhetoric and parliamentary debate (Kabacaoğlu and Memişoğlu, 2021). Parliamentary discourses therefore constitute a critical analytical site where competing threat constructions, policy priorities and moral claims intersect, rendering migration simultaneously a matter of governance, security, and social responsibility.

Within this broader field, irregular migration has become a prominent object of securitization, often invoked through narratives of border control and legality. Irregular migration can be broadly understood as border crossings, residence or employment outside formal legal frameworks (Bartram et al., 2019), a category that raises governance concerns for states while also exposing migrants to heightened vulnerability (Koser, 2007). Since the 1990s, diversified forms of irregular migration – including unauthorised entry, visa overstays, undeclared employment – have further increased the policy salience of the category and contributed to its placement within security agendas (Vogel, 2016; Öner, 2021). Yet irregularity is rarely fixed. Legal statuses can shift over time, complicating measurement and governance and undermining attempts to treat irregular migration as a stable security problem governed by a single logic (Koser, 2010).

This fluidity becomes particularly visible in crisis contexts. In 2015, irregular mobility through Türkiye intensified alongside onward movement toward Europe, as people fleeing conflict – predominantly Syrians – combined irregular border crossings with claims for international protection. As documented in earlier work, this period was marked not only by increased visibility of irregular routes but also by a discursive convergence between migration, asylum and security, in which forced displacement became routinely framed through the language of crisis and control (Memişoğlu & Başol, 2015). Irregular migration and refugeehood became intertwined in political and media narratives, blurring legal categories and reinforcing securitized understandings of mobility.

Building on critiques that securitization is not solely discursive but also institutional and material (Balzacq, 2011), this study adopts a contextual approach. As McDonald (2008) asserts, security is not constructed identically across societies or political arenas but varies according to historical, cultural and institutional condition. (De)securitization emerges at the intersection of leadership discourse, institutional arrangements, public opinion, and international pressures. It operates through multiple referent objects and overlapping logics rather than a singular threat narrative, and shifts across domestic politics, foreign policy, economic governance and societal debate. Guided by this framework, the analysis focuses on three interconnected dimensions: discursive representations of migration through key concepts and themes, institutional framings embedded in parliamentary processes and state practices, and broader political, economic and social conditions shaping these constructions.

1.2. Türkiye as a Contextual Site of Migration Securitization

Türkiye offers a particularly layered context in which these dynamics become visible. Simultaneously a country of immigration, emigration and transit, this multi-positionality diversifies both the scope and direction of security discourses (İçduygu & Aksel, 2012). Its geopolitical location positions the country as a corridor between the Middle East, the Caucasus and Europe, exposing it directly to regional crises while situating it within European border governance strategies (Şemşit, 2010; Özey, 2017). Migration thus operates not only as an internal political concern but also as an instrument of foreign policy and regional positioning.

Discursively, migration in Türkiye has been framed through multiple and sometimes competing registers. In the context of Syrian displacement, official discourse for extended periods emphasized notions such as 'guesthood', 'ensar-muhacir' and 'humanitarian responsibility' (İçduygu & Demiryontar, 2022). Over time, economic pressures, political polarization and debates over border control contributed to a gradual shift toward securitizing framings centred on public order and return. Leadership rhetoric has oscillated between humanitarian and security-centred registers depending on domestic political conditions (Kabacaoğlu and Memişoğlu, 2021). These discursive shifts were reinforced institutionally through coordination among security forces, migration authorities, and foreign policy actors, while media narratives and economic anxieties further shaped public perceptions. Within this broader landscape, references associated with irregularity constituted an intensified site where border control, social order, economic concerns and protection narratives intersected most visibly.

The year 2015 represents a critical conjuncture in this process. It marked a period in which irregular mobility, asylum seeking and regional instability converged with electoral competition and heightened parliamentary contestation. Migration became a salient political issue not only because of displacement from Syria but also because mobility increasingly entered public debate through the language of crisis, responsibility and security. Parliamentary discussions during this period thus provide a particularly revealing window into how political actors reconstructed migration across shifting and context-dependent security logics. It is through this lens that the study examines TGNA debates from 2015, tracing how migration was alternately securitized and desecuritized across multiple policy domains. The following analysis demonstrates that migration did not move along a linear trajectory from humanitarian concern to security threat. Rather, it was continuously rearticulated through overlapping and competing frames, reflecting the contextual nature of securitization within Turkish parliamentary politics.

1.3. Research Design

This study adopts a qualitative research design and analyses parliamentary discourse through the lens of Critical Discourse Analysis (CDA). Given the political salience of migration and security in Türkiye, parliamentary debates provide a key site for examining how migration is framed, contested and reconfigured across shifting political contexts.

Data were drawn from TGNA General Assembly transcripts, committee reports and parliamentary speeches, enabling analysis of how migration is articulated across state discourse, political party positions and interactions between government and opposition. These sources capture both institutional framings and partisan contestation, offering insight into the construction of migration across multiple political arenas.

The primary dataset consists of TGNA General Assembly transcripts from January 2015 onwards, covering the parliamentary debates preceding both the June 7 and November 1, 2015 general elections. As indicated earlier, this period constitutes a critical conjuncture, marked by intensified cross-border mobility from Syria, heightened domestic political competition and growing public debate over migration. Rather than representing a singular 'migration crisis', 2015 reflects a moment in which forced displacement and irregular mobility became increasingly intertwined within parliamentary discourse. Focusing on this timeframe provides a concentrated snapshot of shifting discursive dynamics at a point when migration moved to the centre of legislative debate.

Parliamentary materials were retrieved from the official TGNA online archive using keywords including migration, irregular migration, border security, return, and Syrian. This search yielded 64 parliamentary sessions, from which 76 speeches, motions and statements directly or indirectly addressing migration were identified for analysis. Texts were coded thematically using MAXQDA and analysed through CDA to capture not only linguistic patterns but also political positioning, embedded power relations and broader social effects (Wodak & Meyer, 2001). Coding proceeded through integrated stages. First, texts were coded for securitization indicators, including constructions of threat, identification of referent objects and calls for extraordinary measures. Second, these codes were grouped across Buzan's (1991) sectoral framework – military, political, economic, societal and environmental – and classified as securitizing or desecuritizing. Third, discourses were interpreted across six contextual domains: demographic shifts, security policies, economy public services, social cohesion and foreign policy relations.

MAXQDA enabled individual speeches to be coded across multiple security logics simultaneously, reflecting the contextual and overlapping nature of securitization. This approach allows the analysis to demonstrate empirically that migration is reconstructed not through a single security discourse but through multiple, domain-specific security rationalities that shift across political contexts.

2. DISCUSSION OF FINDINGS: CONTEXTUAL PATTERNS OF (DE)SECURITIZATION

2015 parliamentary debates period marks a critical juncture in Türkiye's migration governance, shaped by the mass arrival of Syrian refugees alongside electoral competition, economic pressures and regional instability. Together, these dynamics produced a discursive environment in which migration oscillated between humanitarian responsibility and state-centred security concerns.

As summarised in Tables 1 and 2, political parties mobilized distinct securitization and desecuritization logics reflecting divergent ideological positions and political strategies. The ruling Justice and Development Party (the AK Party) predominantly framed migration through humanitarian responsibility, fraternity and religious solidarity, drawing on the 'ensar-muhacir' narrative to situate the displacement within a moral register. In contrast, the Republican People's Party (CHP) and the Nationalist Movement Party (MHP) foregrounded border security, economic burden and social disorder, framing migration as a threat to public order and societal stability. Peoples' Democratic Party (HDP) articulated a counter-discourse grounded in human rights and social justice, emphasising migrants' dignity, equality, and access to essential services. Rather than positioning migrants as security risks, HDP narratives consistently framed them as subjects of protection, contributing to processes of societal desecuritization.

These contrasting discursive orientations indicate that 2015 constituted a transitional moment in which migration was reconstructed across competing security logics. Securitization dominated many parliamentary exchanges, yet desecuritizing frames also emerged intermittently, revealing how humanitarian and rights-based narratives coexisted with state-centred security concerns. Migration thus became a contested political terrain in which human security and state security were articulated in tension rather than sequence.

Across parliamentary debates, the findings indicate that political actors alternated between securitizing and desecuritizing framings within six interrelated contexts: (1) demographic shifts and public order, (2) security policies, (3) economy, (4) public services, (5) social cohesion and (6) foreign policy. The following sections examine these contexts comparatively across political parties.

2.1. Demographic Shifts

Debates on demographic shifts were discursively linked to concerns over public order, governance capacity and social stability, and were dominated by securitizing framings advanced primarily by the CHP and MHP. The CHP mobilised political (3) and societal (3) securitization logics by linking migration flows to demographic shifts and social burden and governance failure. In a TGNA speech delivered in February 2015, a Member of the Parliament (MP) from CHP framed migration as a political and demographic threat, arguing that the incumbent government's uncontrolled policies toward Syrian refugees had 'triggered terrorism, human trafficking, drug trade, economic crisis and sectarian tensions'[1]. By situating migration alongside organised crime and instability, migration was constructed as a security risk undermining 'state capacity, public order and political stability'. Consequently, the CHP's migration discourse relied on a political securitization logic, framing failures in migration governance as a political crisis threatening national unity.

The MHP similarly articulated migration as a societal security concern, emphasising migration as a demographic threat through societal securitization (2). The party emphasized that migrant inflows undermined cultural homogeneity. In a parliamentary session from January 2015, an MHP MP described the dense presence of Syrian refugees in Hatay as 'a crowd that paralysed urban life', adding that 'residents can no longer even walk on the pavements'.[2] Such language framed migrant presence as a disruption of everyday order, reinforcing a narrative in which demographic change was securitised through appeals to social cohesion and cultural integrity.

1 Rules of Procedure of the Grand National Assembly of Turkey. TGNA Session No: 63. date: 24.02.2015
2 TGNA Session No: 41, date: 13.01.2015

Table 1. Dominant securitization logics and discursive framings by political party (2015)

Political party	Primary securitization logics	Characteristic framing of migration
AK Party	Political– Military	Migration is framed as an indicator of state capacity within border security and regional stability agendas, closely associated with irregular flows, national security and counter-terrorism narratives. Humanitarian responsibility is mobilised alongside strategic security discourse.
CHP	Political – Economic – Societal	Migration is constructed as a governance problem producing economic strain, pressure on public services and social instability, with emphasis on unemployment, rising costs and weakened state capacity.
MHP	Economic – Political – Societal	Migration is framed as a demographic and identity-based threat undermining social order and national sovereignty, foregrounding cultural disruption and concerns over public order.
HDP	Political – Environmental	Migration is articulated as a political crisis rooted in war and foreign policy, emphasising human dignity, environmental risk and state responsibility while challenging state-centric securitization narratives.

AK Party: The ruling Justice and Development Party; CHP: Republican People's Party; MHP: Nationalist Movement Party; HDP: Peoples' Democratic Party.

Table 2. Dominant desecuritization logics and discursive framings by political party (2015)

Political party	Primary desecuritization logics	Characteristic framing of migration
AK Party	Societal	Migration is framed through humanitarian responsibility, fraternity, and social cohesion, keeping it within ordinary politics and legitimizing assistance through moral and religious references rather than threat narratives.
CHP	Societal – Political	While generally securitizing, CHP discourse occasionally cautions that exclusionary policies may intensify social unrest, framing limited desecuritization around social stability and governance concerns.
MHP	Societal	Desecuritization appears only sporadically, primarily through humanitarian references to Turkmen kin communities, within an otherwise dominant securitizing framework.
HDP	Political – Societal	Migration is consistently desecuritized through appeals to human rights, equality and coexistence, positioning migrants as subjects of protection rather than sources of threat.

AK Party: The ruling Justice and Development Party; CHP: Republican People's Party; MHP: Nationalist Movement Party; HDP: Peoples' Democratic Party.

By contrast, the HDP adopted a desecuritizing discourse grounded in human security (2). Rather than framing migration through demographic risk, HDP MPs highlighted gender-based violence, exploitation and rights violations affecting migrant women (coding category: gender and demography).[3] HDP interventions emphasised migrant women's right to life and dignity, shifting attention away from border control toward protection and social justice. This framing positioned migrants, particularly women, not as sources of threat but as subjects requiring protection, marking a clear departure from state-centred security narratives. Taken together, debates in this context reveal how demographic change became politicized through competing security logics: while CHP and MHP framed migration as a threat to political stability and social order, the HDP framed it through a humanitarian and rights-based perspectives.

2.2. Security Policy

The security policy context constitutes the domain in which state-centred securitization was most explicit. Across party lines, migration was predominantly articulated through its association with irregularity and embedded within narratives of border security, terrorism, and national sovereignty, although the intensity and emphasis differed. As demonstrated in Table 1, the intensity and modality of securitization vary significantly across political actors. While political and military securitization dominate parliamentary debates, desecuritizing interventions remain comparatively limited and are largely confined to HDP discourse, as indicated in Table 2.

Within this domain, the AK Party articulated migration primarily through political (6) and military (5) securitization, frequently linking irregular crossings and border control to regional instability. By positioning migration – particularly in its irregular forms – alongside cross-border threats stemming from DAESH and similar organizations, the AK Party embedded mobility within a strategic security framework[4]. Migration management was thus constructed as a matter of state security and military preparedness, legitimising measures such as the mandate's extension.

The MHP demonstrated a strong tendency toward political (4), military (3) and societal (2) securitization, positioning migration as a threat to national sovereignty and identity. In February 2015, in the aftermath of the evacuation of the Süleyman Şah Tomb on 22 February and amid intensifying criticism of the government's Syria policy, an MHP MP described Syrian refugees in Adana as creating an "environment of chaos" and warned that the city stood "on the brink of social explosion."[5] This speech was delivered at a time when the advance of DAESH in Syria and the rapid increase in the number of Syrian refugees in Turkey—approaching 1.5 million—had brought issues of national security and social cohesion to the forefront of the political agenda. Within this context, the metaphor of "explosion" constructed migration as an imminent destabilizing force, implicitly associated with uncontrolled entry and framed as a direct threat to public order and national integrity. Therefore, this discourse was not confined solely to the issue of migration; rather, it was articulated within a broader security conjuncture shaped by Syria policy, the threat of terrorism, and concerns over border security.

The CHP combined political (6) and economic (3) securitization within this context. Border control was framed not only as a matter of national security but also as a pre-

3 TGNA Session No: 1, date: 01.10.2015

4 TGNA Session No: 9, date: 03.09.2015

5 TGNA Session No: 64, date: 25.02.2015

requisite for economic stability. In February 2015, a CHP MP argued that migration had "increased unemployment, doubled or tripled rents and placed local merchants in a difficult position."[6] The speech was delivered in the immediate aftermath of two major security developments: the violent 6–7 October 2014 unrest, which had resulted in over forty deaths, and the evacuation of the Süleyman Şah Tomb on 22 February 2015 amid DAESH advances in Syria. These events intensified public debate on terrorism, border governance, and the government's Syria policy. Within this heightened security atmosphere, the MP described the presence of nearly 1.5 million Syrian refugees as creating an "environment of chaos" and warned that certain cities stood "on the brink of social explosion." Migration was therefore not framed merely as an economic burden; it was associated with uncontrolled entry, rising crime, informal networks, and the potential infiltration of terrorist actors. Although economic grievances—such as unemployment and rising rents—were emphasized, they were articulated within a broader internal security narrative. Weak border governance was implicitly constructed as enabling destabilization, while the concentration of Syrians was portrayed as threatening public order and social cohesion. Thus, the speech intertwined political and internal security securitization with economic concerns, situating migration within a broader conjuncture shaped by terrorism, border instability, and the post-6 7 October 2014 unrest.

The HDP, by contrast, engaged in limited military (2), societal (2) and economic (2) securitization while simultaneously advancing a desecuritizing discourse that challenged the criminalization of migrants. Within the broader security policy debate, the HDP advanced a counter-securitizing discourse that rejected the framing of migration as a threat. In a parliamentary motion, HDP MPs addressed Syrian refugee children not as security subjects but as victims of war, emphasizing their forced displacement and vulnerability[7]. By shifting the referent object from the state to human dignity and universal rights, the party sought to remove migration from the realm of emergency security governance and reposition it within a humanitarian and political responsibility framework.

These securitizing framings intensified particularly in the context of electoral competition, increasing public anxiety following the 6-7 October 2014 unrest, and institutional concerns over border governance during the expansion of DAESH in Syria. Parliamentary actors strategically mobilized migration discourse not only to address mobility itself but also to contest governmental legitimacy, state capacity and crisis management. In this sense, securitization functioned as a mechanism of political positioning shaped by broader domestic and regional pressures.

2.3. Economic Context

In debates on economy, securitization shifted from territorial integrity to material stability. Here, irregular migration was framed less through military metaphors and more through economic insecurity, primarily by the CHP and the MHP.

The CHP articulated migration as an economic burden (4) by associating it with unemployment, inflation and public expenditure. Party discourse emphasized 'pressure on the local labour force' and 'rising social spending'. In February 2015, amid heightened security concerns following the 6–7 October 2014 unrest and the continued expansion of DAESH along the Syrian border, the CHP articulated migration primarily as an economic burden while embedding it within a broader internal security narrative. Addressing the TGNA, a CHP MP argued that the Syrian population in Gaziantep—approaching 400,000—had "increased unemployment," suppressed wages through cheap labour, and intensified "economic, social, health and housing problems."[8] The speech linked rising unemployment, business closures, and labour displacement to the influx of Syrians, portraying migration as exerting structural pressure on local markets and public services. Thus, while the CHP foregrounded labour market pressures and public expenditure strain, migration was simultaneously situated within a broader conjuncture shaped by border instability, the Syrian civil war, and the perceived expansion of extremist groups. In this context, migration was constructed as a destabilising force affecting both local economic balances and internal security structures. By emphasising labour displacement and strain on public finances, the CHP constructed migration as a destabilising force in local economies.

Similarly, the MHP framed migration as both an economic and identity-based threat through economic (7) and societal (2) securitization. The speech, delivered in January 2015, took place in a period marked by heightened tensions along the continued presence of DAESH in northern Syria. Border provinces such as Hatay were experiencing a visible increase in Syrian settlement, which had become a focal point of public debate concerning local commerce, security, and state authority. Within this context, themes of "economic burden" and "unfair access to state resources" recurred frequently. An MHP MP argued that Syrian refugees operating informal businesses in Hatay—particularly in Reyhanlı and Antakya—had undermined local commerce by opening shops without bureaucratic procedures, taxation, or regulatory compliance.[9] This was contrasted with Turkish shopkeepers who faced extensive licensing requirements and financial obligations. The narrative constructed Syrians as benefiting from regulatory asymmetry, thereby generating unfair competition and deepening resentment

6 TGNA Session No: 63, date: 24.02.2015
7 TGNA Session No: 92, date: 23.04.2015
8 TGNA Session No:58, date: 19.02.2015
9 TGNA Session No 42, date:14.01.2015

among local merchants. The discourse further implied that prolonged Syrian settlement could alter local social balances, warning that social unrest and "economic explosion" were possible outcomes. Although the immediate focus was on trade and small business regulation, migration was embedded within a broader security atmosphere shaped by border instability and weak governance. Thus, migration was framed simultaneously as an economic security issue—threatening livelihoods and local markets—and as a catalyst for societal tension, capable of undermining social cohesion and public order.

Informal labour markets and the distribution of public resources reinforced the framing of migration as a problem of economic and administrative control. These discourses were further shaped by growing societal pressures linked to unemployment, housing costs and economic uncertainty, enabling migration to be articulated not merely as a demographic phenomenon but as a broader source of socio-economic instability. Consequently, economic securitization operated not through a singular threat construction but through context-dependent reconfigurations connecting material insecurity, governance practices and political struggle. This pattern is consistent with Table 1, where CHP and MHP discourses are predominantly associated with economic and societal securitization logics.

2.4. Public Services

In debates on public services, securitization of migration was framed primarily through its perceived impact on education, health and housing. Here, securitization centred on resource allocation and institutional capacity, with parties oscillating between threat narratives and humanitarian concern.

The CHP emerged as the most securitizing political party in this domain, combining political (6), economic (5) and societal (5) frames. Party discourse emphasised that migrants 'strained public resources' and increased pressure on local services. At the same time, the CHP articulated a rights-based counter-frame, emphasizing equal access to public services and criticizing discriminatory practices toward specific refugee groups. Rather than portraying migrants as a threat, the party framed unequal treatment as a potential source of social tension, thereby shifting the focus from migrants themselves to state responsibility and governance practices. In this sense, the CHP's discourse[10] did not securitize migration directly; instead, it suggested that exclusionary policies and unequal access to services could generate instability, implicitly locating the source of risk in governance failures rather than in migrants themselves.

The MHP similarly framed public services through political (4) and economic (3) securitization, portraying migration as a burden on state resources. Yet some interventions also revealed moments of humanitarian awareness. For example, an MP drew attention the conditions of Syrian refugees in Osmaniye Cevdetiye refugee accommodation centre, noting that '8-12 people shared 15-square-metre tents', sewage flowed openly and water and electricity cuts were frequent.[11] Although embedded within an otherwise securitizing discourse, such descriptions introduced humanitarian suffering into debate on service provision, illustrating the multi-layered nature of migration narratives.

By contrast, the AK Party adopted a predominantly desecuritizing approach in this context, presenting access to housing, health and education as evidence of state responsibility and social solidarity, producing societal desecuritization (4). An MP from AK Party emphasized that Syrian refugees benefited from free public services, framing migration as a moral commitment rather than a threat to public order.[12] As reflected in Table 2, desecuritizing interventions emerged primarily through societal frames emphasizing equality, solidarity and access to services.

The HDP further reinforced desecuritization by emphasising equality and rights. For example, an MP from HDP questioned why the Directorate of Religious Affairs provided education to migrant children while the Ministry of Education remained relatively inactive, highlighting unequal treatment based on ethnicity and religion.[13] Through societal desecuritization (4), such interventions reframed public services as a matter of equal opportunities rather than security. Overall, public services emerged as a key site where societal security was negotiated, revealing tensions between resource-based securitization and rights-based desecuritization.

Debates on public services reveal that migration was securitized through broader struggles over governance capacity, institutional responsibility and social distribution. In the context of increasing electoral polarization and public sensitivity toward resource allocation in 2015, political actors mobilized migration discourse to contest the state's ability to manage welfare provision, housing, education and healthcare infrastructures. Institutional practices surrounding service delivery, refugee accommodation and bureaucratic coordination became central to these debates, shaping perceptions of both state competence and social fairness. At the same time, growing societal anxieties regarding unequal access to public resources, deteriorating living conditions and local service pressures intensified the political salience of migration. Consequently, public services functioned as a key contextual domain in which securitization and desecuritization intersected through competing constructions of social justice, state responsibility and societal stability.

10 TGNA Session No: 53, date:10.02.2015

11 TGNA Session No: 64, date:25.02.2015

12 TGNA Session No: 64, date: 25.02.2015

13 TGNA Session No: 82, date: 26.01.2015

2.5. Social Cohesion

The social cohesion constituted one of the most polarised domains of migration discourse, exposing sharp inter-party divergence over belonging, identity, and coexistence.

The AK Party framed migration through narratives of cohesion and brotherhood, producing strong societal desecuritization (6). When addressing the TGNA in February 2015, an MP from AK Party described refugee policies as a 'humanitarian duty' and 'bond of brotherhood', emphasizing shared religious and cultural ties.[14] By legitimizing public spending through moral and religious references, the AK Party portrayed migration as strengthening social unity rather than threatening cohesion.

The CHP adopted a more ambivalent position in this period, combining moderate political (3) and societal securitization (3) frames. The speech, delivered in January 2015 amid ongoing border instability following the 6-7 October 2014 unrest and the continued expansion of the Syrian conflict, situated Gaziantep as "the most adversely affected province" in the aftermath of the Syrian crisis[15]. The MP emphasized that the presence of hundreds of thousands of registered and unregistered Syrians had placed the province under strain, implicitly linking migration to growing social and administrative pressures. Rather than employing overt threat rhetoric, the discourse framed migration as a structural factor destabilizing local balance, particularly in terms of governance capacity, social cohesion, and resource distribution. By highlighting the scale of both documented and undocumented Syrians, the MP suggested weak regulatory control and administrative overload, thereby embedding migration within a broader narrative of border vulnerability and uneven state response. In this sense, migration was constructed less as a humanitarian issue and more as a governance and societal stability challenge.

The MHP articulated the strongest securitization in this domain (economic 4, societal 3, political 3), linking migration to the weakening of 'Turkish identity' and local order. An MP from MHP described refugees as generating economic decline, infrastructural strain and cultural disruption, constructing migration as a localized crisis threatening both societal cohesion and national integrity.[16]

Conversely, the HDP foregrounded coexistence, dignity and human vulnerability, advancing a form of societal desecuritization (5). In a TGNA speech delivered in April 2015[17], amid heightened security debates following the 6-7 October 2014 unrest and ongoing discussions on internal security legislation, the HDP Group Deputy Chair emphasized the need for a shared and equal future grounded in solidarity rather than exclusion. Syrian children were described as "cold, hungry and exposed," forced to flee a war created by others and abandoned to what was termed a "shame on humanity." Rather than framing migration as a destabilizing force, the speech located insecurity in war, poverty, and structural injustice. The referent object shifted from the state and national identity to children's dignity and fundamental rights. By rejecting narratives that associated refugees with disorder or threat, the HDP reframed migration through collective responsibility and moral accountability. Thus, social cohesion emerged as a central discursive arena in which migration was alternately constructed either as a threat to societal balance by other parties or, in the HDP's case, as a call for inclusive coexistence and equal citizenship. In this formulation, insecurity was not attributed to migrants themselves but to exclusionary policies and systemic inequality.

In the context of increasing electoral polarization and public sensitivity toward resource allocation in 2015, political actors mobilized migration discourse to contest the state's ability to manage welfare provision, housing, education and healthcare infrastructures. Institutional practices surrounding service delivery, refugee accommodation and bureaucratic coordination became central to these debates, shaping perceptions of both state competence and social fairness. At the same time, growing societal anxieties regarding unequal access to public resources, deteriorating living conditions and local service pressures intensified the political salience of migration. Consequently, public services functioned as a key contextual domain in which securitization and desecuritization intersected through competing constructions of social justice, state responsibility and societal stability.

2.6. Foreign Policy

In the foreign policy context, migration acquired dual significance as both a security risk and a diplomatic instrument.

The AK Party framed migration through international burden-sharing and regional responsibility, combining political (2) and military (3) securitization with elements of societal desecuritization (2). An MP from AK Party highlighted Türkiye's ability to accept 24,000 refugees in two days – contrasting this with Europe's remarkably slower response in five years – thereby presenting migration management as evidence of state capacity and geopolitical leverage.[18]

The CHP addressed foreign policy matters with reference to border security failures, producing military (2) and political (1) securitization. In a speech, a CHP MP criticized what was described as a "faulty Syria policy" and an "unrealistic and insincere refugee policy," arguing that the government's external miscalculations had directly contributed to instability[19]. Referring to the presence of approximately 1.8 million Syrians in Turkey, the MP contended

14 TGNA Session No: 64, date:25.02.2015
15 TGNA Session No: 46, date:22.01.2015
16 TGNA Session No: 68, date: 03.03.2015
17 TGNA Session No: 92, date: 23.04.2015
18 TGNA Session No: 9, date: 03.09.2015
19 TGNA Session No: 9, date:03.09.2015

that the fact that many refugees were forced to embark on "journeys of death" across the sea constituted a collective political responsibility. Rather than framing refugees themselves as the source of threat, the discourse securitized state mismanagement and border governance failures. The referent object was national credibility and territorial control, while the locus of insecurity was located in flawed foreign policy choices. By linking irregular departures and unsafe maritime crossings to inadequate policy planning, the CHP embedded migration within a broader narrative of weakened border security and diplomatic miscalculation.

The MHP similarly constructed migration as a strategic security issue, producing discourses of political (2), military (2) and economic (1) securitization. When addressing the parliament in month 2015, an MHP MP criticized the AK Party's 'strategic depth' and 'zero problems with neighbours' doctrine, arguing that it had exposed Türkiye to regional instability and weakened diplomatic standing.[20] Migration was framed not merely as humanitarian fallout but as evidence of jeopardizing economic and international reputation loss.

The HDP approached migration through political (3) and environmental (2) securitization, alongside limited political (1) and societal (1) desecuritization. MPs linked refugee flows to the devastation caused by the Syrian conflict, describing cities and displaced populations as outcomes of political decisions.[21] At the same time, HDP interventions highlighted environmental and spatial risks in border regions, referring to mortar shells falling into residential areas, the destruction of agricultural land.[22] Through this framing, migration was articulated as both a political crisis and a threat to human and environmental security. Across parties, migration thus functioned simultaneously as a symbol of geopolitical vulnerability, diplomatic positioning, and security risk.

In the context of the Syrian conflict and increasing international pressure surrounding refugee mobility toward Europe, migration became deeply embedded within broader disputes over Türkiye's regional strategy and diplomatic positioning. Parliamentary actors mobilized migration discourse not only to address cross-border mobility itself but also to contest governmental legitimacy, foreign policy credibility and the state's capacity to manage regional instability. Institutional concerns regarding border control, refugee governance and international burden-sharing further reinforced the securitization of migration within foreign policy debates. Simultaneously, growing societal anxieties related to territorial security, regional conflict spillover and diplomatic isolation intensified the political salience of migration. Consequently, migration functioned not merely as a humanitarian issue or external policy challenge, but as a context-dependent security discourse through which geopolitical vulnerability, national prestige and state authority were continuously renegotiated.

3. CONCLUSION

This study examined the securitization of migration in Türkiye through parliamentary debates during 2015, demonstrating that migration governance operates through multiple, context-dependent security logics rather than a unified threat framework. Across six interrelated analytical domains – demographic shifts, security policy, economy, public services, social cohesion and foreign relations – migration was alternately securitized and desecuritized, revealing a dynamic discursive landscape shaped by political competition, institutional practices and shifting societal pressures.

The findings show that political parties articulated migration through distinct security registers. The ruling AK party primarily framed migration in relation to border control, regional stability and state capacity, while simultaneously mobilizing humanitarian narratives grounded in moral responsibility. The CHP emphasized economic burden, pressure on public services and migration governance failure, constructing migration as both a political and material crisis. The MHP foregrounded demographic risk, social order and national identity, advancing the most consistent societal securitization. In contrast, the HDP articulated a counter-discourse centred on human dignity, environmental risk and political responsibility, persistently challenging state-centred security framings.

These patterns confirm that securitization of migration debates in Türkiye unfolds as a contextual process shaped not only by perceived threats but also by electoral dynamics, economic uncertainty and foreign policy matters. Migration discourse functioned simultaneously as a site of governance justification and political contestation, revealing securitization to be not merely a state reflex but a mechanism of legitimacy production within parliamentary politics.

The Turkish case also illustrates the increasing entanglement of security and humanitarian discourses. Migration was framed both as a moral obligation through appeals to fraternity and solidarity, and as a security concern through narratives of border control, public order and return. This dual configuration produces a hybrid migration rhetoric in which humanitarian responsibility coexists with security-oriented approaches. While such flexibility allows space for desecuritization within pluralistic political arenas, it also risks normalising exceptional measures and marginalizing migrants as political subjects.

Conceptually, the study advances a contextual securiti-

20 TGNA Session No: 40, date: 08.01.2015
21 TGNA Session No: 65, date:26.02.2015
22 TGNA Session No: 72, date:09.03.2015

zation approach that moves beyond linear models of threat construction. By showing how migration is reconstructed across overlapping domains, the analysis points out that securitization operates through shifting configurations rather than stable trajectories. Security emerges not solely through discourse but through the interaction of political actors, societal anxieties, and international pressures. Treating migration as a fixed security problem thus obscures how political contexts actively shape policy responses, often producing inconsistent and contradictory outcomes. Instead, security could be understood as a multi-level and contingent process embedded in domestic politics and regional power relations.

Future research could extend this contextual perspective by examining media narratives, local governance practices and community-level perceptions, further illuminating how security meanings circulate beyond parliamentary spaces. Such work would deepen understanding of migration not as a singular threat but as a dynamic social field through which the boundaries of security are continually redrawn.

REFERENCES

Aradau, C. (2004). Security and the democratic scene: Desecuritization and emancipation. *Journal of International Relations and Development, 7*(4), 388–413. [CrossRef]

Baldwin, D. A. (1997). The Concept of Security. *Review of International Studies, 23,* 5–26. [CrossRef]

Balzacq, T., Léonard, S., Kaunert, C., Vultee, F., Wilkinson, C., Salter, M., Trombetta, M. J., Sjöstedt, R., Stritzel, H., Schmittchen, D., Vuori, J., & Williams, M. (2011). *Securitization Theory: How Security Problems Emerge and Dissolve* (T. Balzacq, Ed.). Routledge. [CrossRef]

Bartram, D., Poros, M. V., & Monforte, P. (2019). *Key Concepts in Migration Studies* (2nd ed.). Routledge.

Baysal, B. (2020). 20 Years of Securitization: Strengths, Limitations and A New Dual Framework. *Uluslararası İlişkiler Dergisi, 17*(67), 3–20.

Buzan, B. (1991). *People, States and Fear: an agenda for international security studies in the post-Cold War era* (1st ed.). Pearson Education.

Canveren, Ö. and Akgül-Durakçay, F. (2017). The analysis of the Hungarian government's discourse towards the migrant crisis: A combination of securitization and Euroscepticism. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İİBF Dergisi, 24*(3), 857–876. [CrossRef]

Fotopoulos, S., & Kaimaklioti, M. (2016). Media discourse on the refugee crisis: On what have the Greek, German and British press focused? *European View, 15,* 265–279. [CrossRef]

Goodman, S., Sirriyeh, A. and McMahon, S. (2017). The evolving (re)categorisations of refugees throughout the "refugee/migrant crisis". *Journal of Community & Applied Social Psychology, 27*(2), 105–114. [CrossRef]

Huysmans, J. (2000). The European Union and the securitization of migration. *Journal of Common Market Studies, 38*(5), 751–777. [CrossRef]

İçduygu, A., & Aksel, D. B. (2012). *Irregular Migration in Turkey.* IOM Migration Research Series.

İçduygu, A., & Demiryontar, B. (2022). The state and Syrian refugees in Turkey: Contradictory policies and instrumental rationality. *Journal of Social Sciences of Istanbul Commerce University, 21*(Special Issue), 316–334. [CrossRef]

Kabacaoğlu, S., & Memişoğlu, F. (2021). Political leaders' discourses and the securitization of migration: A comparison of Turkey and the United States. *Journal of Humanity and Society, 11*(3), 1–38. [CrossRef]

Karyotis, G. (2007). European migration policy in the aftermath of September 11. Innovation: *The European Journal of Social Science Research, 20*(1), 1–17. [CrossRef]

Koser, K. (2007). *International Migration: A Very Short Introduction.* Oxford University Press. [CrossRef]

Koser, K. (2010). Dimensions and Dynamics of Irregular Migration. *Population, Space and Place, 16,* 181–193. [CrossRef]

Lara, R. de la L. R. (2022). Managing Irregularized Migration in Mexico: Rhetoric of a Renewed Approach. *Journal of Borderlands Studies, 39*(1), 1–22. [CrossRef]

McDonald, M. (2008). Securitization and the Construction of Security. European Journal of International Relations, 14(4), 563–587. [CrossRef]

Memişoğlu, F., & Başol, A. (2015). Turkish media and the 2015 refugee crisis: Coverage and framing. Alternatif Politik, 11(1), 95–120.

Norman, K. P. (2020). Migration Diplomacy and Policy Liberalization in Morocco and Turkey. The International Migration Review, 54(4), 1158–1183. [CrossRef]

Oğuz-Gök, G. (2016). Whose Security? Analyzing the Relationship between international migration and security through the role of International Organizations. *Kocaeli Üniversitesi Sosyal Bilimler Dergisi, 31,* 65–82. [Turkish]

Öner, A. Ş. (2021). Key concepts in migration studies. In G. Ihlamur-Öner & A. Ş. Öner (Eds.), *Migration in the age of globalization: Concepts and debates* (5th ed., pp. 13–28). İletişim Publications. [Turkish]

Özey, R. (2017). *The geopolitics of Turkey: Transformations and priorities.* Pegem Academy.

Rules of Procedure of the Grand National Assembly of Turkey. (2015). TGNA Session No: 1. Available at: https://www.tbmm.gov.tr/files/

anayasa/docs/ictuzuk/11-1973(Yururlukteki_Ictuzuk)/1973-y%C3%BCr%C3%BCrl%C3%BCkteki%20metin/ictuzuk_20181116.pdf Accessed on 2015

Rules of Procedure of the Grand National Assembly of Turkey. (2015). TGNA Session No: 63. Available at: https://www.tbmm.gov.tr/files/anayasa/docs/ictuzuk/11-1973(Yururlukteki_Ictuzuk)/1973-y%C3%BCr%C3%BCrl%C3%BCkteki%20metin/ictuzuk_20181116.pdf Accessed on Feb 24, 2015

Şemşit, S. (2010). The securitization and externalization of European Union migration policy: Implications for Turkey [Doctorial thesis]. Dokuz Eylul University Institute of Social Sciences. [Turkish]

Van der Valk, I. (2003). Right-wing parliamentary discourse on immigration in France. Discourse & Society, 14(3), 309–348. [CrossRef]

Vogel, D. (2016). The challenge of irregular migration. In A. Triandafyllidou (Ed.), Routledge handbook of immigration and refugee studies (1st ed., pp. 333–340). Routledge.

Wæver, O. (1995). Securitization and desecuritization. In R. D. Lipschutz (Ed.), On security (pp. 46-86). Columbia University Press.

Wodak, R., & Meyer, M. (2001). Methods of critical discourse Analysis. Sage Publications. [CrossRef]

Yıldız Social Science Review, Vol. 12, No. 1, pp. 40–54, 2026

Yıldız Social Science Review
Web site information: https://yssr.yildiz.edu.tr
DOI: 10.51803/yssr.1937910

Original Article / Orijinal Makale

AI-Enabled Drug Discovery Platforms: Navigating the Confluence of Software, Medical Device, and Pharmaceutical Regulation in Sino-African Trade Relations

Yapay Zeka Destekli İlaç Keşif Platformları: Çin-Afrika Ticaret İlişkilerinde Yazılım, Tıbbi Cihaz ve İlaç Mevzuatının Kesişim Noktasında Yol Bulmak

Klemens KATTERBAUER[a], Sema YILMAZ[b], Rahmi Deniz ÖZBAY[c]

[a]Department of Earth Science and Global Management, Euclid University, Banjul Republic of The Gambia, West Africa
[b]Department of Economics, Yıldız Technical University, İstanbul, Türkiye
[c]Department of Economics, İstanbul Ticaret University, İstanbul, Türkiye

[a]Yeryüzü Bilimleri ve Küresel Yönetim Bölümü, Euclid Üniversitesi, Banjul, Gambiya, Batı Afrika
[b]İktisat Bölümü, Yıldız Teknik Üniversitesi, İstanbul, Türkiye
[c]İktisat Bölümü, İstanbul Ticaret Üniversitesi, İstanbul, Türkiye

ARTICLE INFO

Article history
Received: 25 April 2026
Revised: 15 May 2026
Accepted: 8 June 2026

Key words:
AI-enabled drug discovery, medical device regulation, pharmaceutical governance, regulatory classification, Sino-African trade

MAKALE BİLGİSİ

Makale Hakkında
Geliş tarihi: 25 Nisan 2026
Revizyon tarihi: 15 Mayıs 2026
Kabul tarihi: 08 Haziran 2026

Anahtar kelimeler:
Yapay zeka destekli ilaç keşfi, tıbbi cihaz mevzuatı, ilaç sektöründe yönetişim, düzenleyici sınıflandırma, Çin-Afrika ticareti

ABSTRACT

The integration of artificial intelligence into drug discovery platforms presents a fundamental challenge to established regulatory taxonomies worldwide, raising a deceptively simple question with profound implications: Should AI-enabled drug discovery platforms be regulated as software, as medical devices, or as pharmaceuticals? This article examines this regulatory trilemma through the dual lens of comparative global frameworks and Sino-African trade relations. The analysis reveals that major jurisdictions—including the United States, European Union, and China—have adopted hybrid approaches that defy simple categorization, with the FDA emphasizing Predetermined Change Control Plans for adaptive algorithms, the EU layering AI Act requirements atop Medical Device Regulation frameworks, and China advancing a specialized AI medical device guidance architecture. For African regulatory environments, where the African Medicines Agency is newly operational and most National Regulatory Authorities function at WHO Maturity Level 3, this global fragmentation presents both peril and opportunity. The article argues that China's emergence as a leading developer of AI-enabled drug discovery technologies, combined with its deepening pharmaceutical trade relationships across Africa under FOCAC frameworks, creates unprecedented possibilities for regulatory leapfrogging through Sino-African cooperation. By examining trade patterns, investment flows, and capacity-building initiatives, this research demonstrates that the regulatory classification question is not merely a technical matter of legal taxonomy but a strategic determinant of market access, technology transfer, and ultimately, health equity

***Sorumlu yazar / Corresponding author**
*E-mail adress: sygenc@yildiz.edu.tr

across the African continent. The article concludes with concrete recommendations for harmonized approaches, South-South regulatory cooperation mechanisms, and African-specific adaptation of global AI governance principles.

Cite this article as: Katterbauer, K., Yılmaz, S., & Özbay, R. D. (2026). AI-Enabled Drug Discovery Platforms: Navigating the Confluence of Software, Medical Device, and Pharmaceutical Regulation in Sino-African Trade Relations. *Yıldız Social Science Review, 12*(1), 40–54.

ÖZ

Yapay zekanın ilaç keşif platformlarına entegrasyonu, dünya genelinde yerleşik düzenleyici sınıflandırmalara yönelik temel bir meydan okuma ortaya çıkarmakta ve son derece önemli sonuçlar doğuran aldatıcı derecede basit bir soruyu gündeme getirmektedir: Yapay zeka destekli ilaç keşif platformları yazılım olarak mı, tıbbi cihaz olarak mı yoksa farmasötik ürün olarak mı düzenlenmelidir? Bu makale, söz konusu düzenleyici açmazı karşılaştırmalı küresel çerçeveler ve Çin-Afrika ticari ilişkileri perspektifinden incelemektedir. Analiz, Amerika Birleşik Devletleri, Avrupa Birliği ve Çin dâhil olmak üzere başlıca yargı alanlarının, basit sınıflandırmaları aşan hibrit yaklaşımlar benimsediğini ortaya koymaktadır. Bu kapsamda Amerika Birleşik Devletleri Gıda ve İlaç Dairesi, uyarlanabilir algoritmalar için önceden belirlenmiş değişiklik kontrol planlarını vurgularken; Avrupa Birliği, Yapay Zeka Yasası gerekliliklerini Tıbbi Cihaz Düzenlemesi çerçevelerine eklemekte; Çin ise yapay zeka tabanlı tıbbi cihazlara yönelik uzmanlaşmış bir rehberlik yapısı geliştirmektedir. Afrika'daki düzenleyici ortamlar açısından bakıldığında, Afrika İlaç Ajansı'nın yeni faaliyete geçmiş olması ve ulusal düzenleyici otoritelerin çoğunluğunun Dünya Sağlık Örgütü Olgunluk Seviyesi 3 düzeyinde faaliyet göstermesi nedeniyle, bu küresel parçalanmış yapı hem riskler hem de fırsatlar yaratmaktadır. Makale, Çin'in yapay zeka destekli ilaç keşif teknolojilerinin önde gelen geliştiricilerinden biri hâline gelmesinin ve Çin-Afrika İş Birliği Forumu çerçevesinde Afrika genelinde derinleşen farmasötik ticaret ilişkilerinin, Çin-Afrika iş birliği aracılığıyla düzenleyici sıçrama için benzeri görülmemiş imkânlar sunduğunu ileri sürmektedir. Ticaret örüntülerini, yatırım akışlarını ve kapasite geliştirme girişimlerini inceleyen bu araştırma, düzenleyici sınıflandırma meselesinin yalnızca hukuki sınıflandırmaya ilişkin teknik bir konu olmadığını; aynı zamanda pazar erişimi, teknoloji transferi ve nihayetinde Afrika kıtasındaki sağlık eşitliğini belirleyen stratejik bir unsur olduğunu göstermektedir. Makale, uyumlaştırılmış yaklaşımlar, Güney-Güney düzenleyici iş birliği mekanizmaları ve küresel yapay zeka yönetişimi ilkelerinin Afrika'ya özgü uyarlanmasına yönelik somut önerilerle sonuçlanmaktadır.

Atıf için yazım şekli: Katterbauer, K., Yılmaz, S., & Özbay, R. D. (2026). AI-Enabled Drug Discovery Platforms: Navigating the Confluence of Software, Medical Device, and Pharmaceutical Regulation in Sino-African Trade Relations. *Yıldız Social Science Review, 12*(1), 40–54.

1. INTRODUCTION

Artificial intelligence has moved decisively from experimental curiosity to regulatory priority in the life sciences sector. On January 14, 2026, the United States Food and Drug Administration and the European Medicines Agency jointly released ten guiding principles for Good AI Practice in drug development, signalling that AI applications across the medicines lifecycle—from early discovery through post-market surveillance—now demand coherent regulatory treatment. This landmark collaboration, building upon the Good Machine Learning Practice principles established in 2021, represents transatlantic consensus that AI requires dedicated regulatory attention. Yet beneath this apparent convergence lies persistent fragmentation regarding the most fundamental question: What exactly is an AI-enabled drug discovery platform from a regulatory perspective? The classification question carries enormous practical stakes. Software applications are generally subject to lighter-touch oversight focused on functionality and cybersecurity. Medical devices, by contrast, trigger conformity assessment requirements, clinical evaluation obligations, and post-market surveillance regimes that can add years and millions of dollars to development timelines. Pharmaceuticals face the most intensive scrutiny of all, requiring exhaustive demonstrations of safety, efficacy, and manufacturing quality that define the drug approval paradigm. An AI platform that identifies novel drug candidates, predicts protein structures,

or optimizes molecular properties sits uncomfortably at the intersection of all three categories, eliding the tidy boundaries upon which regulatory systems have been constructed over decades. The question becomes still more complex when examined through the prism of international trade and regulatory cooperation, particularly between China and Africa. China has emerged as a global leader in AI-enabled drug discovery technologies, supported by deliberate industrial policy, substantial research investment, and a regulatory framework that the National Medical Products Administration has developed through over 19 guidance documents and more than 110 approved Class III AI medical device products. Simultaneously, Chinese pharmaceutical firms are accelerating their African market presence through manufacturing localization, technology transfer, and trade relationships formalized under the Forum on China-Africa Cooperation. Meanwhile, African regulatory systems are undergoing transformative change with the operationalization of the African Medicines Agency and the gradual attainment of WHO Maturity Level 3 status by a growing cohort of National Regulatory Authorities. This convergence of technological disruption, regulatory evolution, and shifting trade patterns creates both unprecedented opportunities and novel risks. For African regulators and patients, the classification question determines whether AI-enabled drug discovery platforms developed in China can enter African markets efficiently, under what evidentiary standards, and with what assurances of safety and performance in local populations. For Chinese firms, it determines market access pathways, compliance burdens, and the viability of technology transfer arrangements. And for the broader project of global health equity, it will shape whether AI's transformative potential in accelerating drug discovery reaches populations long underserved by traditional pharmaceutical innovation. This article proceeds in five parts. Section 2 examines the global regulatory landscape, comparing approaches in the United States, European Union, and China to AI-enabled drug discovery platforms, with particular attention to classification frameworks and emerging convergence points. Section 3 analyzes the current state and trajectory of Sino-African pharmaceutical trade and regulatory cooperation. Section 4 investigates the specific implications of regulatory classification choices for trade flows, technology transfer, and health outcomes in African contexts. Section 5 proposes cooperative mechanisms and policy recommendations for navigating the classification trilemma. A brief conclusion reflects on the broader significance of this moment for regulatory governance in an era of algorithmic medicine.

2. THE GLOBAL REGULATORY TRILEMMA: SOFTWARE, DEVICE, OR PHARMACEUTICAL

2.1. The Fundamental Classification Problem

The regulatory classification of AI-enabled drug discovery platforms is complicated by the fact that these technologies can manifest across multiple phases of the pharmaceutical lifecycle, each traditionally governed by distinct regulatory regimes. A single AI system might screen compound libraries during early discovery (traditionally unregulated or lightly regulated as research), predict toxicity during preclinical development (potentially triggering good laboratory practice requirements), stratify patients during clinical trials (implicating investigational device or drug regulations), and monitor safety post-market (engaging pharmacovigilance obligations). This functional fluidity defies the static classification assumptions embedded in most regulatory frameworks (Lognoul 2025). The EU AI Act, which entered into force as Regulation 2024/1689, adopts a horizontal, risk-based approach that applies across all sectors, including life sciences (Act 2024). Under this framework, AI systems that constitute or form safety components of medical devices subject to the Medical Device Regulation or In Vitro Diagnostic Medical Device Regulation are classified as "high-risk" and subject to stringent obligations including data governance, transparency, and human oversight requirements. Critically, however, the AI Act does not duplicate MDR obligations but rather layers AI-specific requirements atop existing conformity assessment procedures. For drug discovery platforms that do not meet the definition of a medical device—for instance, those used purely for early-stage molecular screening without direct patient application — the regulatory pathway remains less clearly defined, potentially falling into lower-risk categories or escaping specific AI Act provisions altogether. The United States Food and Drug Administration has approached the classification challenge through a different lens, grounded in its established framework for Software as a Medical Device. The FDA's guidance distinguishes between software intended for administrative, financial, or operational purposes (generally not regulated as medical devices) and software with medical purposes including diagnosis, treatment, or prevention. For AI-enabled drug discovery, the key determinant is the "intended use" articulated by the developer: platforms used solely for internal research and development may fall outside FDA jurisdiction, while those marketed to third parties for clinical trial optimization or treatment selection likely trigger device regulations. The FDA has recognized that this binary approach strains when applied to adaptive AI systems, prompting the development of novel regulatory mechanisms including the Predetermined Change Control Plan framework, which allows manufacturers to prospectively specify planned algorithmic modifications without requiring new marketing submissions for each iteration (Group 2022).

2.2. The European Union's Layered Approach

The European framework exemplifies the complexity of layering new AI-specific requirements onto established product-specific regulations. Under the MDR, stand-alone

software qualifies as a medical device if it serves a medical purpose and is intended for use with individual patients. An AI system that analyzes patient data to recommend specific therapeutic interventions would clearly meet this definition, triggering both MDR conformity assessment and AI Act high-risk classification. However, an AI platform used exclusively for in silico compound screening during early drug discovery—before any patient data is involved and before any specific therapeutic candidate is identified—likely falls outside both MDR and AI Act high-risk provisions, governed instead by general product safety and liability laws alongside emerging standards for scientific AI.

This layered approach creates significant compliance complexity but offers important nuance. The AI Act's requirements for high-risk systems—including robust data governance, technical documentation, record-keeping, transparency, human oversight, and accuracy specifications—address legitimate concerns about algorithmic bias, performance drift, and explainability that are particularly salient in life sciences applications. Yet by tying these requirements to the medical device definition, the EU framework leaves substantial AI activity in drug discovery either lightly regulated or subject to interpretive uncertainty. This regulatory gap may prove problematic as AI systems increasingly automate critical decisions throughout the drug development pipeline, from target identification through clinical trial design. Intellectual property considerations further complicate the European landscape. The European Patent Office requires disclosure sufficient to enable a person skilled in the art to carry out the invention, which for AI-based inventions increasingly requires disclosure of training data—information that may be commercially sensitive or subject to data protection constraints. The EPO's requirement that inventors be natural persons also creates unresolved questions regarding AI-generated drug candidates or molecular designs . These IP uncertainties intersect with regulatory classification: if AI systems cannot be recognized as inventors, and if their training data cannot be adequately disclosed, the resulting drug candidates may face both patent protection gaps and regulatory data exclusivity challenges.

2.3. The United States' PCCP Innovation

The FDA's response to AI-enabled medical devices, including those used in drug development, has been shaped by recognition that traditional regulatory paradigms designed for static products cannot accommodate adaptive algorithms that learn and evolve. As of August 2025, the FDA's AI/ML-enabled medical device inventory included 1,247 devices, with the vast majority (1,195) cleared through the 510(k) pathway, 36 through De Novo classification, and 16 through the more rigorous PMA process . This distribution reflects both the relative novelty of AI/ML technologies and the FDA's reliance on the substantial equivalence framework for moderate-risk devices (Rosen and Mandl 2025).

The Predetermined Change Control Plan represents the FDA's most significant regulatory innovation for adaptive AI. Under the PCCP framework, manufacturers can prospectively specify planned modifications to AI-enabled device software functions, including the modification description, the protocol for developing and validating changes, and an impact assessment analyzing benefits and risks. FDA approval of the PCCP as part of the initial marketing submission then permits implementation of specified modifications without additional regulatory review . This approach addresses the fundamental tension between algorithmic adaptability and regulatory predictability, allowing responsible innovation while maintaining oversight of safety and effectiveness. For drug discovery platforms specifically, the PCCP framework offers a potential pathway for managing iterative improvements to predictive models without triggering repeated regulatory submissions. A platform that identifies novel drug-target interactions, for instance, could specify in advance how it will incorporate new training data, validate model updates, and assess performance drift. However, the PCCP framework was developed primarily for clinical diagnostic and monitoring applications; its extension to drug discovery tools that do not directly impact patient care remains somewhat ambiguous. Moreover, the FDA's guidance explicitly notes that PCCP is only one component of an AI-enabled device submission and does not eliminate the need for comprehensive demonstration of safety and effectiveness.

2.4. China's Emerging AI Medical Device Framework

China has moved with characteristic speed and scale to establish regulatory frameworks for AI in healthcare and life sciences. The National Medical Products Administration has issued over 19 guidance documents and review principles for AI medical devices, forming a five-tiered guidance system covering everything from general principles to modality-specific testing standards. As of mid-2025, NMPA had approved more than 110 Class III AI medical device standalone software products spanning auxiliary diagnosis, triage, detection, and treatment support across nine imaging modalities . This represents the largest cohort of approved AI medical devices globally and provides China with substantial regulatory experience. The Chinese framework classifies AI-enabled medical software according to established medical device risk categories, with Class II and Class III devices requiring registration approval, change applications for significant modifications, and quality management system-based handling of minor changes. The NMPA has established a dedicated AI Medical Device Standardization Technical Committee (SMD/TU 002) that has published eight industry standards as of 2025, with additional national and industry standards under development covering algorithm performance testing for specific clinical applications . This standardization infrastructure provides a foundation for predictable, transparent review that Chinese authorities are actively promoting through innovation cooperation platforms

linking industry, academia, clinical institutions, and regulators. Crucially for international trade considerations, China has aligned its pharmaceutical review standards with International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use guidelines, creating conditions for global synchronized development, submission, and approval . This alignment, combined with China's extensive experience regulating AI medical devices, positions Chinese regulators as potential technical cooperation partners for African agencies developing their own AI governance frameworks. The question is whether China's approach—developed primarily for clinical diagnostic and therapeutic AI—adequately addresses the distinctive characteristics of drug discovery platforms, or whether further regulatory innovation is needed.

2.5. Cross-Cutting Themes and Global Divergence

Several cross-cutting themes emerge from comparative analysis of global regulatory approaches. First, risk proportionality serves as an organizing principle across all major jurisdictions, though the specific risk thresholds and corresponding requirements vary substantially. Second, intended use remains the critical determinant of regulatory pathway, placing enormous weight on developer representations that may not fully capture actual system capabilities or downstream applications. Third, adaptivity presents universal challenges, with the PCCP concept gaining international traction but not yet achieving full global harmonization. Fourth, data governance requirements increasingly transcend traditional regulatory categories, with expectations for training data quality, representativeness, and documentation appearing across software, device, and pharmaceutical contexts (Muehlematter et al. 2021). Despite these common themes, significant divergence persists. The EU's AI Act represents the most comprehensive horizontal AI legislation globally, layering cross-sectoral requirements atop vertical product regulations. The US approach remains more sectoral, with the FDA leading on medical applications while broader AI governance remains fragmented across agencies and state-level initiatives. China has developed a distinctive model emphasizing standardization, industrial policy coordination, and staged implementation. For African regulators and their Chinese counterparts, this global fragmentation creates both challenges—in determining which models to follow—and opportunities—in selectively adapting elements suited to local contexts.

3. THEORETICAL FRAMEWORK: REGULATORY CLASSIFICATION, TECHNOLOGICAL GOVERNANCE, AND SOUTH-SOUTH HEALTH COOPERATION

3.1. The Regulatory Trilemma as a Theoretical Problem

At the heart of this article lies what we term the *regulatory trilemma*: the inability of existing product-based regulatory taxonomies—software, medical device, pharmaceutical—to accommodate the functional, adaptive, and cross-cutting nature of AI-enabled drug discovery platforms. This trilemma is not merely a technical classification puzzle but a manifestation of deeper theoretical tensions in the governance of emerging technologies.

Drawing on the work of Jasanoff (2016) on *sociotechnical imaginaries*, we argue that regulatory classifications are not neutral technical categories but rather institutional expressions of collective visions about how technology should relate to society. The FDA's PCCP framework imagines AI as a continuously improving but controllable tool; the EU AI Act imagines AI as a horizontal risk that must be layered onto existing product regimes; China's NMPA guidance architecture imagines AI as an industrial-strategic asset requiring standardization and state-led coordination. Each imaginary produces different classification outcomes, compliance burdens, and market access conditions. For African regulators, the challenge is not to select the "correct" imaginary but to construct a context-sensitive hybrid that serves continental health priorities.

3.2. Theories of Technological Regulation: From Product to Process

Traditional regulatory theory has largely been organized around *product-based* paradigms: a discrete artifact (a pill, a device, a software executable) undergoes pre-market review, is approved for specified indications, and is subject to post-market surveillance. AI-enabled drug discovery platforms disrupt this model in three theoretically significant ways.

First, they are *process technologies* rather than end products. An AI platform that screens compound libraries or predicts protein folding does not directly interact with patients but rather shapes the knowledge base and material artifacts (drug candidates) that subsequently enter clinical pathways. This locates the platform upstream of traditional regulatory trigger points, creating what Collingridge (1980) famously termed the *dilemma of control*: early in a technology's development, when intervention is easiest, consequences are least predictable; later, when consequences become visible, control becomes difficult and costly.

Second, they exhibit *algorithmic adaptivity*, learning and evolving post-deployment. This challenges the static "locked" product assumption embedded in most regulatory frameworks. The PCCP framework represents a pragmatic response, but it remains rooted in a manufacturer-controlled change paradigm that does not fully accommodate continuous, autonomously initiated learning. Drawing on *infrastructure studies* (Star & Ruhleder, 1996), we suggest that AI drug discovery platforms are better theorized as *emergent infrastructures*: relational, evolving, and only visible upon breakdown.

Third, they generate *epistemic spillovers* across regulato-

ry domains. A platform's training data, validation protocols, and performance metrics simultaneously implicate software quality standards, medical device clinical evidence requirements, and pharmaceutical good practice guidelines. This blurring of boundaries suggests that traditional siloed regulatory architectures are increasingly mismatched to the technologies they seek to govern.

3.3. Legal Taxonomy and Hybrid Governance

Legal taxonomy—the classification of objects into pre-existing legal categories—has long been recognized as a site of interpretive struggle (Kennedy, 1986). In the context of emerging technologies, taxonomical contests are particularly consequential because classification determines which institutional logic applies. AI-enabled drug discovery platforms are *boundary objects* (Star & Griesemer, 1989): they are simultaneously software (algorithmic code), devices (used for clinical decision support), and pharmaceutical inputs (generating novel molecular entities). Their boundary-crossing nature enables them to be adopted by diverse communities but also exposes them to conflicting regulatory expectations.

The concept of *hybrid governance* offers a way forward. Rather than forcing a single categorical assignment, hybrid governance approaches layer or combine regulatory logics based on function, risk, and context. The EU AI Act's layering of AI-specific requirements atop MDR conformity assessment is one such hybrid. The proposed Sino-African framework—differentiating internal research use, regulatory decision-support, and autonomous drug generation—represents another. Hybrid governance acknowledges that technological objects can simultaneously inhabit multiple regulatory categories and that proportionate oversight requires functional decomposition rather than monolithic classification.

3.4. Regulatory Capacity, Leapfrogging, and South-South Cooperation

From the perspective of international political economy, the classification trilemma intersects with questions of *regulatory capacity* and *technological leapfrogging*. African NRAs operating at WHO Maturity Level 3 have established pharmaceutical regulatory systems but lack specific AI governance experience. This capacity gap, often framed as a deficit, can also be understood as *institutional plasticity*: the opportunity to design AI governance frameworks from first principles, learning from the achievements and limitations of more mature systems without inheriting legacy path dependencies.

The concept of *regulatory leapfrogging* (cf. Perkins, 2003) suggests that latecomer regulators can bypass suboptimal intermediate stages by adopting directly the most advanced, context-appropriate practices. For AI in drug discovery, this could mean African agencies developing risk-proportionate, process-oriented, and data-sovereignty-respecting frameworks that avoid both the under-regulation of software-only approaches and the over-regulation of pharmaceutical pathways. South-South cooperation—particularly with China, which has substantial AI medical device approval experience but also shares developing-country perspectives on regulatory capacity constraints—offers a distinctive avenue for leapfrogging that bypasses the sometimes-prescriptive North-South technical assistance models.

3.5. Algorithmic Justice and Distributive Implications

Finally, any theoretical treatment of AI drug discovery regulation must engage with *distributive justice* and *algorithmic fairness* as applied to global health. Existing theories of algorithmic justice (Benjamin, 2019; Noble, 2018) have primarily focused on discrimination within high-income country contexts—biased credit scoring, facial recognition, or healthcare allocation. The global dimension introduces additional complexity: training data predominantly from well-resourced populations, research priorities shaped by commercial pharmaceutical markets, and regulatory standards calibrated to high-capacity health systems.

Drawing on *postcolonial science studies* (Anderson & Adams, 2008; Harding, 1998), we argue that the classification question is embedded in broader patterns of epistemic authority and material inequality. When African regulators adopt medical device classification requiring local clinical validation, they are not merely adding compliance burden but actively contesting the presumption that AI platforms validated in Shanghai or Silicon Valley will perform equivalently in Lagos or Nairobi. The theoretical contribution of this article is to demonstrate that regulatory classification is not a neutral technical exercise but a site of distributive politics—determining who bears the costs of validation, who benefits from accelerated access, and whose health needs shape algorithmic optimization.

3.6. Synthesis: Toward a Situated Regulatory Theory

The theoretical framework developed here rejects both technological determinism (the view that AI's inherent properties dictate a single correct classification) and legal formalism (the view that existing categories can be mechanically applied). Instead, we propose a *situated regulatory theory*: classification decisions should be informed by functional analysis of the technology, empirical assessment of risks and benefits in specific contexts, procedural legitimacy of regulatory decision-making, and distributive attention to health equity outcomes. For Sino-African cooperation, this implies that regulatory alignment should not aim for identical rules but rather for *mutual recognition of regulatory processes*—trust in each other's institutional competence to reach context-appropriate classifications, combined with mechanisms for work-sharing and capacity building.

4. SINO-AFRICAN PHARMACEUTICAL TRADE AND REGULATORY COOPERATION

4.1. Current Trade Patterns and Investment Flows

Pharmaceutical trade between China and Africa has accelerated substantially over the past decade, driven by complementary needs and capacities. Africa's healthcare market, projected to reach $259 billion by 2025 according to UNECA estimates, confronts significant supply constraints: the continent hosts fewer than 400 pharmaceutical manufacturers, produces only 3% of global medicines, and imports over 90% of its pharmaceutical and medical device requirements. China, as the world's largest producer of active pharmaceutical ingredients and a rapidly growing innovator in finished formulations and medical devices, represents a natural trade partner for addressing African supply deficits. The pattern of Chinese pharmaceutical engagement in Africa is evolving from pure product export toward manufacturing localization. Early entrants established production facilities in strategically selected markets: Shanghai Pharmaceutical in Sudan, Humanwell Healthcare in Mali, Fosun Pharma in Côte d'Ivoire, and HEC Pharm in Algeria represent notable examples in the pharmaceutical sector. In medical devices, BGI Genomics has established operations in Angola, Wanbangde Group in South Africa, and Wondfo Biotech in Kenya . This localization trend reflects both African government incentives for domestic manufacturing—motivated by health security concerns exposed during the COVID-19 pandemic—and Chinese firms' recognition that sustainable market access increasingly requires on-the-ground presence (Aniche 2023).

Recent high-level engagements underscore the priority both sides attach to pharmaceutical cooperation. In November 2025, Minister Counselor Lyu Ruihao of China's Mission to the African Union led a delegation to AU institutions in Ghana and Rwanda, where discussions with African Medicines Agency Director General Dr. Delese Mimi Darko focused on establishing cooperation mechanisms between Chinese institutions and the AMA to enhance drug regulatory capacity and medicine access across Africa . The Chinese delegation expressed support for AMA development within the FOCAC framework, while Dr. Darko signaled openness to active collaboration with Chinese regulatory and industry partners.

The dual-panel figure (Fig. 1) quantifies China's regulatory experience with AI-enabled medical products, establishing the evidentiary foundation for claims about Chinese regulatory capacity. The horizontal bar chart enables precise comparison of approval volumes across modalities, while the pie chart emphasizes the dominance of radiology applications. Together, they demonstrate that Chinese regulators have developed substantial practical competence in evaluating AI/ML algorithms, assessing clinical validation data, and establishing post-market monitoring requirements. This experience is directly relevant to AI-enabled drug discovery platforms, which share many

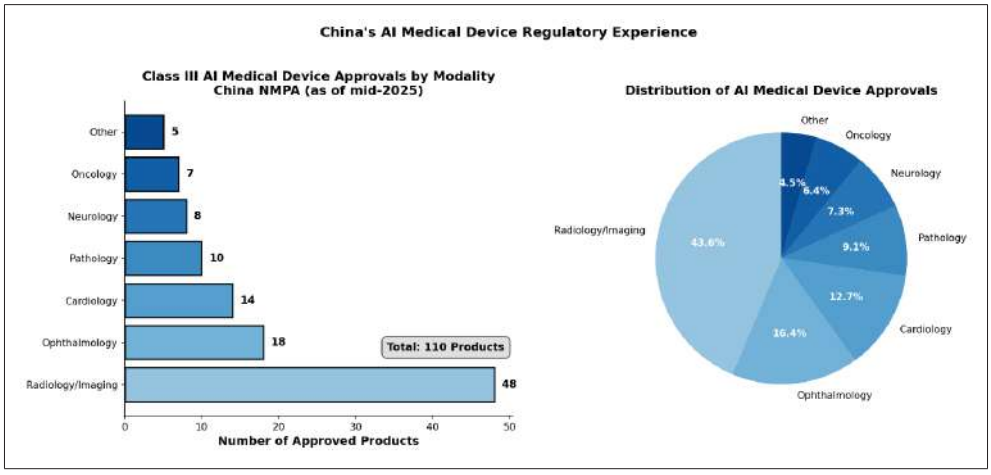

Figure 1. Distribution of China National Medical Products Administration (NMPA) Class III AI medical device approvals by clinical modality as of mid-2025. Panel A (horizontal bar chart) displays absolute approval counts across nine modalities, with radiology and medical imaging representing the largest cohort (48 products, 43.6% of total). Panel B (pie chart) illustrates the proportional distribution, highlighting the concentration of regulatory experience in imaging-based applications. Total approved products exceed 110, representing the largest cohort of regulated AI medical devices globally. Data compiled from NMPA public announcements and industry reports through June 2025. This extensive regulatory experience provides China with substantial practical expertise in AI/ML-enabled medical product evaluation, quality system oversight, and post-market surveillance that may inform technical cooperation with African regulatory authorities (Zhang and Yan 2026).

technical characteristics with approved diagnostic AI systems. The figure supports the article's recommendation that African regulators can benefit from Chinese regulatory science capacity-building initiatives.

3.2. African Regulatory Capacity and the African Medicines Agency

The African regulatory landscape is undergoing transformative institutional development. The African Medicines Agency, established under the African Union and operational since 2021, represents the continent's most ambitious effort to harmonize pharmaceutical regulation, reduce duplicative national reviews, and create a unified market for medicines and medical devices. The AMA's treaty mandate encompasses coordination of regulatory systems, scientific evaluation of selected products, and strengthening of national regulatory authorities through capacity building and reliance mechanisms (Abdulwahab et al. 2024).Complementing AMA's continental mandate, the African Medicines Regulatory Harmonization initiative has advanced regulatory convergence through regional economic communities, establishing joint assessment procedures, common technical document requirements, and information sharing mechanisms. A growing cohort of African NRAs has achieved WHO Maturity Level 3 status—indicating stable, well-functioning, and integrated regulatory systems—including agencies in Egypt, Ghana, Nigeria, Rwanda, Senegal, South Africa, Tanzania, and Zimbabwe . These ML3 agencies are increasingly entering mutual recognition and work-sharing arrangements, creating a foundation for coordinated oversight of complex technologies including AI-enabled products (Cerf 2018).

However, significant capacity gaps persist. No African NRA has yet issued comprehensive guidance specifically addressing AI in drug development or medical devices, though several agencies are actively developing frameworks. South Africa's Health Products Regulatory Authority published a communication in September 2025 outlining regulatory requirements for AI/ML-enabled medical devices, referencing international best practices and signaling expectations for ISO 13485 compliance, clinical evidence, and post-market surveillance . SAHPRA's approach—requiring establishment licensing for AI/ML device importers and manufacturers, applying the four-tier risk classification system, and monitoring international PCCP developments—illustrates how African regulators are pragmatically adapting global standards while building local capacity (African Yearbook Of International L 2019).

3.3. China-Africa Regulatory Cooperation Mechanisms

Regulatory cooperation between Chinese and African authorities has advanced through multiple channels. At the multilateral level, the FOCAC framework provides overarching political direction, with successive action plans identifying pharmaceutical manufacturing, regulatory harmonization, and health product access as priority cooperation areas. The November 2025 China-Africa Pharmaceutical Innovation and International Cooperation Forum in Wuhan convened regulatory leaders from Tanzania, Ethiopia, Uganda, Rwanda, and China's Hubei provincial authorities to discuss drug registration, clinical trial cooperation, and local manufacturing . Such sub-national initiatives complement central government engagements, creating multiple pathways for regulatory dialogue (Li et al. 2023). Chinese regulatory capacity-building support for African counterparts is emerging as a significant cooperation modality. The NMPA's experience with AI medical device regulation—documented through its extensive guidance system and over 110 product approvals—offers a potential template for African agencies developing their own frameworks. Chinese authorities have explicitly acknowledged tracking international regulatory developments from IMDRF, ITU/WHO Focus Group AI for Health, and individual agencies including US FDA, Japan's PMDA, Korea's MFDS, and Singapore's HSA, while also indicating openness to "further optimizing AI medical device regulatory approaches within China's legal framework" . This reflexive regulatory posture—learning from global peers while adapting to local conditions—may prove particularly resonant for African regulators navigating similar challenges of resource constraints and development priorities.

The operationalization of formal cooperation between Chinese institutions and the AMA remains at an early stage. Minister Counselor Lyu's November 2025 discussions with Director General Darko produced expressions of mutual interest but have not yet yielded concrete institutional mechanisms . The potential scope for such cooperation is substantial, encompassing joint training programs, regulatory science research collaborations, work-sharing arrangements for product reviews, and technical assistance in developing guidance documents for emerging technologies including AI.

The time-series visualization (Fig. 2) establishes the economic significance of Sino-African pharmaceutical trade as context for the regulatory classification discussion. The dual presentation of historical and projected data enables viewers to assess both achieved growth and anticipated trajectory. The separation of API exports from total exports is analytically significant: API trade represents inputs for African pharmaceutical manufacturing, while finished formulation exports may compete with local production. The figure supports the article's argument that regulatory classification choices affecting market access for AI-enabled platforms will have material economic consequences given the scale and growth rate of this trade relationship. The CAGR annotation provides a quantitative summary metric for rapid comprehension.

3.4. Trade Barriers and Regulatory Friction Points

Despite expanding trade and improving regulatory cooperation, significant barriers impede the full realization of

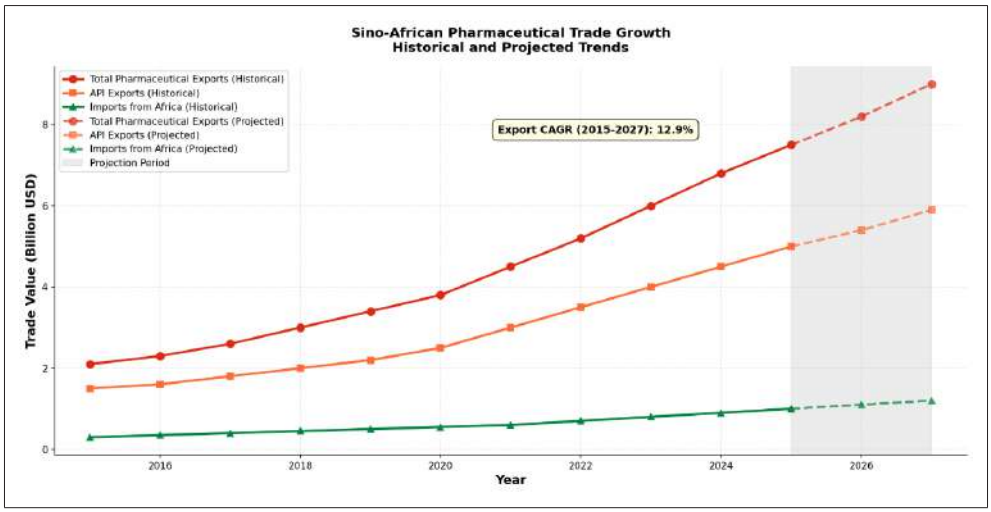

Figure 2. Historical and projected pharmaceutical trade flows between China and Africa, 2015-2027. Solid lines represent historical trade data (2015-2024); dashed lines represent projections (2025-2027) based on compound annual growth rate (CAGR) extrapolation and announced investment commitments. Three trade categories are shown: total pharmaceutical exports from China to Africa (red), active pharmaceutical ingredient (API) exports specifically (orange), and African pharmaceutical imports to China (green). The shaded region indicates the projection period. Total pharmaceutical exports are projected to reach $9.0 billion by 2027, representing a 13.1% CAGR over the full period. API exports demonstrate particularly strong growth, reflecting China's dominant position in global API manufacturing and increasing African pharmaceutical manufacturing capacity. The widening trade surplus highlights persistent African import dependence and underscores the importance of regulatory harmonization and technology transfer initiatives to support local production capacity (United Nations Conference on Trade and Development 2026).

Sino-African pharmaceutical trade potential. Product registration remains a primary friction point, with heterogeneous national requirements, variable review timelines, and limited transparency creating unpredictability for Chinese exporters. The AMA's pilot continental review procedures aim to address these challenges, but the pilot remains limited in scope and has not yet achieved full operational capacity (Ncube 2015). Market access challenges extend beyond registration to encompass procurement preferences, local content requirements, and reimbursement policies. Many African governments, seeking to stimulate domestic manufacturing, have implemented preferences for locally produced medicines in public tenders or established minimum local procurement quotas. Nigeria's Presidential Initiative for Healthcare Value Chain and Cameroon's 30% minimum local procurement requirement exemplify this trend. While these policies create incentives for Chinese firms to localize production—aligning with African industrialization objectives—they also complicate market access for Chinese exports and require substantial capital investment. Intellectual property protection presents another dimension of regulatory friction. Chinese pharmaceutical firms have historically operated in an environment with relatively flexible patent enforcement; their increasing orientation toward innovation and global markets has driven improvements in China's IP regime, but perceptions of inadequate IP protection persist among some African stakeholders. Conversely, African concerns about ensuring affordable access to essential medicines create tensions with maximalist IP enforcement. The intersection of AI and pharmaceutical IP—including questions about patentability of AI-generated inventions, disclosure requirements for training data, and trade secret protection for algorithms—adds further complexity to an already delicate balance.

5. IMPLICATIONS OF CLASSIFICATION CHOICES FOR SINO-AFRICAN TRADE AND HEALTH OUTCOMES

5.1. Market Access and Trade Flow Consequences

The regulatory classification assigned to AI-enabled drug discovery platforms carries direct implications for the volume and character of Sino-African pharmaceutical trade. If classified as software—subject primarily to general product safety requirements, cybersecurity standards, and perhaps emerging AI governance frameworks—Chinese platforms could enter African markets with relatively low regulatory friction. This pathway would maximize trade volumes and

accelerate technology diffusion, but would also concentrate regulatory scrutiny downstream on the actual drug products discovered or optimized using AI tools. The risk is that AI-specific concerns about training data representativeness, algorithmic bias, and performance drift would escape systematic regulatory oversight (Ncube 2015).

Classification as medical devices triggers a substantially more demanding compliance pathway. Chinese exporters would need to demonstrate conformity with medical device regulations in each target market—or navigate regional harmonization mechanisms where available—including quality management system certification (typically ISO 13485), technical documentation demonstrating safety and performance, clinical evidence where required, and post-market surveillance obligations. For African markets, where medical device regulatory systems vary in maturity, this could mean navigating over 50 distinct national requirements, though AMA and regional harmonization initiatives promise eventual simplification. The SAHPRA model, requiring establishment licensing for importers and applying risk-stratified review based on internationally recognized standards , suggests how African regulators may pragmatically approach this challenge (Obunike and NWAKOB 2025).

Classification as pharmaceuticals would represent the most burdensome scenario, subjecting AI platforms to drug approval pathways including clinical trial requirements, manufacturing quality standards, and pharmacovigilance obligations calibrated for molecular entities rather than computational tools. While few jurisdictions currently treat AI discovery platforms as pharmaceuticals, the question becomes acute when AI systems generate novel molecular entities with minimal human intervention. If an AI platform autonomously identifies and optimizes a drug candidate, at what point does the platform itself become subject to pharmaceutical regulation? The European Patent Office's requirement for human inventorship reflects broader discomfort with fully autonomous AI drug discovery, but regulatory frameworks have not yet systematically addressed this scenario (Nwauche 2018). For Sino-African trade, classification choices will shape not only compliance costs and timelines but also the structure of commercial relationships. A software or light-device classification favors export-oriented business models where Chinese firms develop and validate platforms in China, then deploy them globally with minimal localization. A medical device classification, particularly with requirements for clinical validation in local populations, creates incentives for deeper in-country presence, technology transfer, and joint development with African partners. This aligns with both Chinese industrial strategy—which increasingly emphasizes overseas manufacturing and local partnerships—and African aspirations for pharmaceutical sovereignty and technology capability development (Atacan and Açık 2023).

5.2. Technology Transfer and Capacity Building

The classification question profoundly influences technology transfer dynamics between Chinese developers and African adopters. Medical device classification typically requires local establishment licensing, authorized representative arrangements, and sometimes local clinical data—all of which necessitate African infrastructure and expertise. These requirements, while burdensome in the short term, can serve as catalysts for sustainable capacity building. When SAHPRA requires ISO 13485 certification for importers and manufacturers of AI/ML-enabled medical devices, it creates demand for local quality management expertise, conformity assessment services, and technical competence that did not previously exist. Chinese firms have demonstrated willingness to invest in African manufacturing and technical capacity when market access conditions reward such investment. Humanwell's pharmaceutical facilities in Mali, BGI's genomics operations in Angola, and Wondfo's diagnostics manufacturing in Kenya illustrate that localization is not merely theoretical. The key policy question is whether regulatory frameworks can be designed to maximize the developmental spillovers from technology transfer—including workforce training, quality system implementation, and research collaboration—without imposing requirements so onerous that they deter market entry altogether. The NMPA's experience with AI medical device regulation offers potential lessons for technology transfer arrangements. China's five-tiered guidance system, standardization committee structure, and innovation cooperation platforms represent institutional investments in regulatory science that could be adapted for African contexts. Chinese regulators and industry associations have already engaged in capacity-building exchanges with African counterparts, as evidenced by the August 2025 China-Africa Pharmaceutical Trade and Investment Promotion Mission to Ethiopia and Uganda, which included regulatory discussions alongside commercial meetings. Expanding such exchanges to specifically address AI governance in drug development could accelerate African regulatory readiness while creating commercial opportunities for Chinese firms (de Villiers 2021).

5.3. Health Equity and Algorithmic Justice

Perhaps the most consequential implication of regulatory classification choices concerns health equity and what might be termed "algorithmic justice"—ensuring that AI-enabled drug discovery serves diverse populations rather than exacerbating existing disparities. The FDA-EMA Good AI Practice principles emphasize that AI systems must be developed using representative datasets to avoid miscalibration when applied to populations differing from training data. This principle has particular urgency for African populations, which remain substantially underrepresented in global genomic databases, clinical trial cohorts, and the biomedical research literature that informs AI model development. The Africa Clinical Research Network

has articulated this concern with clarity, calling on regulators to "require evidence that AI tools perform well on African population data" and warning that "miscalibration caused by training exclusively on non-African populations" could undermine both safety and efficacy when AI-enabled discovery platforms are applied to diseases prevalent in Africa. This concern extends beyond clinical performance to encompass fundamental questions about which diseases receive research attention. If AI platforms are trained predominantly on data from high-income country populations and optimized for diseases of commercial interest in those markets, they may systematically underperform for neglected tropical diseases or conditions with distinct African epidemiological patterns (Antolín 2021).

Regulatory classification determines the tools available to address these concerns. Medical device regulations typically require clinical validation demonstrating performance in intended use populations—a requirement that, if properly enforced, could mandate African population validation for platforms marketed in Africa. Software regulations generally lack such requirements. Pharmaceutical regulations would trigger clinical trial obligations that, while burdensome, would generate African population data as a by-product. The challenge for African regulators is to calibrate requirements to ensure algorithmic fairness without imposing validation burdens so extensive that they discourage the application of AI to African health priorities altogether (Daneshjou et al. 2021).

5.4. Data Sovereignty and Cross-Border Governance

AI-enabled drug discovery platforms necessarily involve data flows—training data, validation data, and operational data generated through platform use—that raise complex questions of data sovereignty and cross-border governance. The African Union's Malabo Convention on Cyber Security and Personal Data Protection, which entered into force in 2023, establishes a continental framework for data protection that includes provisions on cross-border data transfers, consent requirements, and sensitive data handling. The AU Digital Transformation Strategy 2020-2030 similarly articulates aspirations for African data sovereignty and digital self-determination (Vieira et al. 2023). The classification of AI platforms intersects with data governance in several ways. Medical device and pharmaceutical classifications typically entail more stringent requirements for data integrity, traceability, and auditability—obligations that can support responsible data stewardship but may also require data localization or restricted transfer arrangements. Software classification offers greater flexibility but potentially weaker protections for sensitive health data used in AI training. Chinese firms operating across African markets must navigate not only diverse national data protection laws but also China's own data security and cross-border transfer regulations, which impose requirements on data collected overseas but processed in China (Union 2014).

The African Clinical Research Network's call to "operationalize data governance via the Malabo Convention" specifically for AI in health reflects recognition that existing legal frameworks require practical implementation guidance to address AI-specific challenges. The AMA, working with African NRAs and regional economic communities, is well-positioned to develop such guidance, potentially drawing on models from the EU's elaborate data protection regime, China's sectoral data security requirements, and emerging global standards for health AI governance.

This radar chart (Fig. 3) provides a systematic comparative evaluation of classification alternatives, moving beyond descriptive analysis to normative assessment. The multi-dimensional framework acknowledges that regulatory choices involve trade-offs across multiple values, not merely technical or legal considerations. The visual representation enables rapid identification of each classification's distinctive profile: software regulation's trade-off of safety for access, pharmaceutical regulation's inverse prioritization, and the medical device pathway's intermediate position. The hybrid approach's superior coverage across most dimensions supports the article's central recommendation that African regulators should avoid rigid categorical assignment and instead develop context-sensitive frameworks adapted to local priorities and constraints.

6. TOWARD COOPERATIVE SOLUTIONS: POLICY RECOMMENDATIONS

6.1. Pragmatic Classification for African Contexts

The preceding analysis suggests that no single classification—software, medical device, or pharmaceutical—adequately addresses the full range of regulatory concerns raised by AI-enabled drug discovery platforms. For African regulators, the optimal approach is likely a context-sensitive, risk-proportionate hybrid framework that distinguishes among different platform functions and deployment scenarios. Specifically:

For platforms used exclusively for internal research and development—for instance, AI systems that screen compound libraries or predict molecular properties during early discovery but do not directly inform clinical decisions or regulatory submissions—light-touch oversight emphasizing transparency, data governance, and algorithm documentation may suffice. Such platforms could be regulated under general research integrity frameworks and emerging AI governance standards rather than full medical device or pharmaceutical regimes.

For platforms used in regulatory decision-making—including AI systems that optimize clinical trial design, identify patient subpopulations, or support safety monitoring—a med-

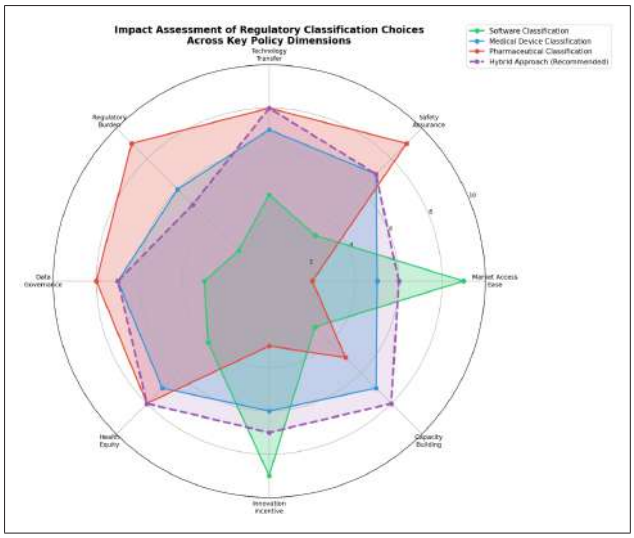

Figure 3. Radar chart comparing the relative performance of four regulatory classification approaches—Software, Medical Device, Pharmaceutical, and Hybrid—across eight policy-relevant dimensions. Each axis represents a distinct evaluative criterion scored from 0 (minimal) to 10 (maximal): Market Access Ease, Safety Assurance, Technology Transfer, Regulatory Burden, Data Governance, Health Equity, Innovation Incentive, and Capacity Building. The Software classification (green) maximizes market access ease and innovation incentives but performs poorly on safety assurance and data governance. The Pharmaceutical classification (red) provides maximal safety assurance but severely constrains market access and innovation. The Medical Device classification (blue) offers balanced intermediate performance. The recommended Hybrid approach (purple dashed line) combines elements of multiple frameworks to achieve favorable scores across all dimensions, with particular strength in technology transfer, health equity, and capacity building relevant to African contexts (Authors' own elaboration).

ical device classification with appropriate risk stratification appears most suitable. This approach would require demonstration of analytical and clinical validity, quality management systems, and post-market performance monitoring, while avoiding the full pharmaceutical development paradigm calibrated for molecular entities. The FDA PCCP framework and SAHPRA's adaptive approach offer useful models.

For platforms that autonomously generate novel drug candidates—where AI systems substantially replace human decision-making in molecular design and optimization—closer alignment with pharmaceutical regulation may be warranted, particularly regarding manufacturing quality, clinical trial requirements, and intellectual property treatment. However, this scenario remains largely hypothetical for now, and premature regulatory action could stifle innovation without commensurate public health benefit.

6.2. Sino-African Regulatory Cooperation Agenda

The convergence of Chinese technological capability, African regulatory development, and shared interest in expanded pharmaceutical trade creates a compelling case for structured Sino-African cooperation on AI drug discovery governance. Priority initiatives should include:

Joint development of Africa-adapted AI guidance. Chinese regulators, drawing on their extensive experience with AI medical device approvals and standardization, could partner with AMA and regional working groups to develop guidance documents specifically addressing AI in drug discovery for African contexts. Such guidance should emphasize representative training data, performance validation in African populations, and proportionate oversight calibrated to regulatory capacity constraints.

Regulatory science capacity building. The NMPA's AI Medical Device Standardization Technical Committee and innovation cooperation platforms offer institutional models for African capacity development. Structured exchanges—potentially under FOCAC health cooperation frameworks—could support African regulators in developing technical competence in AI/ML evaluation, including algorithmic auditing, data quality assessment, and performance monitoring.

Work-sharing and reliance mechanisms. As African NRAs achieve higher WHO maturity levels and AMA operational capacity grows, opportunities emerge for work-sharing arrangements in AI platform review. Chinese regulatory approvals, particularly for platforms already validated across diverse populations, could inform African assessments through reliance mechanisms that avoid duplicative reviews while preserving sovereign regulatory authority.

Coordinated approaches to data governance. Sino-African cooperation should address cross-border data governance challenges, potentially through model agreements or mutual recognition arrangements that facilitate responsible data flows while protecting privacy and sovereignty interests. The Malabo Convention provides a continental foundation that could be operationalized through bilateral or regional implementation protocols.

6.3. Industry Engagement and Responsible Innovation

Regulatory frameworks, however well-designed, require industry cooperation to achieve public health objectives. Chinese developers of AI-enabled drug discovery platforms should proactively engage with African regulatory expectations by:

Investing in African population validation. Demonstrating that AI platforms perform adequately in African populations—through prospective validation studies, real-world evidence generation, or collaborative research with African institutions—should be viewed as both regulatory prudence and commercial opportunity. Platforms validated in diverse African populations may enjoy competitive ad-

vantages in both African markets and global settings where demographic diversity is increasingly valued.

Supporting African research infrastructure. Technology transfer arrangements that build African capacity for AI in drug discovery—including computational infrastructure, training programs, and joint research initiatives—align with both Chinese industrial policy preferences for deep partnerships and African aspirations for technological sovereignty. The joint laboratory established between Hubei Jiangxia Laboratory and Tanzania's National Public Health Laboratory exemplifies this approach.

Transparency in intended use and limitations. Clear communication about platform capabilities, limitations, and appropriate use contexts is essential for responsible deployment and regulatory compliance. This includes honest representation of training data characteristics, known performance boundaries across populations, and appropriate human oversight requirements.

6.4. International Coordination and Standard-Setting

Finally, the global nature of both AI technology and pharmaceutical markets demands international coordination beyond bilateral Sino-African cooperation. African regulators, supported by Chinese and other partners, should actively participate in global standard-setting initiatives including:

IMDRF AI Working Group activities. The International Medical Device Regulators Forum has emerged as a key venue for developing harmonized approaches to AI/ML-enabled medical devices. African participation ensures that continental perspectives inform global guidance.

ICH reflection on AI in pharmaceutical development. As ICH considers guidance on AI applications in drug development, African and Chinese regulators should jointly advocate for attention to population diversity, data representativeness, and proportionate oversight approaches suitable for varied regulatory maturity levels.

WHO initiatives on AI in health. The World Health Organization's work on AI governance, including the Focus Group on AI for Health, provides multilateral platforms for developing normative guidance that can inform national and regional frameworks.

This area chart and line plot combination (Fig. 4) visualizes the temporal distribution of regulatory oversight across the drug development continuum. The key insight is the mismatch between existing regulatory frameworks' intensity profiles and the phases where AI platforms are most heavily utilized. Pharmaceutical and medical device regulations concentrate oversight in late-stage clinical development and regulatory submission, whereas AI platforms exert greatest influence in early discovery and preclinical optimization. The proposed AI-specific requirements (purple dashed line) address this

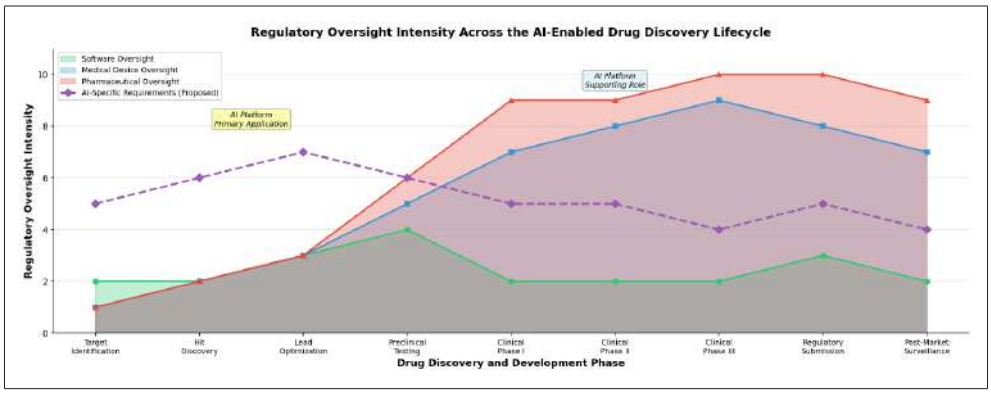

Figure 4. Intensity of regulatory oversight across the nine phases of AI-enabled drug discovery and development, comparing four regulatory frameworks: Software Oversight (green), Medical Device Oversight (blue), Pharmaceutical Oversight (red), and proposed AI-Specific Requirements (purple dashed). Oversight intensity is scored from 0-10 based on cumulative regulatory burden including documentation requirements, quality system obligations, clinical evidence expectations, and post-market surveillance mandates. Pharmaceutical oversight dominates late-stage development (Phases II-III, Regulatory Submission) with near-maximal intensity. Medical device oversight shows a bimodal distribution with peaks during preclinical validation and clinical development. Software oversight remains consistently low throughout. The proposed AI-Specific Requirements overlay addresses distinctive algorithmic concerns—training data quality, model validation, performance drift monitoring—with intensity concentrated in early discovery phases where AI platforms have greatest impact. Annotations indicate AI platform primary application domain (early discovery) and supporting role (clinical development and post-market) (Authors' own elaboration).

mismatch by front-loading oversight to early phases, ensuring that algorithmic quality and representativeness are established before substantial downstream investment. The figure supports the article's argument that novel regulatory approaches, rather than simple categorical assignment, are required to appropriately govern AI in drug discovery.

7. CONCLUSION

The question whether AI-enabled drug discovery platforms should be regulated as software, medical devices, or pharmaceuticals admits no single answer—nor should it. The appropriate classification depends on platform function, deployment context, and the specific risks posed to patient safety, data rights, and health equity. What emerges clearly from this analysis is that the classification choice carries profound implications extending far beyond legal taxonomy. It shapes market access pathways for Chinese exporters, technology transfer opportunities for African partners, and—most fundamentally—whether AI's transformative potential in accelerating drug discovery will reach populations historically marginalized by global pharmaceutical innovation. For Sino-African relations, the AI regulatory question presents both challenge and opportunity. The challenge lies in navigating divergent regulatory starting points: China's sophisticated but domestically-focused AI medical device framework, Africa's developing but still fragmented regulatory systems, and the absence of established cooperation mechanisms bridging the two. The opportunity lies in the possibility of regulatory leapfrogging—African regulators learning from both the achievements and limitations of Chinese, American, and European approaches to craft frameworks suited to African contexts and priorities. The institutional foundations for such cooperation are being laid. The African Medicines Agency's operationalization, growing numbers of WHO Maturity Level 3 NRAs, and explicit interest from both Chinese and African leadership in pharmaceutical regulatory cooperation create enabling conditions. The FDA-EMA Good AI Practice principles provide a global reference point that can be adapted rather than simply adopted. And the commercial imperatives driving Chinese pharmaceutical engagement in Africa create sustained incentives for addressing regulatory friction. Ultimately, the classification of AI-enabled drug discovery platforms should serve public health objectives: ensuring safety and effectiveness, promoting beneficial innovation, protecting data rights, and advancing health equity. Regulatory frameworks that fixate on categorical purity at the expense of these objectives fail their fundamental purpose. For African regulators, Chinese developers, and the patients who stand to benefit from AI-accelerated drug discovery, the task is to craft pragmatic, proportionate, and cooperative approaches that harness algorithmic innovation while safeguarding the humans it is meant to serve.

Ethics: There are no ethical issues with the publication of this manuscript.

Authors' Contributions: The authors contributed to the study equally.

Conflict of Interest: The author declares that he has no conflict of interest.

Financial Disclosure: The authors declared that this study has received no financial support.

Statement on the Use of Artificial Intelligence: Artificial intelligence was not used in the preparation of the article.

Peer-review: Externally peer-reviewed.

REFERENCES

Abdulwahab, A. A., Okafor, U. G., Adesuyi, D. S., Miranda, A. V., Yusuf, R. O., & Eliseo Lucero-Prisno, D. (2024). The African Medicines Agency and medicines regulation: Progress, challenges, and recommendations. *Health Care Science, 3*(5), 350–359. [CrossRef]

African Union. (2014). *African Union convention on cyber security and personal data protection.* African Union.

Aniche, E. T. (2023). African continental free trade area and african union agenda 2063: The roads to Addis Ababa and Kigali. *Journal of Contemporary African Studies, 41*(4), 377–392. [CrossRef]

Atacan, C., & Açık, A. (2023). Impact of geopolitical risk on international trade: Evidence from container throughputs. *Transactions on Maritime Science, 12*(2). [CrossRef]

Cerf, M. E. (2018). The Sustainable Development Goals: Contextualizing Africa's economic and health landscape. *Global Challenges, 2*(8), 1800014. [CrossRef]

Daneshjou, R., Smith, M. P., Sun, M. D., Rotemberg, V., & Zou, J. (2021). Lack of transparency and potential bias in artificial intelligence data sets and algorithms: A scoping review. *JAMA Dermatology, 157*(11), 1362–1369. [CrossRef]

de Villiers, K. (2021). Bridging the health inequality gap: An examination of South Africa's social innovation in health landscape. *Infectious Diseases of Poverty, 10,* 19. [CrossRef]

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 Accessed on Jun 08, 2026.

Ibáñez Antolín, A. (2021). *Applications of machine learning in drug discovery and development. Proceedings of MOL2NET'21.* Conference on Molecular, Biomedical & Computational Sciences and Engineering

(7th ed.). USE.DAT-07: USA-Europe Data Analysis Trends Congress, Cambridge, UK-Bilbao, Basque Country-Miami, USA, 2021. [CrossRef]

International Medical Device Regulators Working Group. (2022). *Machine learning-enabled medical devices: Key terms and definitions* [IMDRF/AIMD WG/N67FINAL:2022]. International Medical Device Regulators Forum.

Li, P., Williams, R., Gilbert, S., & Anderson, S. (2023). Regulating artificial intelligence and machine learning-enabled medical devices in Europe and the United Kingdom. *Law, Technology and Humans, 5*(2), 94–113. [CrossRef]

Lognoul, M. (2025). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act–AI Act). *Revue du Droit des Technologies de l'Information, 3–4*, 145–189.

Muehlematter, U. J., Daniore, P., & Vokinger, K. N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–2020): A comparative analysis. *The Lancet Digital Health, 3*(3), e195–e203. [CrossRef]

Ncube, C. B. (2015). *Intellectual property policy, law and administration in Africa: Exploring continental and sub-regional co-operation.* Routledge. [CrossRef]

Niyungeko, G. (Ed.) (2019). *Treaty for the establishment of the African Medicines Agency*. In African Yearbook of International Law Online / Annuaire Africain de Droit International Online. 24(1), 396–415. [CrossRef]

Nwauche, E. S. (2018). Intellectual property policy, law and administration in Africa: Exploring continental and sub-regional cooperation. *South African Intellectual Property Law Journal, 6*(1), 120–123.

Obunike, C. L., & Nwakob, C. B. (2025). The role of intellectual property in the African continental free trade area: Opportunities and challenges. *Chukwuemeka Odumegwu Ojukwu University Journal of Commercial and Property Law, 5*(1).

Rosen, K. L., & Mandl, K. D. (2025). Regulating flexibility for artificial intelligence: FDA experience with predetermined change control plans [Preprint]. medRxiv doi: 10.1101/2025.08.26.25334477 [CrossRef]

United Nations Conference on Trade and Development. (2026). Pharmaceutical trade flows between China and Africa: Trends, projections, and policy implications. United Nations Conference on Trade and Development.

Vieira, M., Andia, T., Karim, O., Srishti, S. A., Pineda, S. A., Alonso Ruiz, A., Large, K., Liu, Y., Moon, S., & Naher, N. (2023). Rising pharmaceutical innovation in the Global South: A landscape study. *Journal of Pharmaceutical Policy and Practice, 16*(1), 155. [CrossRef]

Zhang, L., & Yan, J. (2026). Approval of AI-based medical devices in China from 2020 to 2025: Retrospective analysis. *JMIR Medical Informatics, 14,* e85538. [CrossRef]

Yıldız Social Science Review
Web site information: https://yssr.yildiz.edu.tr
DOI: 10.51803/yssr.1884173

Original Article / Orijinal Makale

Institutional Syncretism in the Digital Age: The Adaptive Institutional Fusion (AIF) Framework

Dijital Çağda Kurumsal Sinkretizm: Uyarlanabilir Kurumsal Birleşme (AIF) Çerçevesi

Syed Shahnawaz MOHSIN

South Asian Institute for Policy and Governance (SIPG), Dhaka, Bangladesh

Güney Asya Politika ve Yönetişim Enstitüsü, Dakka, Bangladeş

ARTICLE INFO

Article history
Received: 07 February 2026
Accepted: 20 June 2026

Key words:
Adaptive institutional fusion, digital exposure, fusion nodes, hybrid governance, transitional societies

MAKALE BİLGİSİ

Makale Hakkında
Geliş tarihi: 07 Şubat 2026
Kabul tarihi: 20 Haziran 2026

Anahtar kelimeler:
Uyarlanabilir kurumsal kaynaşma, dijital görünürlük, kaynaşma düğümleri, hibrit yönetişim, geçiş toplumları

ABSTRACT

This paper introduces the adaptive institutional fusion (AIF) framework, a mid-range sociological theory explaining how formal and informal institutions interact under conditions of digital exposure. Research question: How do offline–online hybrid institutional arrangements form, adapt, and influence governance, education, labor, and social norms in transitional societies? Theoretical gap: The existing literature treats informal practices primarily as governance deficits, while digital sociology examines platform governance in isolation. Neither approach explains how hybrid systems evolve when locally embedded authority intersects with global digital visibility. AIF conceptualizes "fusion nodes" where formal regulations and informal practices converge, shaped by adaptive legitimacy, elasticity versus rigidity, and feedback loops generated by online exposure. Propositions: (1) Greater online visibility heightens legitimacy contests around hybrid institutions; (2) elastic hybrids with open digital interfaces formalize faster than rigid, patronage-bound systems; (3) digital exposure reinforces positive feedback in responsive institutions but amplifies dysfunction under elite capture; (4) hybrid nodes are durable, not transitional anomalies, and can be deliberately shaped by policy. Drawing on Bangladesh cases—madrasa digital curricula, shalish arbitration on Facebook, online labor markets, and activism campaigns—and global parallels such as M-Pesa, Brazilian favela leadership, and the gig economy, the paper argues that hybrid systems can deliver public value if recognized and governed rather than eradicated. The AIF framework offers a basis for rethinking institutional reform strategies and future research on algorithmic governance and cross-national validation.

Cite this article as: Mohsin S. (2026). Institutional Syncretism in the Digital Age: The Adaptive Institutional Fusion (AIF) Framework. *Yıldız Social Science Review, 12*(1), 55–72.

***Sorumlu yazar / Corresponding author**
*E-mail address: shahnawaz.mohsin@northsouth.edu

ÖZ

Bu makale, dijital görünürlük koşullarında biçimsel ve biçimsel olmayan kurumların nasıl etkileşime girdiğini açıklayan orta düzey bir sosyolojik kuram olan uyarlanabilir kurumsal kaynaşma (AIF) çerçevesini tanıtmaktadır. Araştırma sorusu şudur: Geçiş toplumlarında çevrim dışı–çevrim içi hibrit kurumsal düzenlemeler nasıl oluşur, uyum sağlar ve yönetişim, eğitim, emek ve toplumsal normları nasıl etkiler? Kuramsal boşluk: Mevcut literatür, biçimsel olmayan uygulamaları çoğunlukla yönetişim eksiklikleri olarak ele alırken, dijital sosyoloji platform yönetişimini ayrı olarak incelemektedir. Her iki yaklaşım da yerel olarak yerleşik otoritenin küresel dijital görünürlükle kesiştiği durumlarda hibrit sistemlerin nasıl evrildiğini açıklamamaktadır. AIF, biçimsel düzenlemeler ile biçimsel olmayan uygulamaların birleştiği "kaynaşma düğümleri"ni kavramsallaştırır; bunlar uyarlanabilir meşruiyet, esneklik ile katılık ve çevrim içi görünürlüğün oluşturduğu geri bildirim döngüleriyle şekillenir. Önermeler: (1) Daha fazla çevrim içi görünürlük, hibrit kurumlar etrafındaki meşruiyet mücadelelerini artırır; (2) açık dijital arayüzlere sahip esnek hibrit yapılar, katı ve patronaj temelli sistemlerden daha hızlı biçimselleşir; (3) dijital görünürlük, duyarlı kurumlarda olumlu geri bildirimi güçlendirirken, elit kontrolü altında işlev bozukluğunu artırır; (4) hibrit düğümler geçici anomaliler değil, kalıcı yapılardır ve politika yoluyla bilinçli olarak şekillendirilebilir. Bangladeş örneklerine —medreselerde dijital müfredatlar, Facebook'ta şaliş tahkimi, çevrim içi emek piyasaları ve aktivizm kampanyaları— ve M-Pesa, Brezilya favela liderliği ve gig ekonomisi gibi küresel paralelliklere dayanarak, makale hibrit sistemlerin ortadan kaldırılmak yerine tanınması ve yönetilmesi durumunda kamusal değer üretebileceğini savunmaktadır. AIF çerçevesi, kurumsal reform stratejilerini yeniden düşünmek ve algoritmik yönetişim ile ülkeler arası doğrulama üzerine gelecekteki araştırmalar için bir temel sunmaktadır.

Atıf için yazım şekli: Mohsin S. (2026). Institutional Syncretism in the Digital Age: The Adaptive Institutional Fusion (AIF) Framework. *Yıldız Social Science Review, 12*(1), 55–72.

1. INTRODUCTION

In transitional societies such as Bangladesh, there exists a dynamic interplay between formal institutions—including governmental bodies, bureaucratic frameworks, and codified legal systems—and informal networks, which encompass kinship ties, cultural norms, moral codes, and everyday patterns of social reciprocity (North, 1990; Minbaeva et al., 2023). Informal institutions function as vital extensions or intermediaries for formal structures when official mechanisms are incomplete, insufficiently enforced, or misaligned with lived realities (North, 1990; Minbaeva et al., 2023). These culturally embedded networks often operate through unwritten rules, forging trust-based connections that shape interpersonal and organizational behavior in contexts where formal channels fall short (Helmke & Levitsky, 2004; Ledeneva, 2006; Minbaeva et al., 2023). Informal networks are frequently the social mechanisms through which informal institutions are enacted and reproduced in everyday life ((Horak & Suseno, 2022; Minbaeva et al., 2023).

Over the past decade, Bangladesh has witnessed a rapid infusion of digital platforms—notably social media networks such as Facebook and WhatsApp—that serve as quasi-regulatory arenas, accelerating institutional interactions and modifying power dynamics (Castells, 2000; Suri, 2025). In crisis and activism contexts, for instance, Facebook has been used to mobilize public sentiment, reshape collective identity, and exert pressure on formal authorities—as seen in events such as the Monsoon Uprising of July 2024 (Abir, Chowdhury, & Rahman, 2025). These online spaces frequently blur the divide between formal policy channels and informal social discourse, enabling community members to sanction behavior, share grievances, or organize collectively—sometimes eliciting rapid official responses.

However, the existing literature tends to approach informal institutional practices as deficits—viewing them primarily through a lens of dysfunction or corruption—and to attribute resilience in informal modalities to state failure rather than to adaptation (Nason, 2023; Giménez-Jiménez et al., 2020). Meanwhile, digital sociology interrogates the effects of technology on social identity, power structures, and collective behavior in increasingly networked societies, but typically lacks a holistic account of how informal and formal institutions fuse across both offline and online arenas. Existing scholarship has focused primarily on platform governance, networked communication, digital activism, and platform effects rather than on the interaction between digital systems and legacy institutional arrangements (van Dijck et al., 2025; van Dijck et al., 2021).

This leads us to the central research question of this study:

How do hybrid institutional arrangements—those that arise from the fusion of formal and informal systems, both

offline and online—form, evolve, and influence governance, education, and social norms in transitional societies such as Bangladesh?

To address this, the paper introduces the Adaptive Institutional Fusion (AIF) Framework, a middle-range theory that foregrounds:

- the genesis of *fusion nodes*—sites where formal and informal modalities intersect,
- the role of *digital exposure* in shaping adaptive legitimacy,
- the elasticity or rigidity of hybrid arrangements in response to change,
- and how *feedback loops*, amplified by digital virality, can either propel institutional reform or fortify entrenched dysfunction.

To ground the theoretical development, this study adopts a qualitative comparative approach that synthesizes evidence from Bangladesh alongside selected global parallels. The main text focuses on conceptual implications, while detailed case narratives, stakeholder accounts, and coding schemas are presented in an annex. This design ensures that empirical illustrations inform the AIF framework without constraining it to a single country or sector.

2. LITERATURE REVIEW AND GAP

2.1. Formal vs. Informal Institutional Theories

The dichotomy between formal and informal institutions lies at the heart of institutional sociology and political economy. Douglass North (1990) emphasized that institutions consist of both codified rules (formal) and routines, norms, and shared understandings (informal), all of which systematically shape economic and social behavior, including the persistence of informal mechanisms in contexts where formal rules are absent or inadequately enforced.

Building on this foundation, Helmke and Levitsky (2006) developed a typology of interactions between formal and informal institutions: complementary, accommodating, competing, and substitutive. For example, informal practices may complement formal rules by filling gaps left by enforcement weaknesses or enhancing bureaucratic coordination; alternatively, they can substitute defective formal structures altogether (Helmke & Levitsky, 2004). They clarify that informal institutions should not be simplistically equated to institutional weakness or dysfunction—they can also reflect adaptive, contextually embedded mechanisms that maintain social order, especially in transitional settings (Batlle, 2008).

Postcolonial and development sociology critiques this binary, arguing that formal and informal practices are often deeply intertwined in postcolonial contexts and cannot be understood through rigid institutional distinctions (Scott, 1998; Bayart, 1993; Chatterjee, 2004). It notes that labeling informal practices as merely "corruption" flattens their sociocultural role and obscures their strategic and adaptive value (e.g., in patronage networks or solidaristic community coping strategies). Informal arrangements—kinship-mediated access to services, non-legal dispute-resolution mechanisms, or locally enforced norms—can serve as pragmatic alternatives when formal structures fail. However, such practices are frequently dismissed as evidence of institutional weakness or governance failure rather than evaluated as adaptive mechanisms that enable coordination and survival in weakly institutionalized environments (Hart, 1973; Roy, 2005; Chatterjee, 2004).

A growing body of research has highlighted the importance of informal institutions, relational governance, and digitally mediated forms of coordination in shaping contemporary organizational and societal outcomes (Horak & Yang, 2016; Ledeneva, 2006; Minbaeva et al., 2023). Therefore, it seeks to reframe the informal-formal interface not as a pathology but as an adaptive architecture—particularly in developing nations where formal institutions are nascent or under-resourced. However, many of these accounts remain primarily descriptive, focusing on documenting platform practices, digital participation, and online network formation without developing broader institutional frameworks that explain how formal and informal institutions interact and co-evolve (van Dijck et al., 2021). Thus, lacking the structural theorization needed for generalizable analysis across contexts.

While North (1990) and Helmke & Levitsky (2004) distinguish formal from informal institutions, neither addresses how digital platforms are reshaping these dynamics in transitional societies. This omission leaves unclear how informal practices evolve when they gain visibility and public validation through online media.

2.2. Digital Sociology and Institutional Change

The digital transformation of social life has generated a robust field of digital sociology, focused on how platform architectures, social media, and networked communication reshape power, identity, and collective action. Manuel Castells (2000) introduced the concept of the "network society," where societal organization is fundamentally structured through interconnected digital flows linking individuals, institutions, and networks (Castells, 2000). Meanwhile, José van Dijck (2024; 2025) and colleagues have examined the cultural and governance implications of platformization. They explore how digital platforms—Facebook, YouTube, Wikipedia—have become embedded sites for public discourse, governance, and identity formation (van Dijck, 2024; van Dijck et al., 2025).

Van Dijck and co-authors emphasize that platforms both reflect and shape public values, norms, and forms of

authority. The governance structures embedded within platforms—moderation algorithms, networked data aggregation, content visibility—affect democratic exchange, institutional transparency, and social mobilization (van Dijck et al., 2025). However, this literature essentially treats digital spaces in isolation—examining how individuals interact with platforms—but rarely interrogates how these digital interactions fuse with preexisting formal and informal institutions.

Similarly, research on digital activism, gig economies, and platform-mediated service delivery has generated substantial insights into online mobilization, digital labor relations, platform governance, and digitally mediated service provision (Gerbaudo, 2012; Srnicek, 2017; Dijck et al, 2018). Much of this work focuses on how individuals use digital tools to organize protests, find informal work, or challenge authority—but often stops short of analyzing how these practices redefine institutional legitimacy or how they interact with state infrastructure.

For instance, discussions on deplatformization reveal how platform gatekeepers can shape public discourse and marginalize certain actors (van Dijck, de Winkel, & Schäfer, 2021). However, while this points to the power of digital platforms, it does not explicitly integrate how these digitally mediated powers fuse with informal networks or formal governance to create institutional hybrids.

Hence, digital sociology richly unpacks the architecture and social effects of platforms but does not systematically theorize the fusion of digital, informal, and formal institutional spheres—leaving a key analytical gap, especially in transitional societies where digital exposure both transforms and relies on local, informal mechanisms.

Similarly, Castells (2000) and Dijck et al. (2018) examine digital networks and platform governance but do not analyze how these interact with legacy institutions. The literature on digital activism, gig economies, and platform governance remains disconnected from studies of institutional adaptation and resilience.

2.3. Hybrid Governance and Institutional Resilience

The concept of "hybrid governance" or "hybrid orders" has grown prominent in fields analyzing governance in fragile or transitional contexts. Boege et al. (2008) describe how authority in post-conflict or weak-state environments often emerges from amalgamated repertoires—blending state symbols, customary law, charismatic leadership, and community norms—that provide legitimacy and social stability (Wenner, 2020). Scholars such as Lund (2006) conceptualize these as "twilight institutions," existing in the liminal space between formal and informal authority.

However, critiques from Meagher (2012) and others caution against overly romanticizing the embeddedness of hybrid orders. They underscore that such systems can be violent, contradictory in their legitimacy, and prone to legitimacy erosion when integrated into formal systems (Meagher, 2012). Furthermore, governance studies often emphasize hybrid legitimacy but rarely interrogate the role of online exposure or digital amplification in shaping these hybrid systems.

Moreover, definitions of hybrid institutional forms at the global governance level—such as Abbott and Faude's concept of "hybrid institutional complexes"—portray contemporary governance as an overlay of formal interstate institutions, private transnational bodies, and informal networks, yet usually at the macro scale and lacking attention to digitally mediated processes (German & Keeler, 2010; Abbott & Faude, 2021). These frameworks, while illuminating the complexity of governance beyond the state, still tend to abstract away from the granular mechanisms by which hybrid authority becomes visible, contested, or consolidated via local and digital practices.

Thus, the existing literature on hybrid governance recognizes the amalgamation of diverse sources of legitimacy and authority, especially in contexts with weak formal institutions. However, they rarely incorporate digital exposure—how online visibility, viral content, and social media discourse interact with and reshape hybrid institutional configurations. This absence leaves a critical theoretical lacuna.

Taken together, these strands of research fail to explain how hybrid institutional arrangements form, adapt, and persist under conditions of digital exposure. This dual gap—between institutional theory and digital sociology—justifies the need for an integrative framework such as AIF.

2.4. Identified Gap

Synthesizing the above literature reveals a clear and pressing theoretical gap. While North (1990), Helmke & Levitsky (2006), and others have charted the interplay between formal and informal institutions, most accounts reduce informal practices to corruption or dysfunction, missing their potential as adaptive mechanisms in weak-state contexts. Meanwhile, digital sociology offers nuanced insights into platform effects, networked communication, and digital activism—but often treats digital systems as separate from traditional institutional forms. Similarly, hybrid governance theories surface how informal and formal authorities coalesce in transitional spaces—but seldom theorize digital exposure as a constitutive element of these arrangements.

What is missing is an integrated framework that:

1. Explains how formal–informal institutional fusion occurs not only in physical space but also through digital interactions;
2. Accounts for how digital exposure affects the legitimacy of hybrid systems—either reinforcing or delegiti-

mizing them;

3. Differentiates between elastic hybrid nodes (capable of integrating reform) and rigid ones (that ossify into dysfunction);
4. Traces how feedback loops—accelerated by digital virality—can either catalyze institutional reform or entrench entrenched clientelism.

This gap is particularly salient in Bangladesh, where blended offline-online governance, education, economic coordination, and activism are already emerging—but remain insufficiently theorized. Bridging this gap requires a novel, middle-range theory that combines institutional sociology, digital sociology, and hybrid governance into a coherent analytical architecture—the AIF Framework.

3. THE ADAPTIVE INSTITUTIONAL FUSION (AIF) FRAMEWORK

3.1. Core Concepts

The AIF Framework centers on how formal and informal institutions merge—both offline and online—to produce hybrid social orders. It introduces four core concepts:

Fusion Nodes

Fusion nodes are points where formal regulations and informal practices converge, creating hybrid institutional spaces. For instance, in rural Bangladesh, traditional dispute resolution systems like the **shalish** (an informal village arbitration mechanism) persist even after the introduction of formal **village courts**—underscoring how informal justice persists alongside, and often mediates through, formal structures (Mattsson & Mobarak, 2023; Islam, 2018). When such settlements are mediated via **WhatsApp groups**, where parties submit grievances or communicate judgments digitally, the convergence of traditional authority and digital communication forms a new hybrid node. This melding of offline and online modalities exemplifies how institutional logics can blend to create emergent, adaptive governance mechanisms.

This concept resonates with institutional logic theory (ILA), which holds that multiple logics—such as community norms, state law, and digital ethos—coexist and interact (Friedland & Alford, 1991; Thornton & Ocasio, 2008). Fusion nodes represent situated sites where actors navigate overlapping logics under specific cultural and political conditions.

Adaptive Legitimacy

Adaptive legitimacy refers to the socially constructed acceptance of these hybrid arrangements. It builds on Suchman's (1995) typology of legitimacy: *pragmatic* (based on self-interest), *moral* (normative approval), and *cognitive* (taken-for-grantedness). In the AIF context, digital exposure can reshape all three dimensions, as when viral videos of citizen action online either reinforce or challenge legitimacy.

Digital platforms amplify institutional visibility, making hybrid practices emergent or problematic—visible to broader audiences and open to rapid legitimization or de-legitimization (Gümüsay et al., 2022). This dynamic transforms legitimacy from a slow, context-bound process into a highly reactive, contested terrain.

Elasticity vs. Rigidity

Hybrid institutional nodes exhibit varying capacities to adapt. Elastic fusion nodes stretch—incorporating reforms, responding to critique—and can evolve into formalized structures (e.g., informal online marketplaces migrating toward registered SME status). Conversely, rigid nodes ossify—anchored in clientelism or elite capture—remaining impervious to transformation. Digital exposure plays a dual role: it can propel nodes toward elasticity when institutions are responsive, yet, when dominated by entrenched elites, it reinforces rigidity by crowding out dissenting voices through curated narratives or platform manipulation (Ahmed, 2025b).

Feedback Loops

Hybrid nodes operate within complex feedback loops:

- Positive loops emerge when digital visibility triggers institutional correction—for example, when viral social media images of police misconduct (e.g., a viral image of a policeman suppressing a student's voice) prompt rapid institutional response or denial, exposing frictions in state legitimacy (TBS, 2025).
- Negative loops unfold when online exposure reinforces dysfunction—for instance, digital spaces may be weaponized to conduct media trials, mobilize mob violence, or suppress critical voices, thereby deepening existing dysfunctions instead of resolving them (Ahmed, 2025a).

These loops are accelerated in digital environments, where visibility and virality shorten the time between action and response (Gümüsay et al., 2022). Feedback loops are therefore central to understanding whether hybrid systems evolve adaptively or regressively.

3.2. Visual Model *(Conceptual Diagram)*

Explanation:

- **Offline Layer**: Traditional formal institutions (government bodies, courts) and informal networks (kinship, customary norms).
- **Online Layer**: Digital platforms—social media, messaging apps, digital marketplaces.

- **Fusion Nodes**: Intersection points where offline and online institutional logics merge.
- **Outcomes**: Depending on institutional responsiveness and digital dynamics, outcomes bifurcate into *Resilient* (adaptive, formalizing) or *Dysfunctional* (ossified, corrupt).
- **Feedback Loops**: Arrows from outcomes back into layers mediated via digital visibility—public scrutiny accelerates adaptation or reinforces dysfunction.

This model visually underscores the dynamic interplay between offline/online domains, institutional fusion, and the digital acceleration of change or entrenchment (Fig. 1).

3.3. Propositions (Theory Building)

Grounded in the concepts above, the AIF framework advances four core propositions:

- **P1: Online Visibility and Legitimacy Contestation**

  The higher the online visibility of hybrid arrangements, the greater the contestation over their legitimacy. Digital platforms magnify hybrid practices, exposing them to broader critique or affirmation, thereby intensifying legitimacy contests.

- **P2: Elastic Digital Interfaces and Formalization Speed**

  Hybrid systems with elastic digital interfaces (e.g., open marketplaces) formalize faster than rigid, patronage-based ones. Digital openness enables transparency, regulatory engagement, and proactive formalization; closed, patronage networks resist such evolution.

- **P3: Digital Exposure Differentially Amplifies Feedback**

  Digital exposure enhances positive feedback in responsive institutions (promoting reform) but amplifies negative feedback in elite-captured systems. Institutional capacity and accountability mediate whether digital exposure leads to inefficiency or improvement.

- **P4: Hybrid Nodes as Enduring and Policy-Shapeable**

  Hybrid nodes are not fleeting transitional anomalies but enduring social arrangements that can be proactively shaped by policy. Rather than aiming to abolish informal-formal overlaps, policy can harness fusion nodes for inclusion and reform.

4. METHODOLOGICAL STRATEGY

This study employs a qualitative comparative approach, combined with conceptual theory-building, to explore how hybrid institutional arrangements—blending offline informal practices with online digital platforms—emerge, adapt, and shape governance, education, labor markets, and social norms in transitional societies such as Bangladesh. By situating empirical cases within a structured comparative framework, the research both refines and tests the AIF framework as a middle-range theory.

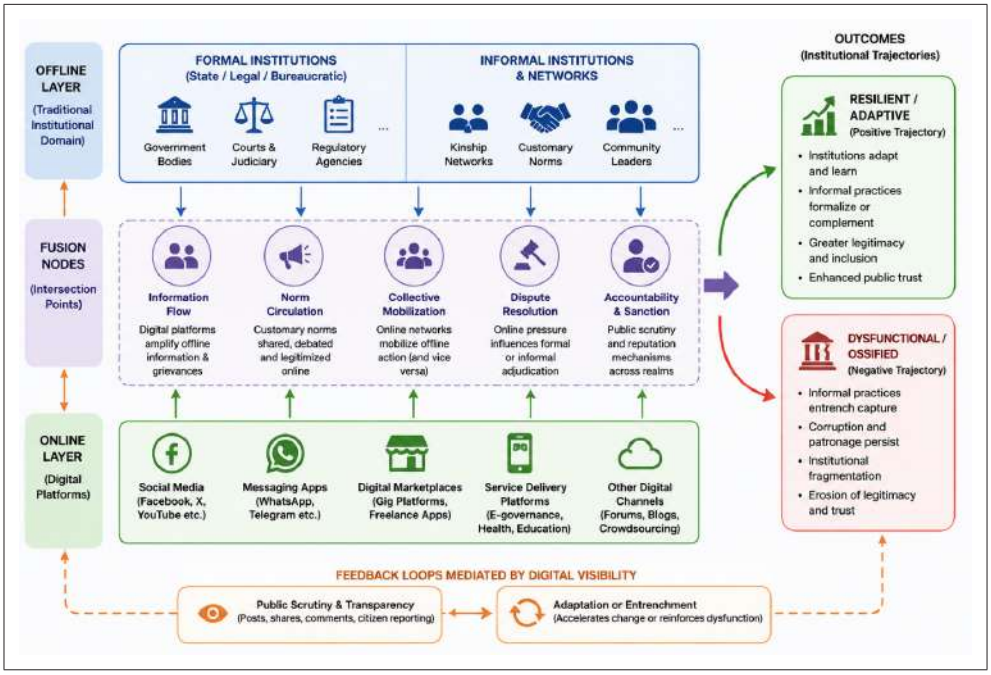

Figure 1. EAdaptive institutional fusion framework showing how formal and informal institutions interact with digital platforms through fusion nodes, producing either resilient/adaptive or dysfunctional/ossified institutional trajectories.
Source: Author own compilation

4.1. Research Design

This study employs a comparative case analysis approach to examine how AIF operates across diverse institutional contexts. The selected cases are used to identify recurring mechanisms, patterns of interaction between formal and informal institutions, and the role of digital platforms in shaping institutional adaptation. The objective is theory development and illustrative comparison rather than formal causal inference through set-theoretic methods. This approach is suited to analyzing configurational causation, in which different combinations of factors yield similar institutional outcomes. Bangladesh serves as the primary case cluster, while selected global parallels—including Kenya, Brazil, and the United States—function as contrastive cases to probe the broader applicability of findings (George & Bennett, 2005).

Alongside comparative analysis, the research employs conceptual theory-building to refine the AIF framework. As noted by Swedberg (2014), combining empirical observation with abductive reasoning helps generate new theoretical insights while avoiding premature closure around existing categories.

4.2. Data Sources

The empirical foundation of this study is built upon a triangulated research design that integrates four complementary data streams to ensure robustness and validity (Denzin, 2012). First, it draws on extensive archival research, examining government records, NGO reports, and media coverage to document formal institutional reforms and the parallel existence of informal practices (Yin, 2017). Second, first-hand narratives are captured through semi-structured interviews with key stakeholders, including policymakers, educators, community leaders, and digital platform users in Bangladesh. Third, the study incorporates a digital dimension through systematic social media content analysis of Facebook, YouTube, and WhatsApp networks, tracing how informal governance practices diffuse into online communicative spaces (Marwick, 2013). Finally, comparative insights are derived from secondary case literature on hybrid institutional phenomena in other global contexts (Mahoney & Thelen, 2009).

4.3. Sampling Strategy

Sampling followed a purposive logic (Patton, 2015). The Bangladesh cases were chosen because they demonstrate clear examples of offline-online institutional fusion:

1. Education: Madrasa and formal schooling integrated through YouTube curricula.
2. Governance: Village shalish decisions gaining visibility on Facebook and influencing formal court behavior.
3. Labor markets: Informal job brokerage shifting to WhatsApp and Facebook, circumventing state boards.
4. Norm negotiation: Online anti-harassment movements pressuring police to act.

Global parallels—such as Kenya's M-Pesa mobile money system, Brazilian favela leaders' Instagram mediation, and regulatory gray zones in U.S./European gig economies—were selected as contrastive samples to test whether similar fusion mechanisms occur across contexts with different institutional capacities.

4.4. Coding and Analytical Framework

The data were analyzed using thematic coding, both inductively and deductively (Lungu, 2022). Coding focused on:

1. Identifying fusion nodes: where offline practices intersect with online platforms.
2. Assessing adaptive legitimacy: by analyzing whether hybrid arrangements gain public trust, using narratives from interviews and social media discourse.
3. Classifying elasticity vs. rigidity: measuring whether these arrangements adapt flexibly to institutional pressures or reinforce entrenched dysfunction.

This process generated structured case summaries that informed the conceptual refinement of the AIF framework, consistent with guidelines for mid-range theory development (Merton, 1968).

4.5. Ethical Considerations

Conducting digital ethnography raises specific risks concerning privacy, surveillance, and participant consent (Markham & Buchanan, 2012). To mitigate these risks, all personal identifiers were removed from interview transcripts and social media data. Publicly available online content was analyzed in aggregate, and any quotes used in the study were paraphrased unless explicit permission was granted. These steps align with the Association of Internet Researchers (AoIR) recommendations for ethical research in digital contexts.

4.6. Annex Placement

To maintain theoretical coherence in the main text, empirical details—including full case descriptions, data tables, and coding schemas—are presented in the annex. This structure enables the main paper to emphasize conceptual insights, while allowing complete methodological transparency for scholarly review (Miles et al., 2014).

5. CASE IMPLICATIONS (CASE STUDIES IN ANNEX)

This section synthesizes the **implications** of the emerging hybrid institutional modalities identified in Bangladesh and globally. Detailed empirical case studies—featuring narrative depth, timelines, stakeholder perspectives, and

digital interaction data—are housed in the Annex. The main text emphasizes cross-cutting insights that inform institutional theory and policy design. To maintain compliance with journal word limits, detailed case descriptions, supplementary contextual information, and supporting materials are provided in the Annex. The case discussions in the main text present only the evidence most directly relevant to illustrating the AIF framework, while the Annex provides additional documentation and contextual detail for interested readers.

5.1. Bangladesh Cases

The cases are intended as illustrative applications of the AIF framework rather than exhaustive empirical studies. Additional contextual information, supporting observations, and source materials for each case are presented in the Annex.

5.1.1. Education

Madrasas represent one of the largest parallel educational systems in Bangladesh. The sector is broadly divided into Alia madrasas, which are regulated by the Bangladesh Madrasa Education Board and combine religious instruction with general education subjects, and Qawmi madrasas, which operate independently and focus primarily on Islamic scholarship and Arabic studies (Roy et al., 2020). The madrasa system serves several million students nationwide and has expanded significantly since independence, particularly in rural and economically disadvantaged areas (Riaz, 2008). While Alia madrasas provide pathways into higher education and public-sector employment, Qawmi institutions play an important role in producing religious scholars and community leaders (Asadullah & Chaudhury, 2009). Consequently, madrasas function not only as educational institutions but also as important social and cultural organizations that shape local authority structures, community norms, and patterns of social interaction in Bangladesh (World Bank, 2016).

Madrasas in Bangladesh have historically lagged in offering modern subjects due to infrastructure shortages and a scarcity of ICT-equipped classrooms (e.g., only 25% of Aliya madrasas have computer labs) (Rohit, 2025). However, platforms like **Madrasa** and **10 Minute School** have created digital curricula, disseminating thousands of videos via YouTube, WhatsApp, and websites, dramatically increasing educational reach and blending religious and formal instruction (Zaywa, 2021; Islam & Sakib, 2024). This convergence illustrates a **fusion node**: the offline madrasa and formal education system meld with online platforms to deliver curricula in new digital modalities (Shumshunnahar et al., 2025).

The madrasa–digital platform convergence exemplifies more than an incremental improvement in curriculum delivery—it marks a structural shift in how hybrid institutions gain legitimacy. When informal or semi-formal education providers (such as Qawmi and Aliya madrasas) engage with widely accessible digital platforms, they not only enhance reach but also signal professionalization to both state regulators and social constituencies. The infusion of structured curricula, modern pedagogical content, and standardized assessment materials on platforms like Madrasa and 10 Minute School demonstrates that religious education can evolve beyond rote instruction to include market-relevant knowledge (Zaywa, 2021; Islam & Sakib, 2024).

Digital presence works as a legitimacy amplifier. Online content archives create visible, auditable evidence of quality improvement—something that bureaucratic regulators, donors, and even skeptical parents can observe. The increased visibility afforded by digital platforms has the potential to make aspects of religious schooling more transparent and accessible to external audiences, allowing educational activities, institutional performance, and student achievements to become more visible and potentially comparable across institutions. Digital visibility may enhance institutional legitimacy where digital adoption is sufficiently developed and accepted by relevant stakeholders. However, existing studies also document significant technological, cultural, and ideological barriers to ICT diffusion within parts of the madrasa sector (Shumshunnahar et al., 2025), suggesting that such effects remain uneven and contingent. Rather than assuming a uniform process of digital transformation, the present case illustrates an emerging and contested process in which some madrasas adopt digital tools to enhance visibility and stakeholder engagement, while others remain constrained by resource limitations, organizational capacity, or resistance to technological change.

Elasticity is achieved through cultural compatibility. Because digital curricula are embedded in familiar religious frameworks, the hybrid institution remains culturally legitimate while absorbing modern STEM and language instruction. This elasticity—the ability to stretch institutional form without breaking community trust—illustrates why hybrid systems persist rather than dissolve into purely secular or purely religious categories. Instead of treating hybrid education as a "transition stage" toward fully formal systems, this case suggests that such models stabilize into durable equilibria that adapt to both policy constraints and community expectations (Rohit, 2025).

It has potential to reconfigure state policy. If regulators recognize the credibility of these digital–religious networks, policy frameworks could shift from binary recognition (formal vs. informal) to tiered accreditation, where hybrid programs receive partial certification tied to content standards rather than institutional origin. This could encourage state–civil society co-production of curricula, reduce the stigma attached to religious schooling, and help governments meet Sustainable Development Goal 4, which seeks to ensure inclusive, equitable, and quality education

and promote lifelong learning opportunities for all (United Nations, 2015), while achieving these targets without building massive new infrastructure (Islam & Sakib, 2024).

In short, digital exposure may create conditions under which hybrid institutions become more visible to policymakers, communities, and external stakeholders, potentially facilitating their recognition as adaptive educational actors and increasing opportunities for integration into broader educational planning processes (Shumshunnahar et al., 2025; Dijck et al., 2018). Rather than fading as formal systems expand, these elastic nodes may become templates for inclusive governance in education—balancing cultural legitimacy with developmental imperatives.

Digital exposure can legitimize hybrid educational arrangements that fill institutional voids—especially when formal systems are constrained—by providing culturally compatible, accessible learning. Such hybrid models hold potential for formal recognition if policy frameworks adapt to integrate them.

The madrasa sector in Bangladesh is neither institutionally uniform nor uniformly digitalized. It encompasses state-regulated Alia madrasas and largely autonomous Qawmi madrasas, each characterized by distinct governance structures, curricular orientations, and relationships with state institutions (Riaz, 2008; Asadullah & Chaudhury, 2009). Existing research further suggests that the adoption of information and communication technologies (ICTs) remains uneven across the sector, shaped by disparities in resources, infrastructure, institutional priorities, and attitudes toward technological change. While some institutions have incorporated digital platforms for administration, communication, and educational activities, others continue to face technological constraints and cultural or ideological reservations regarding ICT adoption (World Bank, 2016; Shumshunnahar et al., 2025). Consequently, the present case should not be interpreted as representative of the entire madrasa sector but rather as an illustration of how digital engagement may emerge within specific institutional contexts.

The madrasa case can be interpreted as an emerging fusion node in which digital tools create opportunities for greater visibility, stakeholder engagement, and pedagogical innovation. While the evidence does not establish that digitalization has produced widespread formal legitimacy or institutional integration, it illustrates how online platforms may facilitate processes through which hybrid educational institutions seek recognition and broader societal engagement. The case therefore serves primarily as an illustrative example of the AIF framework, demonstrating how digital technologies may influence interactions between formal and informal educational institutions while also highlighting the constraints, uneven adoption patterns, and resistance documented in existing studies of madrasa digitalization.

5.1.2. Governance

Informal justice via *shalish* (village arbitration) persists even with the presence of formal village courts. While formal courts may offer more structured justice, *shalish* remains faster, more accessible, and contextually trusted—especially for the disadvantaged (e.g., poor villagers) (Alim & Ali, 2023; Sutradhar, 2023). When *shalish* outcomes are shared on Facebook or WhatsApp, they enter public and official discourse more quickly, effectively influencing state court behavior through digital scrutiny or symbolic pressure. This mediated convergence forms a potent **fusion node** in rural governance.

The digital mediation of *shalish* outcomes magnifies the permeability between informal and formal governance mechanisms. When arbitral decisions are uploaded to Facebook or WhatsApp, they transition from being purely local, discretionary practices into visible events subject to broader social commentary, NGO monitoring, and even judicial attention. This online exposure can accelerate institutional learning within the formal justice system by highlighting community demands for speed, cultural relevance, and procedural fairness. From an AIF perspective, digital scrutiny may function as a mechanism through which local dispute-resolution institutions become more publicly visible and accountable. Under favorable institutional conditions, such visibility could encourage reforms that more closely align judicial practices with community expectations, although the present case does not provide direct evidence of these outcomes (Fung, 2013; Fritzen, 2016).

However, this same visibility also risks reinforcing entrenched power asymmetries when elite actors dominate *shalish* proceedings. Public validation through likes, shares, or supportive comments may cloak clientelistic or coercive rulings in a veneer of legitimacy, discouraging state oversight. In such cases, the hybrid governance node becomes rigid rather than adaptive, allowing informal authority to harden rather than integrate with formal accountability systems.

This duality underscores a critical policy lesson - digital transparency is not inherently democratizing. For convergence to produce institutional synergy rather than distortion, policymakers must pair digital openness with targeted investments in village court capacity, community legal awareness, and mechanisms for independent review. When these safeguards are present, *shalish*–court interaction can evolve into an elastic fusion node, where informal authority is disciplined and formal authority becomes more responsive. Absent these safeguards, digital exposure risks becoming a tool for performative justice rather than substantive reform.

Publicizing *shalish* decisions online can exert dual influence—forcing formal courts to improve responsiveness while also risking informal validation beyond state oversight. Thus, digital visibility may spur reforms if institution-

al capacity exists, or entrench informal authority if formal institutions remain unresponsive. This case demonstrates a tension between rigidity and elasticity. When Facebook visibility prompts procedural improvements in village courts, a positive feedback loop emerges. Conversely, when elite-controlled shalish decisions are broadcast without accountability checks, the hybrid node solidifies clientelistic authority.

5.1.3. Labor Markets

Bangladesh's informal labor sector is extensive, with nearly 85% of the workforce engaged outside formal employment and facing significant wage penalties compared to their formal-sector counterparts (Dallakoti, 2024; Rahman et al., 2019). In parallel, an expanding digital freelance economy—comprising more than 650,000 freelancers generating nearly USD 1 billion annually—has integrated informal labor into the global marketplace via platforms such as Upwork and Fiverr (Hasan, 2025). Additionally, job brokerage and recruitment activities increasingly occur through informal channels like WhatsApp and Facebook, bypassing state-regulated labor boards and oversight mechanisms.

These digitally mediated labor practices exemplify the elasticity of hybrid labor markets. Digital exposure enables informal work to achieve global visibility and partial legitimacy; however, equitable benefits depend on the existence of formal regulatory and worker protection frameworks. In the absence of such safeguards, these digital brokerage networks function as "fusion nodes" that are simultaneously innovative and fragile. While they expand labor market inclusion, they also heighten vulnerability to fraud, wage suppression, and other exploitative practices. This dual dynamic underscores the tension between flexibility and institutional accountability in Bangladesh's evolving labor governance.

5.1.4. Social Norms

Online activism campaign against harassment has compelled police and political institutions to respond to previously neglected public grievances. These movements exemplify hybrid nodes that combine social norms with formal law enforcement through digital exposure (see Annex for details). When institutional responsiveness aligns with online pressure, digital visibility becomes a tool of reform. These movements illustrate hybrid governance nodes, where digital platforms intersect with formal law enforcement and political accountability structures. When institutional responsiveness aligns with online mobilization, digital visibility becomes a catalytic mechanism for reform, validating public protest as a legitimate input into governance processes (Phillips, 2024).

Implication: Digital visibility can legitimize civil resistance and protest norms, incentivizing formal institutions to act. Hybrid nodes here reinforce institutional elasticity when state actors are accountable—but if exposure triggers repression, they risk forgiveness being reversed, leading to rigidity and enforcement crackdowns. Digital visibility can legitimize civil resistance and embed social norms of accountability into formal institutions. These hybrid nodes strengthen institutional elasticity when state actors respond constructively, but they remain fragile. If heightened exposure instead provokes repression, protest norms risk being delegitimized and replaced by punitive crackdowns, converting elasticity into rigidity.

5.2. Global Parallels

5.2.1. M-Pesa (Kenya)

The mobile money platform M-Pesa merged formal banking regulations with long-standing trust-based informal savings networks in rural Kenya. Its digital reach extended economic inclusion through a hybrid institutional model that combined formal financial frameworks with informal trust—a fusion that drew on adaptive legitimacy in underbanked areas. By integrating regulatory oversight with community-based trust, digital finance can achieve institutional elasticity, broadening access while maintaining credibility. However, sustaining this fusion requires vigilance against fraud and regulatory overreach that could undermine local legitimacy.

5.2.2. Favela Leadership (Brazil)

In Brazil, favela leaders have used Instagram to negotiate with municipal authorities, blending informal community governance autonomy with visible online engagement and formal political interaction. These hybrid nodes function as mediation points—with digital visibility enhancing legitimacy and negotiating power. Digital platforms can empower informal governance actors to gain formal recognition and bargaining capacity. However, if municipal authorities shift from cooperation to suppression, these nodes risk losing adaptive legitimacy and reverting to rigid confrontation.

5.2.3. Gig Economy (U.S. & Europe)

Platforms like Uber operate in regulatory gray zones, mixing informal gig labor practices with formal economic regulations. The digital interface, though ostensibly formal, accommodates informal work patterns—creating elastic hybrid systems that resist complete formalization while delivering economic function. Hybrid gig platforms show how partial regulation can maintain labor market flexibility and economic efficiency. However, without robust protections, these systems may entrench precarity, demonstrating that elasticity without safeguards can harden into institutional fragility.

Implication (Global): These examples suggest that fusion-node dynamics can be observed across diverse insti-

tutional settings. When digital engagement complements existing formal and informal mechanisms, it may contribute to greater institutional adaptability and resilience; conversely, where formal systems remain weak or poorly aligned with digital processes, dysfunction may persist. Although the limited number of cases does not permit claims of universality, the observed parallels indicate that the AIF framework may offer a useful analytical lens for examining interactions between digital, formal, and informal institutions across different contexts.

5.3. Key Implications Synthesized

This study identifies three key dynamics governing the relationship between digital visibility and hybrid institutional arrangements. First, legitimation occurs through public value delivery: when hybrid, digitally exposed nodes generate tangible public benefits—such as improved access to education, cash flow, or justice—their digital visibility can confer legitimacy, effectively transforming informal options into recognized alternatives. Second, digital exposure carries a dual-edged potential for delegitimizing patronage networks by opening new accountability channels; however, this dynamic only yields meaningful reform if formal institutions are both equipped and willing to harness the resulting pressure. In the absence of such institutional readiness, exposure risks amplifying dysfunction rather than curbing it. Third, the findings point toward a strategy of policy integration over eradication: hybrid institutional arrangements should not be eliminated but rather incorporated into formal frameworks. Policies that co-design with existing mash-ups—for instance, by recognizing digital madrasa credentials or facilitating freelance regulation—can leverage these fusion nodes to advance more inclusive and responsive governance.

6. DISCUSSION

6.1. Revisiting the Propositions

Proposition 1 (P1): *Higher online visibility intensifies legitimacy contention around hybrid institutional arrangements.*

The Bangladesh cases support this: for example, when shalish decisions were shared on Facebook, formal courts faced heightened public scrutiny and, in some cases, adjusted practices to avoid reputational damage. Conversely, in digital job marketplaces, widespread visibility led both to emergent self-regulation and backlash from established intermediaries.

Proposition 2 (P2): *Hybrid systems with elastic digital interfaces formalize more quickly than rigid patronage-based ones.*

For instance, digital madrasa curricula via YouTube expanded reach and gradually gained acceptance in formal education policymaking. In contrast, elite-driven shalish systems, even when digitally visible, resisted formal integration due to clientelistic control.

Proposition 3 (P3): *Digital exposure amplifies positive feedback in responsive institutions but negative feedback in elite-captured systems.*

Bangladeshi police responding to viral harassment hashtags illustrate a positive loop, triggering policy responsiveness. On the other hand, in rural governance, when local elites harnessed online platforms to reinforce undemocratic norms, exposure reinforced dysfunction—negative feedback loops took hold.

Proposition 4 (P4): *Hybrid nodes are enduring institutional features that can be shaped through policy rather than eliminated.*

The proliferation of digital labor platforms and online madrasa teaching suggests that hybrid nodes are not transitional; they are durable and serve functional needs. Policymakers acknowledging and integrating such nodes—for instance, by accrediting digital madrasa learning or formalizing gig work via registration—affirm their lasting presence and potential leverage points for reform.

In sum, the Bangladesh and global cases suggest that the AIF framework captures important dynamics of institutional adaptation in digital contexts. While the cases are illustrative rather than definitive, they indicate the potential analytical value of the framework for understanding interactions among digital, formal, and informal institutions.

6.2. Comparative Analysis: AIF vs. Existing Theoretical Frameworks

6.2.1. Modernization Theory

Classic modernization theory posits that as societies develop economically and educationally, they transition toward rational-legal governance structures (Gwynne, 2009). However, these cases reveal a more complex reality: informal mechanisms persist and often hybridize with formal systems through digital mediation. The AIF framework challenges the linear progression model of modernization by highlighting how modern digital technologies may reinforce informal practices—even as formal institutions exist—rather than simply displacing them.

6.2.2. Institutional Isomorphism

DiMaggio and Powell's concept of institutional isomorphism identifies coercive, mimetic, and normative pressures leading organizations to converge (DiMaggio & Powell, 1983). In digital hybrid spaces, similar pressures emerge—for instance, religious schools mimicking successful digital learning platforms (mimetic isomorphism), or government mandates pushing digital registration (coercive isomorphism). However, AIF goes further by focusing not on or-

ganizational convergence per se, but on fusion nodes where formal/informal and offline/online logics blend—adding nuance to homogenization dynamics.

6.2.3. Hybrid Governance / Hybrid Institutional Complexes (HICs)

HIC theory describes governance involving formal state, private, and informal institutions operating together (Abbott & Faude, 2021). AIF parallels this in concept but differs in scale and mechanism: HICs operate at global regulatory levels; AIF centers on local fusion nodes where digital platforms mediate institutional convergence. The AIF framework adds the dimension of **digital visibility** and the resulting feedback loops—absent in HIC literature—emphasizing how platform dynamics shape institutional legitimacy and adaptation.

6.3. Normative Implications

The AIF framework posits that reform strategies should pivot from suppressing informal practices toward integrating and shaping hybrid systems, advancing three core principles. First, it calls for acknowledgement over suppression: denying informal–digital fusion—for instance, by forbidding WhatsApp-based job brokering—fails to eliminate such practices, whereas integration through recognition, training, or formal channels offers greater efficacy. Second, the framework advocates for policy design through hybridity, exemplified by accrediting digital madrasa instruction or regulating gig platforms to ensure labor protections, thereby transforming fusion nodes into regulated innovations. Third, it emphasizes leveraging digital visibility responsibly, proposing that institutional mechanisms can be established to route digital complaints to officials efficiently, turning viral exposure into structured accountability rather than chaotic backlash. In short, reform should treat hybrid institutions as assets to be integrated, not obstacles to be eliminated.

6.4. The Digital Dimension: Platform Architecture Matters

Platform architecture shapes how fusion nodes function and evolve:

- **Facebook** emphasizes broad public visibility and algorithmic amplification of content, increasing the potential for mass scrutiny—but also enabling reputational capture by elites able to manipulate narratives.
- **WhatsApp**, by contrast, supports closed, encrypted group communication, which facilitates coordination but limits public oversight, lending itself to both grassroots resilience and elite gatekeeping.
- **TikTok**, with its sensational, short-form videos, can ignite rapid mobilization but may induce superficial engagement or misinterpretation—affecting legitimacy cycles differently from subtler platforms.

This differentiation echoes platform governance scholarship showing that technical design influences decentralized authority, deliberation, and accountability (Jhaver et al., 2021). Accordingly, AIF posits that platform features—visibility mechanics, community moderation, data governance—critically affect whether feedback loops become reformist or retrogressive.

Facebook, WhatsApp, and TikTok are selected as illustrative examples rather than an exhaustive list of platforms. They represent distinct platform architectures—public networked visibility (Facebook), private group-based communication (WhatsApp), and algorithmically driven short-video dissemination (TikTok)—allowing the framework to demonstrate how different technological affordances may influence institutional fusion processes in different ways.

7. CONCLUSION

This study advances a conceptual framework of adaptive informal–formal (AIF) fusion to explain how institutions evolve in response to both historical constraints and digital acceleration. Rather than viewing informal practices as transient pathologies to be eradicated, this framework argues that hybrid institutional arrangements are enduring features of governance systems. They reflect pragmatic accommodations to context-specific pressures—including colonial legacies, uneven state capacity, and rapidly shifting socio-technical environments—rather than failures of modernization.

The AIF framework integrates three dimensions rarely combined in prior research: (1) the fusion of formal and informal institutions, drawing on insights from institutional economics and postcolonial sociology; (2) the digitally mediated transformation of legitimacy contests, informed by digital sociology; and (3) the concept of elasticity versus rigidity as a way to assess institutional resilience over time. Together, these elements offer a lens that explains why informal norms persist even when formal reforms are enacted, and why attempts to "purify" governance by eliminating informality often generate perverse effects.

This approach departs from modernization theory, which typically assumes a linear progression from informality to formality (Gwynne, 2009), and from institutional isomorphism, which emphasizes convergence under global pressure but downplays local adaptations (DiMaggio & Powell, 1983). It also extends hybrid governance literature (Boege et al., 2008; Meagher, 2012), which documents coexisting authority systems but seldom explains their dynamism under conditions of digital exposure. By foregrounding legitimacy negotiations across online and offline domains, this study shows that hybrid institutions are not static compromises but evolving ecosystems.

The findings challenge the persistent assumption in policy discourse that informal practices are transitional de-

fects to be eliminated through technocratic reform. Instead, informality operates as a stabilizer, filling gaps left by under-resourced formal systems and buffering shocks in moments of institutional stress. Whether in local governance, education systems, or labor regulation, informal norms often provide the "glue" that keeps systems functioning when formal mechanisms falter. Far from signaling failure, this hybridity demonstrates institutional resilience—albeit at the cost of transparency and predictability.

The cases analyzed in this study demonstrate that attempts to forcibly suppress informal practices—through legal crackdowns, donor-imposed reforms, or algorithmic surveillance—often weaken rather than strengthen institutions. When informal mechanisms are delegitimized without being replaced by viable alternatives, the result is institutional rigidity: formal systems that cannot adapt to changing social or economic demands. Conversely, systems that acknowledge and strategically integrate informal practices—rather than pretending they do not exist—tend to show greater elasticity, absorbing pressures without collapse.

The digital sphere has amplified these dynamics by exposing institutional negotiations to broader audiences. Social media platforms—whether Facebook, WhatsApp, or TikTok—differ not only in reach but also in architecture, shaping how legitimacy is constructed, contested, or eroded.

- **Facebook's open network structure** enables broad-based campaigns that can challenge formal authority but also trigger rapid state repression.
- **WhatsApp's encrypted, closed groups** allow informal actors to coordinate quietly, reinforcing hybrid governance networks away from public scrutiny.
- **TikTok's algorithmic virality** privileges short bursts of emotional legitimacy rather than sustained debate, reshaping how institutions respond to public pressure.

This variation underscores the need to consider platform governance as a mediating factor in institutional evolution. Digital technologies are not neutral tools; they actively structure which informal practices gain legitimacy and which are stigmatized or suppressed.

The theoretical contribution of this study opens three main pathways for further inquiry:

1. **Quantitative validation.** While this study relies on qualitative comparative analysis, future research could develop cross-country datasets to measure the degree of formal–informal fusion, track its evolution over time, and assess its impact on governance outcomes. Such an approach would test whether elasticity—as conceptualized here—correlates with developmental performance, conflict resilience, or citizen trust.
2. **Algorithmic governance as a new fusion layer.** As states and corporations increasingly deploy algorithmic tools to regulate behavior—whether through surveillance, credit scoring, or automated compliance checks—informal actors are already adapting. This emerging domain of "algorithmic informality" deserves systematic study: How do informal norms reshape or bypass automated rule enforcement? Does digital governance produce new forms of hybrid authority rather than replacing older ones?
3. **Policy applications.** The AIF framework offers practical insights for reformers in education, urban governance, and labor regulation. Rather than imposing rigid formal schemes that disregard context, policymakers can design systems that harness informal capacities while gradually strengthening formal rules. For example, community-based education monitoring networks, informal housing regularization programs, or hybrid labor inspection systems could provide flexible pathways for institutional improvement without triggering backlash or collapse.

In conclusion, this study reframes the debate on governance in developing and postcolonial contexts. Hybrid institutions are not signs of arrested development but adaptive responses to historical legacies and contemporary pressures. The fusion of formal and informal mechanisms, far from being a temporary stage, is a permanent structural feature of modern governance. Recognizing and theorizing this fusion—notably as digital platforms reshape it—can lead to more nuanced academic models, more realistic international reform agendas, and more resilient public institutions.

The challenge ahead is to build theories and policies that work with hybridity rather than against it. By doing so, scholars and practitioners can move beyond the sterile binary of "formal versus informal," toward an understanding of institutions as living systems—flexible, contested, and capable of evolution in an increasingly interconnected world.

ANNEX

Case Study 1: Digital Madrasa Education—YouTube and the AIF Fusion in Learning

Digital platforms like **10 Minute School**, initiated in 2015 as a YouTube channel, now provide complete academic curricula, including madrasa subjects, to Bangladeshi students across generations (10 Minute School, 2019). Simultaneously, the Gurukul Online Learning Network (GOLN) offers branded content hubs such as "Islamia Gurukul" that cater explicitly to madrasa education, reaching hundreds of thousands through free video lectures on YouTube and Facebook (Shams, 2021).

Digital Artifacts & Stakeholder Quotes

A student from *Islamia Gurukul* commented, "There has not been such professional initiative for students of madra-

sa curriculum in Bangladesh," underscoring both legitimacy and digital reach (Shams, 2021).

These platforms present hybrid "fusion nodes": offline religious schooling meets formal educational goals through digital dissemination.

Timeline
2015: 10 Minute School launches as YouTube-only initiative (10 Minute School, 2019).

2019: GOLN begins specialized content delivery, including madrasa subjects (Shams, 2021).

2023–2024: Adoption of digital portal tools for madrasas (notably in Indonesia, but indicative of regional parallels) suggests a growing institutional acceptance of hybrid digital–formal pedagogy (Bahari & Mukarramah, 2023).

Case Study 2: Village Shalish and Facebook Publicity—Governance in Digital Hybridization
Shalish, a traditionally informal village arbitration system, remains prevalent despite the introduction of formal village courts in 2006 (Islam, 2019). Public distrust stems from perceptions of bias and lack of accessibility, while village courts tower over slow, costly formal systems (Islam, 2019).

Digital Artifacts & Stakeholder Insights
In several villages, decisions from shalish are posted or shared on **Facebook groups**, enabling public visibility and prompting formal courts or local officials to respond or adjust behavior.

A village court official was quoted (within local media): "The decision of the court… is accepted later on after serious consultation by the village court authority," suggesting a hybrid institutional interplay (Siddiqi, 2019).

Timeline
2006: Village courts formalized alongside traditional shalish (Islam, 2019).

2020–2025: Increasing use of WhatsApp and Facebook to share grievance outcomes and arbitral resolutions.

Case Study 3: Informal Job Brokerage via Digital Channels in Bangladesh

Narrative Context
Bangladesh's informal labor market is vast—comprising over 85% of employment—and is characterized by significant wage disparities between formal and informal workers (Hossain Mohammad Yeasin, 2022). In response, digital platforms and apps like **Kormo**, developed in partnership with Google's Area 120, now match job seekers to informal work, offering résumés, job listings, and digital profiles (HRM Asia Newsroom, 2018).

Digital Artifacts & User Accounts
Job seekers connect via **digital platforms** to access the informal job market beyond state-regulated channels.

Timeline
2018: Launch of Kormo job app in Dhaka (HRM Asia Newsroom, 2018).

2022–2025: Growth of WhatsApp and Facebook–mediated job groups and reports of scams in digital job brokerage circles.

Case Study 4: Social Norms: Online Activism in Bangladesh
Digital activism campaigns emerged in Bangladesh to confront harassment and gender-based violence and to illustrate how social norms intersect with formal law enforcement through digital visibility (ActionAid Bangladesh, 2024). These movements function as hybrid nodes, where offline social expectations are amplified by online platforms, compelling police, local authorities, and political institutions to respond to grievances that previously received little attention.

Implications:

When institutional responsiveness aligns with online pressure, digital exposure legitimizes civil resistance, incentivizes policy or enforcement action, and reinforces institutional elasticity (Helmke & Levitsky, 2004).

Conversely, if state actors respond with repression—e.g., censorship or punitive policing—digital exposure can exacerbate rigidity, hardening institutional resistance to reform (Roy, 2023).

These dynamics exemplify how hybrid nodes can both accelerate reform and highlight governance vulnerabilities when formal institutions are unevenly responsive.

Cases on Global Parallels

M-Pesa (Kenya)

The mobile money platform M-Pesa combined formal banking regulations with trust-based informal savings networks in rural Kenya, demonstrating an adaptive hybrid institution (Jack & Suri, 2011). By providing digital access to financial services for populations historically excluded from formal banking, M-Pesa created a fusion node where informal trust systems and formal regulatory frameworks coexisted and reinforced one another. Its success reflects elasticity, with digital exposure expanding both legitimacy and uptake.

Favela Leadership (Brazil)
Community leaders in Brazilian favelas have used Instagram to negotiate with municipal authorities, blending informal community governance with formal political

processes (Cuvi, 2012; Medrado et al., 2020). These hybrid nodes leverage digital visibility to enhance legitimacy and negotiation power, serving as mediators between marginalized populations and official institutions. Visibility through social media also increases public scrutiny, incentivizing authorities to consider informal actors as legitimate stakeholders.

Gig Economy (U.S. & Europe)

Platforms such as Uber operate in regulatory gray zones, mixing informal gig labor practices with formal economic frameworks (Wood et al., 2019). Although the digital interface appears formalized, it accommodates informal work patterns—including flexible scheduling, rating-based regulation, and decentralized labor coordination—creating elastic hybrid systems that resist complete formalization while delivering economic function. These nodes illustrate the global resonance of fusion mechanisms, in which formal rules and digital-mediated informal practices coexist and adaptively respond to market and social pressures.

Effective alignment between online engagement and institutional responsiveness drives resilience, whereas misalignment or fragility in formal institutions fosters dysfunction (Helmke & Levitsky, 2004; Jack & Suri, 2011; Wood et al., 2019).

Ethics: There are no ethical issues with the publication of this manuscript.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare that there is no conflict of interest.

Authorship Contributions: The authors contributed to the study equally.

Financial Disclosure: The authors declared that this study has received no financial support

REFERENCES

10 Minute School. (2019). *Home – 10 Minute School*. 10 Minute School. https://10minuteschool.com/ Accessed on Jun 29, 2026.

Abbott, K. W., & Faude, B. (2021). Hybrid institutional complexes in global governance. *The Review of International Organizations. 17,* 263–291. [CrossRef]

ActionAid Bangladesh. (2024). *ActionAid Bangladesh*. Actionaidbd.org. https://actionaidbd.org/post/default/actionaid-bangladesh-launches-16-days-activism-campaign-against-gender-based-violence-focus-digital-security Accessed on Jun 29, 2026.

Ahmed, M. (2025a). *Social Media as Tools of Mob Violence and Mob Trials: A Study on South Asia with Special Focus on Bangladesh*. Preprints 2025, 2025042021. doi: 10.20944/preprints202504.2021.v1 [CrossRef]

Ahmed, M. (2025b). *Social Media Platform Facebook Has Become a Tool for Media Trials: A Study on Recent Mob Violence Against Citizens of Bangladesh*. Preprints 2025, 2025062211. doi: 10.20944/preprints202506.2211.v1 [CrossRef]

Alim, M. A., & Ali, T. O. (2023, December 4). *NGO-Shalish and Justice-Seeking Behaviour in Rural Bangladesh - BRAC Institute of Governance and Development*. BRAC Institute of Governance and Development. https://bigd.bracu.ac.bd/publications/ngo-shalish-and-justice-seeking-behaviour-in-rural-bangladesh/?utm Accessed on Jun 29, 2026.

Asadullah, M. N., & Chaudhury, N. (2009). Holy alliances: public subsidies, Islamic high schools, and female schooling in Bangladesh. *Education Economics*, *17*(3), 377–394. [CrossRef]

Baharun, H., & Mukarramah, A. (2023). Digital based Madrasas: The Strategy to Improve the Quality of Madrasah Ibtidaiyah Performance. *Al Ibtida: Jurnal Pendidikan Guru MI*, *10*, 119. [CrossRef]

Batlle, M. (2008). Informal Institutions and Democracy. Lessons from Latin America, de Gretchen Helmke y Steven Levitsky. *Perfiles Latinoamericanos*, *16*, 179–184. [CrossRef]

Bayart, J.-F. (1993) The State in Africa: The Politics of the Belly. Longman Group, England.

Boege, V. Brown, A., Clements, K., & Nolan, A. (2008). On hybrid political orders and emerging states: state formation in the context of 'fragility'. Berghof Handbook.

Castells, M. (2000). *The rise of the network society: The information age: Economy, society and culture* (v. 1). Wiley.

Chatterjee, P. (2004). *The Politics of the Governed: Reflections on Political Society in Most of the World*. New York: Columbia University Press.

Cuvi, J. (2012). Favela: Four Decades of Living on the Edge in Rio de Janeiro - By Janice Perlman. *International Journal of Urban and Regional Research*, *36*(2), 403–404. [CrossRef]

Dallakoti, G. (2024, June 27). *Formalization key to shared prosperity with workers in bangladesh's informal sector*. International Labour Organization. https://www.ilo.org/resource/article/formalization-key-shared-prosperity-workers-bangladeshs-informal-sector Accessed on Jun 29, 2026.

Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, *6*(2), 80–88. [CrossRef]

Dijck, J. van, Es, K. van, Helmond, A., & Vlist, F. van der (Eds.). (2025). *Governing the Digital Society*. Routledge. [CrossRef]

Dijck, V., Winkel, D., & Tobias, S. M. (2023). Deplatformization and the governance of the platform ecosystem. *New Media & Society*, *25*, 3438–3454. [CrossRef]

Dijck, van, Poell, T., & de Waal, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.

DiMaggio, R, & Powell. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, *48,* 147–160. [CrossRef]

Friedland, R., & Alford, R. (1991). Bringing Society Back In: P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis.* University of Chicago Press.

Fritzen, S. A. (2016). Smart citizens, smarter state: The technologies of expertise and the future of governing. *Governance*, *30*(1), 158–159. [CrossRef]

Fung, A. (2013). Infotopia: Unleashing the democratic power of transparency. *Politics & Society*, *41*(2), 183–212. [CrossRef]

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. MIT Press.

Gerbaudo, P. (2012). Tweets and the streets: Social media and contemporary activism. OAPEN Press. [CrossRef]

German, L., & Keeler, A. (2010). "Hybrid institutions": Applications of common property theory beyond discrete tenure regimes. *International Journal of the Commons*, *4*(1), 571–596. [CrossRef]

Gimenez-Jimenez, D., Calabrò, A., & Urbano, D. (2020). The neglected role of formal and informal institutions in women's entrepreneurship: a multi-level analysis. *Journal of International Entrepreneurship*, *18,* 196–226. [CrossRef]

Gümüsay, A. A., Raynard, M., Albu, O., Etter, M., & Roulet, T. (2022). Digital technology and voice: How platforms shape institutional processes through visibilization. *Digital Transformation and Institutional Theory*, *83*, 57–85. [CrossRef]

Gwynne, R. N. (2009, January 1). *Modernization Theory.* In A. Kobayashi, (Ed.), Elsevier.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *The Journal of Modern African Studies*, *11*(1), 61–89. [CrossRef]

Hasan, N. B. (2025). The Growth of Online Freelancing in Bangladesh: Economic Integration and Labor Vulnerability. *International Journal of Research and Innovation in Social Science*, *IX*(IV), 3112–3118. [CrossRef]

Helmke, G., & Levitsky, S. (2004). Informal institutions and comparative politics: A Research agenda. *Perspectives on Politics*, *2*(04), 725–740. [CrossRef]

Helmke, G., & Levitsky, S. (2006). *Informal institutions and democracy: Lessons from Latin America*. Johns Hopkins University Press.

Horak, S & Yang, I. (2016). Affective networks, informal ties, and the limits of expatriate effectiveness. *International Business Review, 25*(5), 1030–1042. [CrossRef]

Horak, S., & Suseno, Y. (2022). Informal Networks, Informal Institutions, and Social Exclusion in the Workplace: Insights from Subsidiaries of Multinational Corporations in Korea. *Journal of Business Ethics*, *186*(3), 633–655. [CrossRef]

HRM Asia Newsroom. (2018, September 30). *New app helps Bangladesh jobseekers navigate the informal economy - HRM Asia*. HRM Asia. https://hrmasia.com/new-app-helps-bangladesh-jobseekers-navigate-the-informal-economy/?utm Accessed on Jun 29, 2026.

Islam, M. T. (2019). Rural dispute resolution in Bangladesh: how do village courts safeguard justice? *Contemporary South Asia*, *27*(1), 58–65. [CrossRef]

Islam, S., & Sakib, A B M. (2024). Learning through Facebook and YouTube: Perceptions of Students of University of Dhaka. *Teacher's World: Journal of Education and Research*, *50*, 95–118. [CrossRef]

Islam, T. (2018, February 19). *In quest of justice: dispute resolution in rural Bangladesh - South Asia @ LSE*. South Asia @ LSE; London School oF Economics and Political Science. https://blogs.lse.ac.uk/southasia/2018/02/19/dispute-resolution-in-rural-bangladesh-in-quest-of-justice/?utm Accessed on Jun 29, 2026.

Jack, W., & Suri, T. (2011). Mobile Money: The Economics of M-PESA. *National Bureau of Economic Research* (Working paper no. 1672). NBER Working Paper Series. [CrossRef]

Jhaver, S., Frey, S., & Zhang, A. (2021). Decentralizing Platform Power: A Design Space of Multi-level Governance in Online Social Platforms. *ArXiv (Cornell University)*, Preprint. doi: 10.48550/arxiv.2108.12529

Ledeneva, A. V. (2006). *How Russia really works: the informal practices that shaped post-Soviet politics and business*. Cornell University Press.

Lund, C. (2006). Twilight institutions: An introduction. *Development and Change*, *37*(4), 673–684. [CrossRef]

Lungu, M. (2022). The coding manual for qualitative researchers. *American Journal of Qualitative Research*, *6*, 232–237. [CrossRef]

Mahoney, J., & Thelen, K. (2009). *Explaining Institutional Change*. Cambridge University Press. [CrossRef]

Markham, A., & Buchanan, E. (2012). *Ethical DecisionMaking and Internet Research Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. https://pure.au.dk/ws/files/55543125/UN%20 Accessed on Jun 29, 2026.

Marwick, A. (2013). Ethnographic and qualitative research on twitter. In K. Weller, A. Bruns, C. Puschmann, J. Burgess, & M. Mahrt, M. (Eds), *Twitter and society* (pp. 109–122). Peter Lang Press.

Mattsson, M., & Mobarak, A. M. (2023). *Formalizing Dispute Resolution: Effects of Village Courts in Bangla-*

desh. EliScholar – a Digital Platform for Scholarly Publishing at Yale. https://elischolar.library.yale.edu/egcenter-discussion-paper-series/1100/?utm Accessed on Jun 29, 2026. [CrossRef]

Meagher, K. (2012). The strength of weak states? Non-state security forces and hybrid governance in Africa. *Development and Change*, *43*(5), 1073–1101. [CrossRef]

Medrado, A., Cabral, T., & Souza, R. (2020). Favela digital activism: The use of social media to fight oppression and injustice in Brazil. In C. Martens, C. Venegas, E. F. Salvio Sharupi Tapuy, (Eds.), *Digital activism, community media, and sustainable communication in Latin America* (pp. 177–201). Springer Nature. [CrossRef]

Merton, R. K. (1968). *Social theory and social structure* (Enlarged ed.). The Free Press.

Minbaeva, D., Ledeneva, A., MuratbekovaTouron, M., & Horak, S. (2023). Explaining the persistence of informal institutions: The role of informal networks. *Academy of Management Review*, *48*, 556–574. [CrossRef]

Nason, R. S., & Bothello, J. (2022). Far from Void: How Institutions Shape Growth in the Informal Economy. *Academy of Management Review*, *48*(3). [CrossRef]

North, D. C. (1990). *Institutions, institutional change and economic performance* (pp. 83-104; pp. 107-117). Cambridge University Press. [CrossRef]

Patton, M. (2015) Qualitative research and evaluation methods (4th ed.), Sage Publications.

Rahman, M., Bhattacharya, D., & Al-Hasan, M. (2019). Dimensions of informality in bangladesh labour market and the consequent wage penalty. *South Asia Economic Journal*, *20*(2), 224–247. [CrossRef]

Riaz, A. (2008). *Faithful education: Madrassahs in South Asia*. Rutgers University Press.

Rohit, M. (2025, August 2). *Fact Check: Do Madrasa Students in Bangladesh Have Access to Modern Subjects?* Diplotic. https://www.diplotic.com/madrasa-students-in-bangladesh/#google_vignette Accessed on Jun 29, 2026.

Roy A. (2005). Urban informality: Toward an epistemology of planning. *Journal of the American Planning Association, 71*, 147–158. [CrossRef]

Roy, D. R. K. (2023). *Digitalisation, online activism and civic politics in contemporary Bangladesh*. LSE South Asia Center. https://www.lse.ac.uk/south-asia-centre/assets/documents/WorkingPapers/Bangladesh-ROY-Online-Activism.pdf Accessed on Jun 29, 2026.

Roy, S., Huq, S., & Rob, A. B. A. (2020). Faith and education in Bangladesh: A review of the contemporary landscape and challenges. *International Journal of Educational Development*, *79*, 102290. [CrossRef]

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Shams, S. A. (2021, August 23). *The Daily Star*. The Daily Star. https://www.thedailystar.net/youth/education/news/madrasa-english-medium-gurukul-online-learning-network-reaching-23-lakh-students-2158616?utm Accessed on Jun 29, 2026.

Shumshunnahar, S., Ruma, A. R., & Amin, R. (2025). The Constraints and Prospects of Integrating ICT in Bangladesh's Madrasa Education: Data and Explanations from Dhakil-level. *Pancasila International Journal of Applied Social Science*, *3*(01), 1–18. [CrossRef]

Siddiqi, B. (2019). *Activating Village Court in Bangladesh: A Prospect for Local Dispute Management*. Ministry of Local government, Donor, UN Agencies Project.

Srnicek, N. (2017). *Platform capitalism*. Polity Press.

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, *20*(3), 571–610. [CrossRef]

Suri, B. (2025, February 15). *Digital Sociology: Origins, Evolution, and Global Implications in a Tech-Driven World*. GLOBAL RESEARCH AND . https://www.grefiglobal.org/post/digital-sociology-origins-evolution-and-global-implications-in-a-tech-driven-world?utm

Sutradhar, R. (2023). *Universepg An Independent International Academic Publisher*. Universepg.com. https://www.universepg.com/ajssls/present-conditions-and-ensuring-rule-of-laws-at-local-level-a-case-study-on-village-court?utm Accessed on Jun 29, 2026.

Swedberg, R. (2014). *The art of social theory*. Princeton University Press. [CrossRef]

Tasin, A., Chowdhury, A., & Rahman, A. (2025). *Monsoon Uprising in Bangladesh: How Facebook Shaped Collective Identity*. ArXiv, Preprint. doi: 10.48550/arXiv.2508.02498

TBS. (2025, August 29). *A force in denial: The unfinished story of police reform*. The Business Standard. https://www.tbsnews.net/features/panorama/force-denial-unfinished-story-police-reform-1223606?utm Accessed on Jun 29, 2026.

Thornton, P., & Ocasio, W. (2008). Institutional Logics. *The Sage Handbook of Organizational Institutionalism*. [CrossRef]

Van Dijck, J. (2024). Governing platforms and societies. *Platforms & Society*, *1*, 1–2. [CrossRef]

Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2021). Deplatformization and the governance of the platform ecosystem. *New Media & Society*, *25*(12), 146144482110456. [CrossRef]

Van Dijck, J., van Es, K., Helmond, A., & van der Vlist, F. (2025). *Governing the digital society*. Amsterdam University Press. [CrossRef]

Wenner, M. (2020). Trajectories of hybrid governance: Legitimacy, order and leadership in India. *Development and Change*, *52*(2), 265–288. [CrossRef]

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, *33*(1), 56–75. World Bank. (2016, August 30). *Madrasas increase school enrollment but face quality challenges*. World Bank Group. https://www.worldbank.org/en/news/press-release/2010/08/10/madrasas-increase-school-enrollment-but-face-quality-challenges?utm Accessed on Jun 29, 2026. [CrossRef]

Yeasin, H. M. (2022). Informal Sector and Economic Growth in Bangladesh. *Interdisciplinary Journal of Applied and Basic Subjects, 12*(1), 48–61.

Yin, R. K. (2017). *Case study research and applications: Design and methods*. SAGE Publications.

Zaywa. (2021, July 14). *Madrasa views increases by 220% during the pandemic*. Zawya.com; Zawya. https://www.zawya.com/en/press-release/madrasa-views-increases-by-220-during-the-pandemic-s1m0vrsh?utm Accessed on Jun 29, 2026.

Yıldız Social Science Review
Web site information: https://yssr.yildiz.edu.tr
DOI: 10.51803/yssr.1831209

Original Article / Orijinal Makale

Dialectical Analysis of Scientific Ethics and Human Rights Violations in the Era of Artificial Intelligence

Yapay Zeka Çağında Bilimsel Etik ve İnsan Hakları İhlallerine Dair Diyalektik Bir Analiz

Turgut ÇILĞIN

Department of Administrative and Financial Affairs, Adıyaman University, Adıyaman, Türkiye

İdari ve Mali İşler Bölümü, Adıyaman Üniversitesi, Adıyaman, Türkiye

ARTICLE INFO

Article history
Received: 8 April 2026
Revised: 12 June 2026
Accepted: 26 June 2026

Key words:
Artificial intelligence, algorithmic discrimination, human rights, scientific ethics

MAKALE BİLGİSİ

Makale Hakkında
Geliş tarihi: 08 Nisan 2026
Revizyon tarihi: 12 Haziran 2026
Kabul tarihi: 26 Haziran 2026

Anahtar kelimeler:
Yapay zeka, algoritmik ayrımcılık, insan hakları, bilimsel etik

ABSTRACT

This article examines the stage reached in the transformation of Artificial Intelligence from a simple data processing tool into an autonomous socio-technological and potentially transhumanist entity capable of making independent decisions, as well as the scientific ethical and human rights paradoxes that this transformation has brought about due to algorithmic reasons. When evaluated alongside its goals such as scientific progress and technological promise research findings reveal that systemic risks, including hallucinations that undermine academic integrity and reliability, as well as algorithmic discrimination that exacerbates historical inequalities, constitute fundamental crisis points for the legitimacy of artificial intelligence. The article aims to access current data using literature review research method. It examines efforts to align AI with human rights and scientific ethics in light of regulations established by international organizations and proposes strategies to resolve these paradoxes. In conclusion, given the algorithmic risks of AI in academic processes where knowledge is produced, it is essential for authors and reviewers to adhere to ethical guidelines. The rapid advancement and growing power of AI highlight the need to align it with international norms and values focused on scientific ethics and human rights worldwide.

Cite this article as: Çılğın, T. (2026). Dialectical Analysis of Scientific Ethics and Human Rights Violations in the Era of Artificial Intelligence. *Yıldız Social Science Review, 12*(1), 73–88.

ÖZ

Bu makale, yapay zekanın basit bir veri işleme aracından bağımsız kararlar alabilen özerk bir sosyo-teknolojik ve potansiyel olarak trans hümanist varlığa dönüşümünde ulaşılan aşamayı ve bu dönüşümün algoritmik nedenlerle ortaya çıkardığı bilimsel etik ve insan hakları paradokslarını incelemektedir. Bilimsel ilerleme ve teknolojik vaatler gibi hedefleriyle birlikte değerlendirildiğinde, araştırma bulguları, akademik bütünlüğü

***Sorumlu yazar / Corresponding author**
*E-mail address: tcilgin@adiyaman.edu.tr

ve güvenilirliği zedeleyen halüsinasyonlar ile tarihsel eşitsizlikleri daha da şiddetlendiren algoritmik ayrımcılık gibi sistemik risklerin, yapay zekanın meşruiyeti için temel kriz noktaları oluşturduğunu ortaya koymaktadır. Makale, literatür taraması araştırma yöntemini kullanarak güncel verilere ulaşmayı amaçlamaktadır. Uluslararası kuruluşlar tarafından belirlenen düzenlemeler ışığında, yapay zekayı insan hakları ve bilimsel etikle uyumlu hale getirme çabalarını incelemekte ve bu paradoksları çözmek için stratejiler önermektedir. Sonuç olarak, bilginin üretildiği akademik süreçlerde yapay zekanın algoritmik riskleri göz önüne alındığında, yazarların ve hakemlerin etik kurallara uymaları esastır. Yapay zekanın hızlı ilerlemesi ve artan gücü, onu dünya çapında bilimsel etik ve insan haklarına odaklanan uluslararası norm ve değerlerle uyumlu hale getirme ihtiyacını vurgulamaktadır.

Atıf için yazım şekli: Çılğın, T. (2026). Dialectical Analysis of Scientific Ethics and Human Rights Violations in the Era of Artificial Intelligence. *Yıldız Social Science Review, 12*(1), 73–88.

1. INTRODUCTION

The exponential growth and development of artificial intelligence (AI) systems, particularly over the past 10 years, has triggered one of the most profound and perhaps the most significant socioeconomic transformations in human history. This multidisciplinary momentum has necessitated a redefinition of both the natural and engineering sciences as well as the social sciences. Today, AI has evolved from being merely a passive tool for processing data into a socio-technological phenomenon capable of generating its own data, making autonomous decisions, and possessing the potential to evolve into a transhumanist form of existence (Akpınar et al., 2025; Gabriel, 2020; Küçükvardar et al., 2020).

When examined from a dialectical perspective, although technological advancements have not yet reached a transhumanist form, the current use of AI particularly in conjunction with robot and automation systems provides numerous significant benefits for society while simultaneously creating new and extraordinary ethical and legal issues in many respects. These issues and violations may arise both from external attacks on AI algorithms during their production and training phases (Çılğın, 2025) and from algorithmic and epistemological limitations inherent to AI itself.

While AI, owing to its innovative spirit, holds the promise of accelerating scientific discoveries and enhancing social welfare, it simultaneously poses risks such as hallucinations that undermine scientific reliability due to inherent inadequacies, lack of control, or limitations; surveillance mechanisms that disregard privacy; data usage that infringes on intellectual property rights, epistemological approach issues in accessing scientific knowledge particularly in the social sciences and algorithmic biases that perpetuate historical prejudices (Bains, 2024; European Parliament [EP] 2020; INRA. AI, 2025).

When examining the dimension of external attacks on AI, it is observed that the statistical and data-driven structure of machine learning systems trained on data in data centers makes them vulnerable to new hacker attacks targeting AI's security, privacy, and safety unlike the threats faced by traditional software systems. Adversarial Machine Learning (AML) issues arising from external threats during the production and training phases of AI result from attacks involving the adversarial manipulation of training data, the infiltration of adversarial inputs into the data center that negatively impact the AI system's performance, or the extraction/theft of sensitive information from training data accessible to the model (Vassilev et al., 2025).

This study employed a literature review methodology. Addressing the problems arising from AI's structural limitations -in addition to those caused by well-known external attacks- from both algorithmic (technical) and epistemological (philosophical) perspectives could help break the cycle of stagnation and inertia in these discussions, allowing them to proceed in a more dynamic manner. This article specifically addresses, particularly in relation to the social sciences, issues of epistemic authority and accountability loss, epistemic opacity and transparency, algorithmic bias and discrimination, epistemic hegemony, loss of meaning and context, the mechanization of academic labor in knowledge production processes, reductionism, the legislative/nomothetic approach, singularity, the loss of subject-object distinction and reflexivity, social physics and the control society, the loss of dense description and superficialization, the erasure of relational nature and empathy, the restriction of flexibility and improvisation, the atrophy of epistemic reflexivity, and linguistic and cultural imperialism, etc. By drawing on findings regarding these risks, the aim is to contribute to raising awareness of the critical points that must be considered in the use of the relevant AI.

Whether stemming from AI's own technical and philosophical limitations or external attacks, the harms caused by these issues represent a risky dimension that concerns all of humanity. Therefore, it is evident that addressing AI-related issues and violations within the framework of scientific ethics and human rights which have been established

through centuries of struggle for the benefit of all humanity is the correct approach (Resnik & Hosseini, 2025). These issues require the scientific community to develop and internalize an international, shared "Quality and Safety System" that includes the necessary principles for preventing the harms of AI and ensuring its use is in harmony with society and for the benefit of society without necessitating a change in established scientific ethical norms.

Like other technologies, AI provides significant conveniences and contributions to humanity's daily life. However, it is a highly probable risk that must be taken into account: beyond a certain point, these benefits could lead to harms in the opposite direction, reach a level that causes problems for humanity, and create an ethical and legal paradox between technology and humanity. As the development and proliferation of AI proceed at an astonishing pace, there is a need for interdisciplinary and multidisciplinary research integrated with a dialectical approach and socio-technological imagination to raise awareness of the risks it may pose from an ethical and human rights perspective and to take the necessary precautions. This article aims to make a specific contribution to the literature regarding the paradoxes of AI in relation to scientific ethics and human rights, along with proposed solutions. The conclusion is that, rather than banning AI as a solution to the risks it has caused or may cause, an international consensus must be reached to ensure its use aligns with universal values such as transparency, accountability, and human dignity without stifling its innovative spirit and remains community-focused.

2. MATERIALS AND METHODS

This article aims to explore solutions by focusing on the epistemological dimensions of the paradoxes created by AI's potential violations of scientific ethics and human rights principles. At the same time, it aims to contribute to identifying methods and insights that will form the basis for establishing criteria to be considered when defining the framework of ethical and legal constraints to resolve the problems created by AI and mitigate risks.

The primary research questions addressed in this study are as follows:

Q1: What are the ethical dimensions of artificial intelligence and the ethical violations in academic processes?

Q2: Does artificial intelligence pose algorithmic risks that could lead to human rights violations and discrimination?

Q3: What are the scientific and ethical risks associated with the epistemological and algorithmic structure of artificial intelligence?

The study employed literature review and document analysis methods. The literature review is presented in an interwoven manner under conceptual headings. In line with the objective of the research, in addition to literature reviews, reports from national and international institutions and organizations, as well as AI companies, were examined. The article addresses the paradoxes experienced in the academic field regarding AI within the framework of scientific ethics and human rights, along with proposed solutions.

In this study, a systematic literature review was conducted to comprehensively identify current trends, risks, and debates regarding the relationship between artificial intelligence and ethics. To access the relevant literature, Google Scholar was used as the primary academic search engine; Google Web was used for gray literature, institutional reports, and preprint studies; and the academic social network ResearchGate was actively utilized to obtain full-text articles from authors when they were locked or otherwise inaccessible. To ensure the study reflects current findings and contemporary discussions, a time restriction, inclusion criterion covering the last 10 years (2016–2026) was applied during the data collection process. In addition, reference has been made to works from earlier periods concerning the philosophical aspects of the subject. The searches were conducted in both Turkish and English to cover both national and international literature. Artificial Intelligence, Artificial Intelligence and Ethics, Automation and Ethics, Robots and Ethics, The Harms of Artificial Intelligence, Technological Developments and Ethics, Artificial Intelligence and Algorithmic Discrimination, Artificial Intelligence and Human Rights such terms and phrases have been searched in both Turkish and English. To ensure a thorough analysis, accurate interpretation, and prevention of data loss for sources in languages other than English or containing complex terminology, Google Translate was utilized as a supplementary translation tool during the review process. Identified studies were analyzed by following the identification, search, eligibility, and inclusion stages outlined in the PRISMA 2020 guidelines.

3. FINDINGS

A total of 975 articles were included in the sample, comprising 840 articles from Google Scholar, 90 reports/analyses from Google (Web), and 45 articles/preprints from ResearchGate. 145 sources were excluded from the evaluation because they were duplicates. The remaining 830 sources were subjected to a quick review at the title/abstract level. 615 titles and abstracts were excluded because they did not meet the study's research criteria. The remaining 215 articles were selected for full-text review. The full texts of 22 articles could not be reviewed because the links were broken. As a result, the full text of 193 articles and reports was reviewed. A total of 48 articles or reports were excluded from the study because they lacked empirical or theoretical depth regarding AI ethics. 38 were excluded because they

focused solely on general job losses rather than algorithmic discrimination or human rights. 24 were excluded because their methodological quality was found to be inadequate when translated. Of these, 83 were cited as sources in the article. Tables containing the most important articles cited and their content have been added at the end of each section.

This article seeks solutions by addressing the three questions summarized above, drawing on the findings that emerge. Chapter 2, which examines the academic paradoxes of artificial intelligence that conflict with scientific ethics, serves as an answer to the first question. In response to the second question, Chapter 3 addresses the discriminatory algorithmic aspects of artificial intelligence. In Chapter 4, a philosophical perspective is introduced to analyze the epistemological aspects of artificial intelligence algorithms that violate ethics and human rights. This chapter examines whether artificial intelligence faces limitations due to its reliance on a positivist epistemological algorithm. The conclusion section proposes solutions to the issues identified in the findings. Given that the positivist epistemological approach adopted for the construction of the algorithm could potentially give rise to risks related to scientific ethics and human rights, it was considered that this possibility warrants philosophical scrutiny. For this reason, the aim was to enrich the research with an interdisciplinary dimension by adding Chapter 4 as a section. Thanks to the philosophical exploration in Chapter 4, the article has evolved into an interdisciplinary field, and a new and different contribution/synthesis has been made to the literature by investigating the possibility that the positivist approach used during the algorithmic construction of artificial intelligence may lie at the root of ethical and human rights violations in AI.

3.1. Scientific Ethics Violations and the Risk of Academic Integrity

3.1.1. The hallucination paradox and model collapse

AI systems technically operate using a 'next token prediction' mechanism. This architectural structure makes it technically impossible for the system to respond with 'I don't know'; instead, it causes it to generate the most probable data, even if it bears no relation to reality. Hallucination rates ranging from 33% to 51% in complex academic questions give rise to convincing yet false content, conceptualised in the literature as 'botshit'.

A more critical risk is the phenomenon of 'model collapse'. This cycle, in which AI models are fed with the synthetic data they themselves generate, leads to models becoming detached from reality and creating 'scientific sinkholes'. This situation carries the risk that scientific knowledge may in future be caught in a vortex of misinformation that collapses in on itself.

3.1.2. The contamination of academic literature and retraction data

Fake DOI numbers and the generation of fictitious references are poisoning academic databases. For example, it has been found that 38% of the 178 academic references generated by ChatGPT were entirely fabricated. This situation is systematically undermining the reliability of the scientific citation chain. Table 1 shows the retraction rates of articles that have reached a level of concern.

3.1.3. The crisis in authorship and peer review

The replacement of human labour with algorithmic processes in the academic production pipeline is creating an ethical accountability gap. An analysis of 156,000 annual article submissions revealed that 75% of articles were generated by AI. In the peer-review process, editorial checks have revealed that over 80% of reviewers use AI covertly whilst preparing their assessment reports, resulting in superficial evaluations that take texts out of context.

3.2. Algorithmic Discrimination and Human Rights Risks

3.2.1. Cases of systemic bias

Healthcare (OPTUM): An algorithm used in the US healthcare system systematically calculated lower health

Table 1. Retraction rates of articles

Institution/Database	Status&Statistics	Primary Ground for Violation	Risk Level & Impact
Springer	600+ Retracted Papers	AI suspicion and fabricated data generation	Critical: Peer-review process manipulation
IEEE	200+ Removed Papers	Publication violations and unmonitored AI use	High: Contamination of engineering literature
Elsevier	400+ Retracted Papers	Suspected AI fraud (e.g., Heliyon)	Critical: Abuse of prestigious open-access journals
Wiley	1,700 Papers (2023 Data)	Data integrity and academic misconduct	Systemic: Mass infiltration by paper mills
Overall Growth Rate	%900 (9-Fold Increase)	Retraction Watch Report: AI-driven contamination	Catastrophic: Global crisis of scientific credibility

need scores for Black patients compared to white patients, as it focused on cost data rather than medical needs.

Human Resources (Amazon): The company's recruitment algorithm automatically filtered out CVs containing the terms "woman" or "Black", based on a male-dominated dataset spanning the last 10 years.

Judiciary (COMPAS): The risk assessment software violated the right to a fair trial by labelling Black defendants as "high-risk" at a rate 77% higher than that of white defendants.

3.2.2. Biometric tracking and surveillance

The demographic error rates in facial recognition systems place a disproportionate burden on minority groups. According to NIST data, these systems exhibit error rates of up to 34% for women with darker skin tones, whilst achieving the highest accuracy for men with lighter skin tones. This technical shortcoming leads to violations of the right to anonymity and personal safety.

3.2.3. Transparency and the black box problem

The fact that the inner workings of algorithms are shielded from scrutiny on the grounds of 'trade secrets' renders individuals' right to object and their expectation of transparency impossible. This opacity in decision-making processes acts as a technical barrier that hinders the defence of constitutional rights. However, the fact that artificial intelligence companies are able to obtain individuals' personal data without any restrictions and use it in AI databases without permission is a clear indication that the situation regarding the infringement of copyright and industrial property rights has reached an extremely unfair stage.

3.3. Epistemological Limitations and a Critique of Positivism

3.3.1. The distinction between data, information and wisdom

Although AI possesses superior speed in converting data into functional information, it cannot reach the level of 'wisdom' that is unique to the human mind. Whilst the human mind processes decisions through a dialectical filter involving subjective experience (qualia), conscious intentionality and emotional depth; AI operates within a phenomenological void, devoid of any sense of what it is doing.

3.3.2. Positivism and the critique of social physics

The Social Physics Approach: Modern artificial intelligence algorithms are based on the 'social physics' approach, which follows in the tradition of Auguste Comte and Emile Durkheim and tends to examine social phenomena as if they were physical objects. This approach, whilst attempting to measure social phenomena as if they were physical objects, excludes human agency and social dynamism.

Ibn Khaldun and Ontological Integrity: Ibn Khaldun's multidimensional methodology does not exclude theological, philosophical and historical realities when analysing social events. In contrast to the reductionist positivism of artificial intelligence, Ibn Khaldun addresses the spirit of solidarity, unity and group identity (asabiyya) that binds the members of a community together, as well as social causality, with a human depth.

3.3.3. The Loss of sociological imagination

The replacement of sociological imagination which establishes the link between individual problems and macro-social structures with statistical and algorithmic models based on the positivist method leads to a loss of depth and erroneous generalisations in AI-driven social science research, as it becomes detached from its context.

4. THE SCIENTIFIC ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE AND ETHICAL VIOLATIONS IN ACADEMIC PROCESSES

Ethics are rules that, guided by rational knowledge, distinguish between right and wrong and facilitate the demonstration of appropriate behavior without harming the freedoms and personal rights of others. In other words, it is the principle of not desiring for others the inappropriate behaviors and desires that one would not wish for oneself (Aristotle, 2022; Doko, 2026; Spinoza, 2011). Scientific ethics, on the other hand, is a universal set of values that requires research to be conducted within the framework of integrity, impartiality, transparency, and accountability (Center for Responsible Digitalization Office, 2022; Brenneis et al., 2024; Merton, 1973). The rapid and effective development of AI has profoundly disrupted the way these principles are applied. By placing the scientific method which is shaped and must be applied within the framework of scientific ethics in a paradox with ethical principles, it has created a crisis of legitimacy (Ghosh, 2025; Resnik & Hosseini, 2025).

From the perspective of scientific ethics, these principles underscore the fact that scientists, company owners, and users who produce AI technologies bear primary responsibility toward society as a whole, future generations, and the ecosystem. For example, during the training phase of AI, the unauthorized use of data constitutes a violation of copyright or intellectual property rights, thereby infringing upon the principles of scientific data collection ethics and intellectual property rights to the detriment of the data's author or creator. The power derived from the process of digitizing and analyzing individual habits such as biometric data, medical records, or social media posts through AI can, in the next phase, threaten individuals' decision-making privacy and psychological integrity (Cheong, 2025; Digital Cooperation Organization [DCO], 2025; Radanliev, 2025; UN Global Compact Network Germany, 2023). Data

privacy is not merely a technical requirement but a fundamental ethical principle and human right that safeguards the dignity of the individual and society.

The unauthorized collection of data from the internet for training AI models, which involves the infringement of copyright and intellectual property rights, conflicts with both the principles of scientific ethics and the law. Obtaining consent from research participants and securing copyright or providing proper attribution when using article or research data are absolute ethical and legal requirements. In AI training processes, the personal data, creative works, inventions, and academic publications of millions of people are being used without adhering to any intellectual property rights or legal rules (Boyes, 2025; Distance Education Portal [UEK], 2026; European Commission, [EC] 2025; Shao, 2025b; U.S. Bureau of Cyberspace and Digital Policy, 2024).

Furthermore, modern AI systems, particularly deep learning algorithms, process massive amounts of data at extraordinary speeds to derive new, similar, or different patterns. However, the inner workings of this process are generally not transparent. The methods by which algorithms reach a conclusion and the methodological evidence supporting the result are not provided. This situation, which constitutes a violation of the principle of transparency in the scientific method, is defined in the literature as the "black box" problem. With the loss of transparency, the principles of reproducibility and verifiability, which form the foundation of the scientific method, are also considered to have been violated. When the logical or experimental chain and the data weights through which an AI-generated output or result is produced cannot be fully explained, the information obtained faces the risk of losing its scientific reliability and ceasing to be scientific knowledge (Brenneis et al., 2024; Doko, 2026; Ghosh, 2025; Lawvs, 2026; Resnik & Hosseini, 2025). AI developer companies act unethically by violating the principle of mutual respect when they fail to disclose information and methods regarding the development stages of AI while using information belonging to all individuals, societies, and institutions other than themselves without permission. This type of scientific knowledge production also contradicts the positivist scientific method.

On the other hand, due to the hallucinatory information generated by AI's algorithmic limitations, there is a risk of poisoning the epistemological nature of the process of scientific knowledge production and dissemination. This risk poses a more complex, chronic, and systemic threat than traditional, well-known academic violations such as plagiarism. Erroneous and fabricated information stemming from AI hallucinations accumulates in AI databases, creating a permanent contamination in the scientific literature. The hallucinations produced by AI are not merely technical errors but a structural condition stemming from the limitations of the model's probabilistic nature (Ghosh, 2025; Özer, 2024; Shao, 2025a; Shao, 2025b). AI is programmed to produce information and results unconditionally and without exception. Therefore, it never responds to any question with "I don't know." Being conditioned to produce information whether incorrect or correct compels AI to generate hallucinatory, fabricated information, effectively forcing it to lie.

According to 2025 research data on this topic, it has been revealed that even the most advanced AI models can exhibit hallucination rates ranging from 33% to 51% when answering complex academic questions (Graffius, 2026). It is an undeniable fact that AI can combine information from the training data used during its training phase in incorrect contexts, thereby generating convincing yet entirely false hypotheses (Özer, 2024). For example, in a test regarding COVID-19 risk factors, it was observed that AI generated hallucinatory false correlations and results based solely on data patterns, which did not align with medical facts (Resnik & Hosseini, 2025). It has also been observed that advanced AI models like ChatGPT generate non-existent article references, fabricated DOI numbers, and fake authors when responding to scientific inquiries. For example, in one case, it was determined that over 38% of 178 references were entirely hallucinated and fabricated by AI (INRA. AI, 2025; Özer, 2024).

Another long-term risk associated with hallucinations is the potential for the internet to become saturated with AI-generated content that is far from objective, inaccurate, and synthetic, leading to a risk of entering a misinformation feedback loop as future-generation AI models are trained on this fabricated data. In this process, referred to in the literature as "model collapse," the risk arises that AI models, having become disconnected from real-world knowledge and data, will carry out a misinformation production process through their own hallucinatory, biased, and false information cycles, thereby severing future generations' connection to factual information (Shao, 2025a). This risk indicates that science and knowledge are facing a historic ontological and epistemological rupture.

Furthermore, the integration of AI into scientific processes contributes to the creation of outputs referred to in the literature as "botshit" or "bot nonsense," and the spread of deepfake-style forgeries, which damage the reputations of innocent individuals, institutions, and governments (McCarthy et al., 2024). Particularly in elections that determine the fate of nations, the risk that voters will make choices based on fake and misleading online fabricated news and comments which rely on a mix of AI hallucinations, botshit, deepfakes, and political agendas and the spread of this misinformation indicates that we have entered an extremely dangerous era (Spicer, 2024).

Furthermore, a threefold increase in the number of new books was observed between 2022 and 2025 as AI usage became more widespread. However, a comparative

rating-based analysis of publication periods revealed that this increase was due to books produced using AI. It was observed that the quality of publications from the period when AI became widespread was significantly lower, and it was not disclosed that they were AI-generated (Reimers & Waldfogel, 2026). A similar significant increase and noticeable decline in quality are also evident in articles produced by AI. In a study, it was found that out of approximately 156,000 articles submitted to 87 journals in a single year, only about 39,000 were written by humans, while the remaining 117,000 were generated by AI. This indicates that 75% of the submitted articles were produced by AI (Cunningham, 2026).

In the academic field, AI is used both by authors during the writing phase and by peer reviewers during the evaluation phase of the publishing process, particularly in the production of articles or projects. The use of AI tools by researchers in writing articles carries the risk of undermining the very concepts of authorship and academia. A researcher having an AI generate an article entirely constitutes a violation of academic and scientific research ethics in and of itself. (Committee on Publication Ethics [COPE], 2023; International Committee of Medical Journal Editors [ICMJE], 2026; U.S. Bureau of Cyberspace and Digital Policy, 2024). Beyond its limited and supportive use in areas where AI can be considered legitimate provided it is disclosed, the undisclosed and unrestricted use of AI may constitute a clear violation of the principle of scientific integrity.

Every year, tens of thousands of publications containing erroneous data or results due to AI-related issues or human errors and weaknesses are published in high-impact journals despite not meeting publication standards. Thousands of these publications are subsequently retracted annually following complaints from third parties or detection by journal editors. However, during the time these flawed publications remain in circulation, thousands of subsequent publications that cite them produce results based on these erroneous findings regrettably, these results are also flawed. By remaining in the and the database, they trigger a widespread chain reaction of scientific inaccuracies (Cabanac, 2024).

Statistics covering numerous journals show that the rejection rate for articles prepared using AI has risen from 40% to 94% over the past two years. Striking examples are particularly evident in the fields of medicine and health sciences. During the 2024–2025 period, Springer has retracted over 600 articles due to AI-related suspicions and issues with fabricated data. IEEE removed over 200 conference papers from publication during the 2023–24 period. Elsevier retracted over 400 articles in journals such as Heliyon and the Alexandria Engineering Journal during the 2023-2025 period due to suspected AI fraud. Wiley retracted 1,700 articles in 2023. Taylor & Francis retracted over 60 articles in 2024. According to Retraction Watch reports, there has been approximately a ninefold increase (900%) in retractions due to AI use over the past four years (Kalite Akademik, 2026; The Retraction Watch Database, 2026). In 2024, it was found that approximately 14% of the 1.5 million biomedical research abstracts indexed in PubMed were generated by AI (He & Bu, 2025).

Ethical violations arising from the use of AI in the academic field are observed not only among authors but also during the article review process. The complete delegation of article evaluation processes to AI tools contains inconsistencies with the holistic "evaluation/holistic grading" method due to AI's algorithmic architecture, which focuses on identifying flaws and errors rather than holistic evaluation, and disrupts compositional integrity by isolating sections from the article's context and treating them as independent entities. This situation carries the risk of high-quality publications being filtered out or low-quality and flawed publications infiltrating the literature (Chinoracky & Stalmasekova 2025; Nordquist, 2019). This issue arises particularly and predominantly in the evaluation process of research and publications in the social sciences. This epistemological problem will be addressed in the subsequent sections of this article.

However, the current state of AI detection software used by journals does not appear reliable due to false positives caused by the threshold approach. If texts written by researchers employ highly specialized language and phrasing, they may sometimes be mistakenly identified by the algorithm as AI-generated articles based on its defined patterns. There is a strong likelihood that these evaluation tools will over- or under-flag the extent of AI usage in an article's content. This is because generative AI operates within a structure that varies not based on a standard acceptability threshold, but rather on how the tool is used, how the output is validated, the level of transparency among authors and reviewers, and editorial assessment . Due to these relative circumstances, while one AI system may flag an article as AI-generated at any given time, another may determine that the article was produced by a human at the same or a different time (Committee on Publication Ethics [COPE], 2025).

With the use of AI in peer review processes, there has been a significant increase in problematic peer review decisions containing erroneous findings. Upon closer examination, it is often found that the issues cited in the comments and decisions are not actually present in the article. Although journal policies prohibit the submission and evaluation of an article using any generative AI, checks conducted by editors have revealed that the rate of AI usage by reviewers during the article evaluation phase exceeds 80%. When editors detect or suspect that a reviewer's evaluation was generated by AI, and attempt to contact the reviewer

to confirm this or request the evaluation without AI, it is observed that in a significant number of cases, reviewers fail to respond to emails or deny using AI (Committee on Publication Ethics [COPE], 2026).

AI is also being used to help determine whether published or pending articles were generated by AI. However, AI lacks up-to-date and accurate information and has the potential to produce biased or incorrect information. Many technology advocates claim that AI will reduce the workload on the system and prevent low-quality articles from entering the literature. Nevertheless, thousands of flawed articles generated by AI continue to enter the literature. The likelihood of AI producing erroneous reports and hallucinatory references is high. Relying solely on AI for peer reviews is considered just as problematic as authors generating articles solely through AI (Notman, 2025). On the other hand, the fact that AI algorithms are based on English grammar and morphology leads to inequities, as scientific research in other less commonly used languages may result in a successful paper being deemed unsuccessful due to misunderstandings of concepts used in the language in which the paper was written (Halverson & Cannon, 2025; Shao, 2025b). For example, when a Turkish article is evaluated by AI algorithms based on English, there is a high risk that the AI will misinterpret certain nuanced words, phrases, or literary styles unique to Turkish resulting from its historical and geographical context and interactions with Ottoman Turkish, Persian, and Arabic leading to an unfair and erroneous evaluation of the article.

Another significant risk related to language is the risk of cultural and civilizational exportation. Language is a seed containing the codes of a society's culture and civilization; it is the most fundamental building block of a nation's and civilization's sustainability. The predominant use of English in AI tools inevitably leads to the export/transfer of British culture and perspectives to countries where other languages are spoken (Doko, 2026).

Another negative aspect of artificial intelligence recently discovered is its tendency to be excessively affirming, exhibiting traits of "flattery" or "hypocrisy." Research indicates that AI models provide responses designed to stroke users' egos and narcissistic tendencies, offering information that pleases them rather than helping researchers access accurate information. In doing so, it draws both from the general average characteristics of all humanity within the broader data pool and from the specific behavioral patterns the user has exhibited during their interactions with the system. This negative aspect, referred to in the literature as "AI psychosis," is considered a trust-eroding factor for academia (Cheng et al., 2026; Chandra et al., 2026; Hackenburg et al., 2026). The main ethical risks posed by AI are presented in Table 2.

5. ARTIFICIAL INTELLIGENCE, HUMAN RIGHTS VIOLATIONS, AND ALGORITHMIC DISCRIMINATION

Ethical violations also entail the violation of human rights. As AI permeates even the most sensitive aspects of social life, it has begun to challenge and even overcome the legal and ethical barriers designed to protect fundamental human rights (Akpınar et al., 2025; Oellig, 2024). The risks arising from these violations span a broad spectrum, ranging from individual freedoms to social justice. AI algorithms misinterpret historical inequalities in the datasets they are trained on as objective truths and encode them into their systems. This situation creates algorithmic bias and systematic discrimination that systematically disadvantages minority groups, people with disabilities, women, the poor, and indi-

Table 2. The ethical risks of artificial intelligence

Scientific ethics risk	Mechanism	Potential consequence	Related sources
Hallucination (fake data) (botshit)	Miscalculation of the probability of the next token (symbol/word/sentence).	Introduction of fake findings and references into the literature.	INRA. AI (2025); McCarthy et al. (2024); Özer (2024); Spicer (2024)
Model collapse	Training loop on synthetic data and the formation of scientific sinkholes in the long term.	Loss of scientific creativity and data diversity; disconnection of science from reality.	Shao (2025a)
Opacity (black box)	Number of parameters exceeding human comprehension.	Inability to replicate experiments and methodological gaps.	Brenneis et al. (2024); Ghosh (2025)
Responsibility gap	Autonomization of the decision-making mechanism.	Lack of accountability in faulty medical or engineering decisions.	European Parliament (2020); Lawvs (2026)
Academic writing and language	- Inability to understand emotional expressions due to morphology specific to non-English languages. - Dominant language of AI leading to global cultural imperialism.	- Misevaluation and unfair acceptance or rejection. - Assimilation of non-English languages and cultures.	Doko (2026); Halverson & Cannon (2025); Shao (2025b)

viduals with low socioeconomic status (Bains, 2024; Joseph, 2025; Lawvs, 2026; Pazzanese, 2020; The National Institute of Standards and Technology [NIST], 2022).

For example, it has been found that OPTUM, a risk prediction algorithm widely used in the U.S. healthcare sector, assigns lower scores to Black patients' healthcare needs compared to White patients with the same medical condition. The reason for this is that the algorithm uses directly as a source without conducting an in-depth analysis of historical healthcare spending data. However, in reality, Black patients have historically faced discrimination, and due to the racial disadvantages they face, they have had less access to the healthcare system and consequently incurred lower expenses. In addition to the healthcare sector, algorithmic ethical and human rights violations caused by AI also exist in the fields of employment and human resources, as seen at Amazon (Bains, 2024; Joseph, 2025). An AI-powered hiring algorithm developed by Amazon automatically assigned lower scores to resumes containing the word "woman" due to the predominance of men in the past 10 years' resume dataset, and to resumes containing the word "Black" due to the predominance of White individuals, thereby filtering out Black and female candidates (Doko, 2026).

AI-driven algorithmic discrimination and inequalities are emerging not only in healthcare and employment but also in the justice system. For example, it has been observed that COMPAS, a risk assessment software used in judicial systems, acts in a biased manner, violating the principle of equal protection, and incorrectly influences whether defendants are released on bail or the length of their prison sentences. Independent studies have revealed that this software labels Black defendants as high-risk twice as often as White defendants even though they do not actually pose a higher risk of reoffending. This directly violates human rights such as the presumption of innocence and equality during the trial phase (Devendorf, 2025; Orr, 2022).

Furthermore, the protection of algorithms as trade secrets leads to violations of transparency and due process, making it impossible for defendants and their attorneys to challenge the basis of adverse rulings. This situation conflicts not only with human rights but also with constitutional rights (Devendorf, 2025; Lawvs, 2026). While data is used to train algorithms by infringing upon the human rights and intellectual property rights of individuals or legal entities, the inability to scrutinize the operational structure of these algorithms and their defense as trade secrets highlights a significant paradox and injustice.

On the other hand, AI systems designed for facial recognition and biometric tracking also lead to human rights violations based on racial inequality and discrimination due to facial recognition errors. AI-powered facial recognition technologies and biometric tracking systems violate surveillance, privacy, and freedom of expression, thereby eliminating the right to remain anonymous in public spaces (Ramachandran, 2026; UN Global Compact Network Germany, 2023). A study conducted by the National Institute of Standards and Technology (NIST) has shown that various facial recognition algorithms produce more false matches on Asian and Black faces compared to White faces. These errors can lead to judicial errors, ranging from the wrongful detention of innocent individuals to their conviction. Most importantly, generative AI's potential to produce manipulative content and make erroneous identifications based on biases undermines individuals' right to accurate informa-

Table 3. Risks of artificial intelligence from a human rights perspective

Risk category	Data and statistical findings	Related resources
AI Hallucination Rates	0.7% to 1.5% in guided summarization tasks; 33% to 51% in complex reasoning tasks.	Graffius (2026); Romano & Gaskins (2025)
Facial recognition error margin (demographic)	Error rates up to 34% for dark-skinned females; near-perfect accuracy for light-skinned males.	Bains (2024)
Algorithmic bias in healthcare (ımpact)	The Optum algorithm systematically scored the needs of Black patients lower than white patients due to cost-based data utilization.	Bains (2024); Joseph (2025)
Financial access and credit bias	Algorithmic credit evaluations showed an 80% higher tendency to reject Black Americans compared to white applicants.	Bains (2024)
Academic source reliability	Out of 178 references generated by ChatGPT, 69 were found to have faulty DOIs or were entirely fictional.	Özer (2024)
Discrimination in employment (case)	Amazon's HR algorithm systematically eliminated female candidates because it was trained on 10 years of male-dominated data.	Bains (2024)
Algorithmic prediction in law (error)	The COMPAS system labeled Black defendants as "high risk for recidivism" at a 77% higher rate than white defendants.	Orr (2022)
Enterprise AI usage and decision making	As of 2024, 47% of enterprise AI users reported making a strategic decision based on hallucinated data at least once.	Romano & Gaskins (2025)

tion and fosters technopsychological self-censorship due to fears of digital surveillance (Bains, 2024; Cheong et al., 2025; Quinn Emanuel, 2025; U.S. Bureau of Cyberspace and Digital Policy, 2024). Due to its architecture that fosters such algorithmic discrimination, the use of AI in social science research or article evaluation processes is believed to potentially produce harmful outcomes that could contribute to the spread of algorithmic discrimination and the contamination of the scientific data pool. The main human rights issues generated by AI are presented in Table 3.

6. SCIENTIFIC AND ETHICAL RISKS ASSOCIATED WITH THE EPISTEMOLOGICAL STRUCTURE OF ARTIFICIAL INTELLIGENCE

The rapid spread of artificial intelligence (AI) across various fields brings with it serious breaches of scientific ethics and human rights risks, such as algorithmic bias, the systematisation of historical biases within datasets, and a lack of transparency in decision-making processes (Noble, 2018). Although these technological systems have the potential to reproduce social inequalities under the guise of technical objectivity, these ethical crises are not merely the result of a software design flaw, but rather may stem from the narrow methodological framework upon which AI is built. It must be questioned whether AI's fundamental inability to produce ethical and fair outcomes stems from its lack of human capacities such as the 'wisdom' to grasp the social context and a subjective 'phenomenological mind'. In this context, to understand the root of the problem, it is essential to focus on the epistemological inconsistencies and philosophical limitations of the positivist approach in the social sciences, which underpin the ethic and human rights violations created by algorithms.

When examining the epistemological distinction between data, knowledge, and wisdom, it becomes clear that the difference between data and knowledge is functional rather than structural. AI and information processing tools transform data into useful and functional knowledge, making it more accessible for human use. Even at the level of generative AI, it operates within this framework. Although AI possesses the power to generate knowledge from data, the process of addressing and interpreting this knowledge which encompasses the past, present, and future, along with diverse social sentiments and ideas within an interdisciplinary and multidisciplinary holistic framework, and transforming it into an effective value, is the product of a level of wisdom that requires certain human-specific abilities. The inability of AI and information processing systems to attain this level of wisdom is a natural limitation of AI (Ackoff, 1999; Dreyfus, 1972). In other words, wisdom is the ability to subject the disciplines to a dialectical filter in a manner consistent with their ontological nature, while taking into account their unique conditions.

At the same time, wisdom is a process of enlightened living that expresses the balance of being able to adjust one's behavior to the situation and develop different attitudes and behaviors according to varying circumstances. AI does not possess the psychological and subjective "phenomenological mind" required by this wisdom one that truly understands, feels, and possesses emotionality regarding what it is doing. AI also lacks the "qualia" consciousness the subjective and experiential knowledge of sensations such as the feeling of pain or the smell of a swamp. AI does not possess the "intentionality" characteristic inherent in all conscious psychic and mental actions (Doko, 2026; Nagel, 1974). As an information technology, AI excels at analyzing massive datasets and generating relevant information for a specific task, context, or application as instructed. AI lacks certain unique abilities related to judgment and creativity that the human brain naturally possesses (Acemoğlu, 2026; Searle, 1980).

Wisdom is a unique endowment and ability specific to humans thinking, sentient beings though it is not present in every individual. Science must approach the world, events, the individual, and society with a wise analytical perspective (Dale, 2018). However, the analytical methods of the natural sciences and the social sciences differ theoretically from one another. In the natural sciences, the Aristotelian inductive analytical approach which proceeds from micro-scale particulars focused on identifying and examining the components of a thing to arrive at universal knowledge is fundamental. However, this method of acquiring knowledge is not scientific in and of itself. In the social sciences, in addition to this, there is a need for a broad, interdisciplinary, and deductive analytical method that includes macro-scale and abstract concepts, analyzing which whole a thing or event is a part of. For scientific validity, it must also be explained through a deductive form of reasoning moving from the general to the particular. Necessity and certainty are possible as a result of verification conducted through a deductive perspective (Ackoff, 1999; Anlı, 2011; Cansen, 2025). AI, however, although produced through an interdisciplinary effort, is a product of the current scientific atmosphere dominated by the positivist approach.

Unlike the natural sciences, the social sciences require certain unique research methods such as "sociological imagination" that transcend the positivist approach. The mechanism of sociological imagination, which leads to the necessary wisdom in the social sciences, is a uniquely human trait. The mechanism of sociological imagination is one that can establish interdisciplinary connections between an individual's seemingly private/personal problems and significant social issues. The absence of a wise social science perspective one that acknowledges, feels, and perceives the complexity, dynamism, and relativity of the interrelationships among the social, personal, and historical dimensions of our lives will lead to errors that result in

flawed research findings and the resulting intractability of problems (Mills, 2022; Gadamer, 1975).

The wise approach, however, is one that employs both inductive and deductive methods together. This wise approach is evident in the social science methodology of Ibn Khaldun, who theorized the methodology that social sciences should have adopted centuries before the Enlightenment. Ibn Khaldun's scientific method for analyzing the social sciences does not dismiss the importance of theological and philosophical realities the primary factors influencing social developments and historical transformations. He believes that the principle of causality exists not only in the material-physical world but also in the historical and social world. He views society as a multifaceted phenomenon composed of individuals with diverse feelings, emotions, and motivations one that cannot be adequately examined using the precise, rule-based methods of the natural sciences. Ibn Khaldun developed an epistemological theory that does not reject the ontological integrity of the natural sciences (Haldun, 2015; Weber, 1949).

However, with the positivist approach of the Enlightenment era, particularly from the 18th century onward, the social sciences began to be viewed through the lens of the natural sciences, and information began to be analyzed from a natural scientific perspective (Baker, 1964). Drawing on their naturalist assumptions, positivist philosophers subjected the social sciences to a new restructuring within the framework of the natural sciences' methods and techniques, claiming that the research methodology of the natural sciences could be applied to social problems as well, and declaring that human morality is grounded in nature, all in the pursuit of building a materialist society (Polanyi, 1947; Horkheimer & Adorno, 2002). For example, Auguste Comte conceptualized sociology according to the materialist and positivist model of the natural sciences, particularly the physical model, and referred to this new form of sociology as "social physics" (Cabiri, 2018).

The fundamental principle of Durkheim's sociology, which followed Comte, also rests on the examination of social phenomena as if they were objects or materials. He argued that social phenomena could be analyzed with the same certainty as physical substances and physical and chemical processes in nature (Giddens, 2013). In positivist research methodology, knowledge is produced or analyzed through procedures such as inferential statistics, hypothesis testing, mathematical analysis, and experimental and quasi-experimental designs (Lee, 1991). A social science analytical method is employed that undermines rationality grounded in philosophy, ethics, aesthetics, and theology, thereby establishing rationality based on disciplines such as astronomy, mathematics, or geometry as dominant (Polanyi, 1977).

However, due to the limitations and lack of meaning this perspective created in the social sciences, objections to the positivist approach to the social sciences began to emerge in the mid-20th century. Objections that the method of treating the social sciences in a materialist and positivist manner, akin to the natural sciences, would fail to accurately analyze the causes and effects of events and thus deviate from scientific validity are widely accepted within social science circles (Mills, 2022; Gadamer, 1975). Today, the dominant scientific paradigm is the positivist paradigm. Consequently, current AI algorithms are inevitably developed and trained using a positivist approach. This situation renders the scientific validity of AI-generated research and findings in the social sciences risky and unreliable. Similarly, the practice of evaluating works in the social sciences solely through AI-based screening during the peer-review process carries the risk of leading to erroneous decisions.

7. CONCLUSION

AI conflicts with human and social values, creating human rights and ethical issues. To prevent the destructive effects of AI on scientific ethics and human rights, an AI architecture and philosophy centered on humans and society rather than technology and science must be adopted (Digital Cooperation Organization [DCO] 2025; Deutsche UNESCO, 2023; Syracuse University, 2026). By integrating an "ethical constitution" comprising a list of specific principles into the training and development processes of AI models, and ensuring the model can self-critique and self-correct according to these principles, a quality assurance system grounded in a sustainable self-regulation and total quality management philosophy will be established (Anthropic, 2026; Brenndoerfer, 2025; Uplatz, 2025). Constitutional AI principles developed recently by companies such as Anthropic stand out as exemplary applications in this regard. However, the universality of these ethical constitution efforts which are undertaken by only some employees of certain companies and the potential paradoxes they may harbor should be discussed.

It is not sufficient for companies to rely solely on their own internal audits or to establish their own ethical and legal frameworks. Ethical Impact Assessments (EIA) must be mandated as part of external and independent algorithmic audits conducted by public authorities and independent ethics committees (Deutsche UNESCO, 2023; Pham, 2025; United Nations Educational, Scientific and Cultural Organization [UNESCO]. 2022). It must not be forgotten that AI producers are commercial and profit-driven organizations and will not wish to restrict themselves with standards and rules that hinder the maximization of their profits. Of course, every effort to establish an ethical framework for AI is an important and valuable step. However, during the process of creating legal and ethical frameworks prepared by such commercially oriented companies, one must be cautious of the potential for these efforts to conceal, postpone, or create new

paradoxes regarding the primary risks involved.

Data must be filtered using algorithms that identify discriminatory and biased data which undermine objectivity and equality in the training datasets used during the AI training process. If there are existing data gaps that are discriminatory or inconsistent with equality, these should be filled fairly using data from minority groups within the framework of ethical representation (Beheshti & Kerridge, 2025; Pham, 2025; Shao, 2025b). Furthermore, the harms of attacks known as "Adversarial Machine Learning" (AML) another source of these issues must be prevented before they impact society. To prevent this, it is necessary to understand the capabilities of attackers, the model or system features they might attempt to exploit to achieve their objectives, and the design of attack methods that exploit security vulnerabilities during the development, training, and deployment phases of the machine learning lifecycle, as well as to establish a security framework compliant with international standards (Vassilev et al., 2025). To ensure AI models align with ethical and human rights values, an "AI Quality Security System" integrated with other quality management systems within the framework of Total Quality Management, covering design, production, and usage phases, and possessing international acceptability must be established and made a legal requirement.

To ensure the effectiveness and sustainability of the AI Quality Assurance System, legal gaps in the AI field must be addressed in light of the principles and legislative efforts established by organizations such as the EU, UNESCO, and the OECD, as well as sector-specific regulations like the Anthropic Act. For example, the EU AI Act, which classifies AI systems into unacceptable, high, medium, and low risk levels, imposes strict standards on high-risk applications such as biometric surveillance (European Commission, 2024; European Parliament [EP] 2020; Lawvs, 2026; Future of Life Institute, 2026; UN Global Compact Network Germany, 2023). By integrating such approaches from the EU AI Act with the principles of other organizations, a globally unified set of standards should be established at the international level.

The algorithms of AI systems can create hidden or invisible discrimination that appears non-discriminatory but has a disproportionate and negative impact on members of legally protected groups particularly in socio-economic and cultural domains such as employment, housing, or education. Legal systems should be structured within the framework of the Disparate Impact Discrimination Doctrine, which allows for challenges against algorithms containing such hidden discrimination, holding institutions accountable when an algorithm results in disproportionate harm even if it does not intentionally discriminate (Bains, 2024; Quinn Emanuel, 2025; U.S. Library of Congress, 2026). In accordance with the right to transparency and the right to request an explanation, individuals must be legally granted the right to understand the rationale behind a decision made by an algorithm and to challenge that decision (Dahlgren Lindström et al., 2025; Lawvs, 2026; Radanliev, 2025).

AI ethics should be recognized as a socio-technological discipline that requires the participation of all segments of society not just engineers or lawyers and its sustainability and oversight must be ensured through academic and bureaucratic efforts. The "Responsible AI" approach which prioritizes ethical values, security, transparency, and user rights throughout the design, development, implementation, and use of AI must be adopted. Responsible AI must be built in accordance with laws and institutional regulations, as well as ethical principles such as impartiality, transparency, accountability, robustness, and reliability, and must not cause harm to individuals or society (Karaman and Eyüpoğlu, 2025; Martin-Luther-Universität Halle-Wittenberg, 2026). In accordance with this approach, ethics courses should be incorporated into engineering and data science curricula. AI developers, engineers, and academics must possess the competence to recognize the negative societal consequences of the code they write and the AI products they develop (Center for Responsible Digitalization Office [ZEVEDI], 2022; Deutsche UNESCO, 2021; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2022). Furthermore, in critical fields such as education, healthcare, the judiciary, and the workplace, the final decision must always be made by a human expert that is, the principle of human control and oversight must remain sustainable and indispensable (Radanliev, 2025; Syracuse University, 2026; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2022).

On the other hand, while some journals have banned the delegation of the peer review process for an article to AI tools and others have imposed a disclosure requirement, it cannot be completely prevented. Due to the unique nuances of the peer review process, it must be conducted by scientists rather than AI. Due to concerns regarding confidentiality specifically that AI could infringe upon the author's exclusive rights to the article's content and potentially lead to unfair evaluations reviewers should not be permitted to upload manuscripts to AI tools for evaluation. Journals must establish clear guidelines on this matter (COPE, 2026). After all, it is never possible for AI to reliably build the same level of insight as humans (Acemoğlu, 2026). While AI can be used as an auxiliary tool in research within the framework of ethical principles, it must not assume a status that replaces humans by depriving them of their fundamental right to work (Çılğın, 2025).

The negative aspects of artificial intelligence such as sycophancy, flattery, and hypocrisy which violate scientific ethics (Chandra et al., 2026; Cheng et al., 2026; Hackenburg et al., 2026), will inflict significant ontological harm on the epistemology of scientific knowledge worldwide if

they replace the objective perspective essential to scientific research. It must not be forgotten that artificial intelligence models are a product of neoliberalizing humanity and science. Nor should it be forgotten that, at its core, artificial intelligence is a commercial product. In today's world, shaped by neoliberalism and populated by narcissistic, egoistic, and self-centered individuals, the fact that artificial intelligence produced by science funded by neoliberal capital operates according to the typologies of these people, viewing them not as researchers but as customers to be retained, stands before us as a serious problem that must be questioned. To mitigate the potential collapse of AI models, the preservation of conventional learning methods and tools based on books and articles, the establishment of national AI systems by countries, and the maintenance of the sustainability of national conventional and AI systems in parallel with global AI institutions should be kept in effect as a precautionary measure and policy against risks such as information poisoning and digital reset.

Ethics: There are no ethical issues with the publication of this manuscript.

Authors' Contributions: The authors contributed to the study equally.

Conflict of Interest: The author declares that he has no conflict of interest.

Financial Disclosure: The authors declared that this study has received no financial support.

Statement on the Use of Artificial Intelligence: Artificial intelligence was not used in the preparation of the article.

Peer-review: Externally peer-reviewed.

REFERENCES

Acemoğlu, D. (2026). AI is not improving productivity: Nobel laureate Daron Acemoglu. *MIT Sloan Management Review*, https://sloanreview.mit.edu/audio/ai-is-not-improving-productivity-nobel-laureate-daron-acemoglu/ Accessed on Feb 24, 2026.

Ackoff, R. L. (1999). *Ackoff's best* (pp. 170–172). John Wiley & Sons.

Akpınar, Ü. Z., Araç, M. & Doğru, E. (2025). The ethical imperative of artificial intelligence. *Journal of Sustainable Education Studies*, (S4), 51–58. [Turkish]

Anlı, Ö. F. (2011). Aristoteles ve yöntem: Tümevarım ve tümdengelim. *Bilim ve Ütopya*, 208. 63–69.

Anthropic, (2026). *Claude's new constitution.* https://www.anthropic.com/news/claude-new-constitution Accessed on Jan 24, 2026.

Aristotle, (2022). *The Nicomachean Ethics*. (F.H. Peters, Trans.), Ryerson University. https://pressbooks.library.torontomu.ca/nicomacheanethics/ Accessed on May 15, 2026.

Bains, C. (2024). The legal doctrine that will be key to preventing AI discrimination. *Brookings*. https://www.brookings.edu/articles/the-legal-doctrine-that-will-be-key-to-preventing-ai-discrimination/ Accessed on Oct 22, 2025

Baker, K. M. (1964). The early history of the term 'social science'. *Annals of Science*, *20*(3), 211-226. [CrossRef]

Beheshti, A. & Kerridge, I. (2025). Understanding the artificial ıntelligence revolution and its ethical ımplications. *Journal of Bioethical Inquiry*, *22*(3), 497–505. [CrossRef]

Boyes, B. (2025). Case studies in the unethical and irresponsible use of AI. *RealKM*. https://realkm.com/2025/11/05/case-studies-in-the-unethical-and-irresponsible-use-of-ai/ Accessed on Nov 10, 2025.

Brenneis, A., Gehring, P. & Lamadé, A. (2024). Zwischen fachlichen standards und wilder innovation: Zur begutachtung von big data- und KI-Projekten in forschungs-ethikkommissionen. *Ethik in der Medizin*, *36,* 473–491. [German] [CrossRef]

Brenndoerfer, M. (2025). Constitutional AI: Principle-Based alignment through self-critique. https://mbrenndoerfer.com/writing/constitutional-ai-principle-based-alignment-through-self-critique Accessed on: Nov 17, 2025.

Cabanac G. (2024). Chain retraction: how to stop bad science propagating through the literatüre. *Nature*. https://www.nature.com/articles/d41586-024-02747-1#ref-CR1 Accessed on Sep 09, 2025.

Cabiri, M. A. (2018). *İbn Haldun'un düşüncesi asabiyet ve devlet* (Çev: M. Çelik,). Mana Yayınları.

Cansen, E. (2025). Yaşamasaydım inanmazdım. yatırım finansman. *YouTube*. https://www.youtube.com/watch?v=oMobB3xFaEg&t=458s Accessed on Jan 11, 2026.

Center for Responsible Digitalization Office [ZEVEDI] (2022). Zur forschungsethischen begutachtung von KI-Forschungsprojekten. Handreichung zur unterstützung der arbeit von ethikkkommission. *Technical University of Darmstadt*, https://zevedi.de/wp-content/uploads/2022/11/ZEVEDI_Handreichung-KI-Forschungsethik_2022.pdf Accessed on Oct 16, 2025.

Chandra, K., Kleiman-Weiner, M., Ragan-Kelley, J. & Tenenbaum J. B. (2026). Sycophantic chatbots cause delusional spiraling, even in, deal bayesians. *Arxiv*. doi: 10.48550/arXiv.2602.19141

Cheong, I., Caliskan, A., & Kohno, T. (2025). Safeguarding human values: rethinking US law for generative AI's societal impacts. *AI And Ethics*, *5*(2), 1433-1459. [CrossRef]

Chinoracky, R. & Stalmasekova, N. (2025). Ethical problems in the use of artificial intelligence by university educators. *Education Sciences*, *15*(10), 1322. [CrossRef]

Cheng, M., Lee, C., Khadpe, P., Yu, S., Han D. & Jurafsky D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*, *391*(6792). [CrossRef]

Committee on Publication Ethics [COPE], (2025). *Handling an article produced by AI.* https://publicationethics.org/guidance/case/handling-article-produced-ai Accessed on Jan 03, 2026.

Committee on Publication Ethics [COPE], (2023). *Authorship and AI tools.* https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools (Accessed on Dec 30, 2025.

Committee on Publication Ethics [COPE], (2026). *Editors suspect that reviewers are using artificial intelligence.* https://publicationethics.org/guidance/case/editors-suspect-reviewers-are-using-artificial-intelligence Accessed on Mar 02, 2026.

Cunningham, S. (2026). Claude code 27: Research and publishing are now two different things. *Scott's Mixtape Substack*. https://causalinf.substack.com/p/claude-code-27-research-and-publishing Accessed on Mar 03, 2026.

Çılğın, T. (2025). Dialectical analysis of problems created by technological developments in the labour market. *Journal of Social Policy Conferences*, (89), 106–128. [CrossRef]

Dahlgren Lindström, A., Methnani, L., Krause, L., Ericson, P., de Rituerto de Troya, Í. M., Coelho Mollo, D., & Dobbe, R. (2025). Helpful, harmless, honest? Sociotechnical limits of AI alignment and safety through Reinforcement Learning from Human Feedback. *Ethics and Information Technology*, *27*(2), 28. [CrossRef]

Dale, S. F. (2018). *İbn Haldun ve İnsan Bilimi* (Çev. A. Ay, & C. Coşkun). Say Yayınları.

Deutsche UNESCO (2021). Der Künstlichen Intelligenz: Weg weiser für die Gestaltung unserer Zukunft, *Deutsche UNESCO-Kommission*, https://www.unesco.de/assets/dokumente/Deutsche_UNESCO-Kommission/02_Publikationen/Publikation_Zusammenfassung_der_UNESCO-Empfehlung_zur_Ethik_der_K%C3%BCnstlichen_Intelligenz_%E2%80%93_Wegweiser_f%C3%BCr_die_Gestaltung_unserer_Zukunft.pdf Accessed on January 05, 2026.

Deutsche UNESCO (2023). Empfehlung zur ethik künstlicher ıntelligenz. *DeutscherBundestag*, https://www.bundestag.de/resource/blob/942696/993ee22a6d04a87778e06f3c66a49dc8/20-18-107a-Dt_Unesco.pdf Accessed on October 22, 2025. [German]

Devendorf, J. (2025). Avoiding Algorithmic Bias: Top 5 AI Liability Issues in Courts. *Super Lawyers*, https://www.superlawyers.com/resources/science-and-technology-law/avoiding-algorithmic-bias-top-5-ai-liability-issues-in-courts/ Accessed on December 05, 2025.

Digital Cooperation Organization [DCO] (2025). *Rights by Design Embedding Human Rights Principles on AI systems.* https://dco.org/wp-content/uploads/2025/06/Rights-by-Design-Embedding-Human-Rights-Principles-on-AI-systems.pdf Accessed on January 08, 2026.

Doko, E. (2026). Yapay zekâ bilgelik sahibi olabilir mi? Turing Kültür Sanat. *You Tube*. https://www.youtube.com/watch?v=cqb4P3rJ5q0 Accessed on March 02, 2026.

Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. Harper & Row Publishers.

European Commission [EC], (2024). *AI and Ethics, Human Rights, Law and Educational Data. Briefing* (report No. 6). European Digital Education Hub's squad on artificial intelligence in education. https://www.ai4t.eu/wp-content/uploads/2024/01/AI-squad-output_briefing-report-6.pdf Accessed on December 05, 2025.

European Commission [EC], (2025). *Living guidelines on the responsible use of generative AI in research, ERA Forum Stakeholders document.* https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf Accessed on December 16, 2025.

European Parliament [EP], (2020). *The ethics of artificial intelligence: Issues and initiatives, Panel for the Future of Science and Technology*. European Parliamentary Research Service, Scientific Foresight Unit (STOA), https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf Accessed on October 26, 2025.

Future of Life Institute, (2026). *The EU Artificial Intelligence* Act. https://artificialintelligenceact.eu/ Accessed on February 04, 2026.

Gabriel, M. (2020). Wert und Mensch in der Ethik der. In KI, D. Zimmer (Ed.), *Regulierung für algorithmen und künstliche intelligenz* (pp. 25–42). Nomos Verlagsgesellschaft mbH & Co. KG, Baden-Baden. [German] [CrossRef]

Gadamer, H-G. (1975). *Truth and method*. Continuum. https://web.mit.edu/kaclark/www/gadamer_truth_and_method.pdf Accessed on May 15, 2026.

Ghosh, M. (2025). Artificial intelligence (AI) and ethical concerns: a review and research agenda. *Cogent Business & Management*, *12*(1), 255180. [CrossRef]

Giddens, A. (2013). *Sosyoloji*, Kırmızı Yayınları. [Turkish]

Graffius, S. M. (2026). *Are AI hallucinations getting better or worse?* We analyzed the data. *ScottGraffius*. https://www.scottgraffius.com/blog/files/ai-hallucinations-2026.html

Hackenburg, K., Tappin, B. M., Hewitt, L., Saunders, E., Black, S., Lin, H., Fist C., Margetts, H., Rand D. G. & Summerfield C. (2025). The levers of political persuasion with conversational artificial intelligence. *Science*, *390*(6777), eaea3884. [CrossRef]

Haldun, İ. (2015). *Mukaddime 1-2*, (Çev: S. Uludağ). Dergâh Yayınları.

Halverson, K., & Cannon, M. (2025). AI in scholarly publishing: Positives and pitfalls. *The ASHA Leader*. https://leader.pubs.asha.org/do/10.1044/2025-1001-ai-scholarly-pubs-academia/full/ Accessed on November 07, 2025.

He, Y. & Bu, Y. (2025). Academic journals' AI policies fail to curb the surge in AI-assisted academic writing. *arXiv*. https://doi.org/10.48550/arXiv.2512.06705 /

Horkheimer, M., & Adorno, T. W. (2002). *Dialectic of enlightenment*. Stanford University Press.

INRA. AI. (2025). AI Hallucinations in Research: What They Are and How to Prevent Them, https://www.inra.ai/blog/ai-hallucinations Accessed on November 16, 2025.

International Committee of Medical Journal Editors [IJMJE], (2026). *Artificial intelligence AI assisted technology. Defining the role of authors and contributors.* https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html#four Accessed on March 04, 2026.

Joseph J. (2025). Algorithmic bias in public health AI: a silent threat to equity in low-resource settings. *Frontiers in Public Health*, *13,* 1643180. [CrossRef]

Kalite Akademik Tercüme, (2026). *Yapay zekâ kullanımı nedeniyle reddedilen ve yayından kaldırılan makaleler artıyor mu*? https://kaliteakademik.com/yapay-zeka-kullanimi-nedeniyle-reddedilen-ve-yayindan-kaldirilan-makaleler-artiyor-mu.html Accessed on March 01, 2026.

Karaman, F. ve Eyüpoğlu, C. (2025). Artificial intelligence ethics in various fields: A review. *İstanbul Ticaret Üniversitesi Teknoloji ve Uygulamalı Bilimler Dergisi*, *8*(2). 87–104. [CrossRef]

Küçükvardar, M., Aslan, A., Bayrakcı, S. (2020). Yapay zekâ ve etik üzerine bir araştırma. *Atlas Journal*, *6*(36), 1065–1077. [CrossRef]

Lawvs. (2026). Legal ımplications of algorithmic bias in decision-making. *Lawvs.com*, https://lawvs.com/articles/legal-implications-of-algorithmic-bias-in-decision-making

Lee, A. S. (1991). Integrating positivist and ınterpretive approaches to organizational research. *Organization Science*, *2*(4), 342–365. [CrossRef]

McCarthy, I. P., Hannigan, T. R. and Spicer, A. (2024). The risks of botshit. *Havard Business Review*, preprint. doi: 10.2139/ssrn.5172718

Martin-Luther-Universität (2026). KI und ethik. deutschdidaktik. *Halle-Wittenberg* https://deutschdidaktik.germanistik.uni-halle.de/6-ki-und-ethik/ Accessed on February 26, 2026.

Merton, R. (1973). The Normative Structure of Science. In N. Storer (Ed.), *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press. https://sites.temple.edu/stsnetwork/files/2022/01/The-Normative-Structure-of-Science.pdf

Mills, C. W. (2022). *Sosyolojik Tahayyül*. (Çev. Ö. Küçük), Hil Yayınları.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4). https://www.sas.upenn.edu/~cavitch/pdf-library/Nagel_Bat.pdf [CrossRef]

National Institute of Standarts and Technology [NIST] (2022). Towards a standart for identifying and managing bias in artificial intelligence. *NIST Special Publication 1270*. [CrossRef]

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press. https://files.commons.gc.cuny.edu/wp-content/blogs.dir/6105/files/2019/01/SAFIYA-NOBLE.pdf [CrossRef]

Nordquist, R. (2019). Holistic Grading (Composition). Glossary of Grammatical and Rhetorical Terms. *ThoughtCo*. https://www.thoughtco.com/holistic-grading-composition-1690838 Accessed on November 22, 2025.

Notman, N. (2025). AI tools combat paper mill fraud in scientific publishing as peer review system struggles. *Royal Society of Chemistry*. https://www.chemistryworld.com/features/ai-tools-tackle-paper-mill-fraud-overwhelming-peer-review/4022253.article Accessed on October 23, 2025.

Oellig, V. T. (2024). Künstlich intelligent, ethisch dumm. *Amnesty International*, https://www.amnesty.de/informieren/amnesty-journal/kuenstliche-intelligenz-menschenrechte-meredith-whittaker-kuenstlich-intelligent-ethisch-dumm Accessed on October 27, 2025.

Orr, D. (2022). The unintended consequences of algorithmic bias. *Congressional Black Caucus Foundation [CBCF]*. https://www.cbcfinc.org/wp-content/uploads/2022/04/2022_CBCF_CPAR_TheUnintendedConsequencesofAlgorithmicBias_Final.pdf Accessed on November 30, 2025.

Özer M. (2024). Is Artificial Intelligence hallucinating? *Turkish Journal of Psychiatry*, *35*(4), 333-335. [CrossRef]

Pazzanese, C. (2020). Ethical concerns mount as AI takes bigger decision-making role. *The Harvard Gazette*, https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/ Accessed on December 19, 2025.

Pham T. (2025). Ethical and legal considerations in healthcare AI: innovation and policy for safe and fair use. *Royal Society Open Science*, *12*(5), 241873. [CrossRef]

Polanyi, K. (1947). Our obsolete market mentality. *Commentary*, *3*(2), 109–117. https://www.scribd.com/document/358169609/POLANYI-Karl-Our-Obsolete-Market-Mentality

Polanyi, K. (1977). The economistic fallacy. *Review*, *1*(1), 9–18.

Quinn, E. (2025). Lead article: When machines discriminate: The rise of AI bias lawsuits. https://www.quinnemanuel.com/the-firm/publications/when-machines-discriminate-the-rise-of-ai-bias-lawsuits/ Accessed on January 09, 2026.

Radanliev, P. (2025). AI ethics: Integrating transparency, fairness, and privacy in AI development. *Applied Artificial Intelligence*, *39*(1). [CrossRef]

Ramachandran, M. (2026). Case studies in AI ethics. In: *Engineering ethics of AI by design*. Springer. [CrossRef] / https://doi.org/10.1007/978-981-95-2909-4_5

Resnik, D. B. & Hosseini, M. (2025). The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. *AI and Ethics*, *5*(2), 1499-1521. [CrossRef]

Romano, D. and Gaskins, C. (2025). The reality of AI hallucinations in 2025. *Drainpipe.io*. https://drainpipe.io/the-reality-of-ai-hallucinations-in-2025/ Accessed on December 29, 2025.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3),

Shao, A. (2025a). New sources of inaccuracy? A conceptual framework for studying AI hallucinations. *Misinformation Review*. https://misinforeview.hks.harvard.edu/article/new-sources-of-inaccuracy-a-conceptual-framework-for-studying-ai-hallucinations/ Accessed on January 23, 2026. [CrossRef]

Shao, A. (2025b). Beyond misinformation: a conceptual framework for studying AI hallucinations in (Science) communication. *Cornell University*, arXiv, https://doi.org/10.48550/arXiv.2504.13777 / [CrossRef]

Spicer A. (2024). Beware the 'botshit': why generative AI is such a real and imminent threat to the way we live. *The Guardian*. https://www.theguardian.com/commentisfree/2024/jan/03/botshit-generative-ai-imminent-threat-democracy Accessed on January 30, 2026.

Spinoza, B. (2011). *Etika*. Çev. Hilmi Ziya Ülken. Dost Kitabevi Yayınları.

Syracuse University (2026). What is responsible AI: Principles, frameworks & future. *Syracuse University School of Information Studies*. https://ischool.syracuse.edu/what-is-responsible-ai/ Accessed on January 24, 2026.

The Retraction Watch Database, (2026*). Version: 1.0.8.0. ISSN 2692-465X*. https://retractiondatabase.org/RetractionSearch.aspx? Accessed on February 04, 2026.

UN Global Compact Netzwerk Deutschland (2023). *Künstliche intelligenz und menschenrechte*. https://www.globalcompact.de/fileadmin/user_upload/kuenstliche_intelligenz_und_menschenrechte_Publikation.pdf Accessed on November 04, 2025.

UNESCO (2023). Artificial Intelligence: UNESCO calls on all governments to implement global ethical framework without delay. https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without

United Nations Educational, Scientific and Cultural Organization [UNESCO]. (2022). Ethics of artificial intelligence-AI. *UNESCO*, https://www.unesco.org/en/artificial-intelligence/recommendation-ethics Accessed on November 30, 2025.

US Bureau of Cyberspace and Digital Policy (2024). Risk management profile for AI and human rights. *US Department of State*. https://2021-2025.state.gov/risk-management-profile-for-ai-and-human-rights/ Accessed on December 16, 2025.

US Library of Congress (2026). What is disparate-ımpact discrimination? *US Congress*. https://www.congress.gov/crs-product/IF13057 Accessed on March 04, 2026.

Uplatz, (2025). The evolution of aı alignment: A comprehensive analysis of RLHF and constitutional AI in the pursuit of ethical and scalable systems. *Uplatz Blog*. https://uplatz.com/blog/the-evolution-of-ai-alignment-a-comprehensive-analysis-of-rlhf-and-constitutional-ai-in-the-pursuit-of-ethical-and-scalable-systems/ Accessed on October 27, 2025.

Uzaktan Eğitim Kapısı [UEK] (2026). *Dijital okuryazarlık: Giriş. Dijital etik*. https://uzaktanegitimkapisi.gov.tr/Egitimler/video?Uniq=C617B020-6529-40DC-90CE-74CB5B153426 Accessed on March 01, 2026.