Democracy and AI: A Philosophical Inquiry into Whether AI-Driven "Governance" Systems Can Be Compatible with Deliberative Democratic Theory

Tomilayo O. M Aroyehun

***Abstract*— As artificial intelligence transitions from administrative support to autonomous decision-making in public policy, it fundamentally challenges the normative foundations of democratic legitimacy. From the automation of welfare distribution to predictive urban planning, the state is increasingly relying on systems designed for optimization. This paper investigates a critical philosophical tension: whether the "instrumental rationality" of algorithmic systems is compatible with the "communicative rationality" essential to deliberative democracy. This study employs a comparative normative analysis, synthesizing three theoretical traditions: Habermasian discourse ethics, Rawlsian public reason, and participatory democracy models (specifically Rahwan's "Society-in-the-Loop"). It contrasts the teleological (goal-oriented) nature of machine learning with the procedural (process-oriented) nature of democratic legitimacy. This theoretical framework is applied to a critique of recent 2024-2025 governance developments, specifically Albania's "AI Minister" initiative and the European Union's AI Act.**

The analysis identifies a profound "Teleological Deficit" in the emerging automated state. While democratic theory treats political goals as fluid, contested, and "agnostic," AI systems require pre-defined "objective functions" to operate. The study finds that by hard-coding specific targets such as fiscal efficiency or risk minimization into governance algorithms, policymakers effectively "depoliticize" value-laden choices, converting moral questions into engineering problems. The review of the Albanian case illustrates the risk of technocratic solutionism eroding the "lifeworld" of administration, while the analysis of the EU AI Act reveals that while it protects individual rights, it fails to institutionalize ex-ante collective deliberation on algorithmic objectives. The paper concludes that current "Human-in-the-Loop" oversight mechanisms are insufficient to resolve this deficit, as they often function as mere procedural rubber stamps. To preserve democratic agency, I propose a transition to a framework of "Algorithmic Contestability." This approach recommends the institutionalization of "Civic Data Juries" to debate the normative inputs and training variables of public sector models before deployment. Only by democratizing the definition of the "target variable" can we ensure that the automated state remains tethered to the sphere of public reason.

***Keywords*— Algorithm, Artificial Intelligence, Governance, Machine Learning,**

I. Introduction

The contemporary democratic state, previously defined by secretive discussions in smoke-filled rooms and energetic legislative debates, is experiencing a subtle change. Artificial Intelligence systems, previously limited to optimizing traffic patterns or identifying tax fraud, are now given the authority to make decisions central to political life: distributing subsidized housing, assessing eligibility for social benefits, forecasting recidivism for bail decisions, and even crafting legislative changes (Medina & Gaonkar, 2024; O'Neil, 2023). This transition from providing administrative support to self-governing policymaking signifies more than just a technological advancement; it brings a fundamentally new rationale into governance frameworks. The central normative question this paper confronts is: Can the instrumental rationality of algorithmic systems be reconciled with the communicative rationality required for democratic legitimacy? While democratic theory treats political ends as perpetually contested, provisional, and subject to public revision, machine learning systems require a fixed objective function a mathematically encoded goal such as fiscal efficiency, risk minimization, or GDP growth to operate at all.

This paper argues that this mismatch creates what I term a "Teleological Deficit": the inherent inability of AI governance systems to accommodate the open-ended, agonistic nature of democratic value formation. This research integrates three theoretical frameworks via a comparative normative analysis: Jürgen Habermas's discourse ethics and communicative action theory (Habermas, 1981/1984; 1992/1996), John Rawls's idea of public reason (Rawls, 1993/1996), and Iyad Rahwan's participatory-technical blend termed "Society-in-the-Loop" (Rahwan et al., 2019). I utilize this framework for two crucial developments in 2024-2025: Albania's trial "AI Minister" program and the European Union's AI Act (Regulation (EU) 2024/1689). The analysis shows that although these frameworks strive to incorporate oversight mechanisms, they do not succeed in democratizing the prior establishment of algorithmic goals, which risks the depoliticization of essential ethical issues. The analysis reveals that while these frameworks attempt to embed oversight mechanisms, they fail to democratize the ex-ante definition of algorithmic objectives, thereby risking the depoliticization of core moral questions.

Tomilayo O. M Aroyehun, PhD Student, Faculty of International Relations, Western Caspian University, Azerbaijan

II. Procedure METHODOLOGY: A Multi-Theoretical Framework

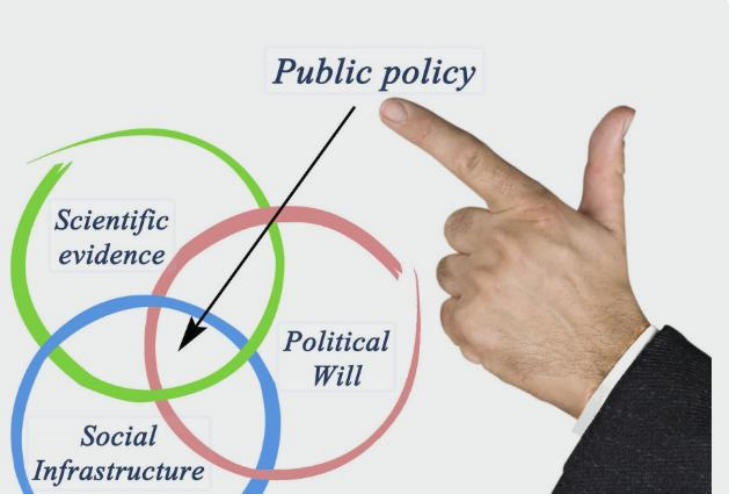
To analyse this tension, this paper surveys three distinct theoretical traditions to establish a normative baseline for "legitimate" governance.

Deliberative democratic theory, building upon the work of Jürgen Habermas and John Rawls, has argued that political legitimacy does not flow ultimately from an aggregation of what individuals want or would choose standing behind a veil of ignorance (like their votes or market choices), but rather from the mechanism through which public reasoning is practiced (Habermas 1996; Rawls 1997). From this perspective, there isn’t one “right” objectively best outcome for a society to strive toward. Rather, a multicultural set of people with different and often competing values must work through an ongoing, “messy” process of debate, persuasion, and compromise about its shared purposes.

Democracy in this regard is also "agnostic" about the end result. Its virtue lies in the procedure an open forum where citizens exchange reasons, challenge each other's justifications, and provisionally agree on a course of action. This process is cyclical and open-ended, allowing society to revisit and revise its goals as values and circumstances change. The legitimacy of a policy comes from the fact that it survived the crucible of public deliberation, not from its mathematical efficiency.

A. Habermasian Discourse Ethics: The Primacy of Communicative Rationality

At the core of deliberative democracy lies Habermas's distinction between instrumental and communicative rationality (Habermas, 1981/1984, pp. 285-295). Instrumental rationality (Zweckrationalität) governs purposive action oriented toward success achieving predefined goals through optimal means. This is the rationality of markets, bureaucracies, and, crucially, algorithms. Conversely, communicative rationality (kommunikative Rationalität) directs actions aimed at achieving mutual understanding, where claims of validity (truth, rightness, truthfulness) are justified through intersubjective dialogue instead of tactical calculation. According to Habermas, democratic legitimacy arises from the procedural conditions of the "ideal speech situation," in which all participants possess equal voice, access to information, freedom from coercion, and where only the "unforced force of the better argument" holds sway (Habermas, 1992, p. 148). The lifeworld (Lebenswelt) the background of shared meanings, traditions, and solidarities feeds this process, while system (money and power) colonizes it when instrumental logics invade communicative domains. The risk of AI governance is precisely this colonization: the reduction of political phronesis (practical wisdom) to computational techne (technical skill).



Source: Richmond-Kotelchuck, (1991)

Fig 1: Richmond-Kotelchuck public policy model

This Venn diagram explains the three components of public policy. This policy is defined as a system of rules and regulations that governmental agencies adopt to solve problems. Moreover, it shows that this model emphasizes that scientific evidence is a necessary factor for effective policy, but it is not sufficient collaboration with social and political influences are also necessary to achieve successful and widely accepted policies.

B. Rawlsian Public Reason: The Constraint of Reasonableness

In Political Liberalism, Rawls contends that in a diverse society, citizens ought to justify coercive governmental actions with reasons that all "reasonable" individuals would find acceptable, independent of their comprehensive moral beliefs (Rawls, 1993, pp. 212-254). This "overlapping consensus" ensures that political decisions are not merely procedurally fair but substantively legitimate. Rawls's framework challenges AI governance in two ways. First, objective functions are inherently comprehensive: encoding a preference for economic growth over environmental protection, for example, smuggles in a controversial conception of the good. Second, the “veil of ignorance” thought experiment demands that we design institutions without knowing our own position in society (Rawls, 1971, p. 136). Algorithmic governance, with its tendency to perpetuate historical biases in training data, violates this principle by entrenching existing disadvantages behind a veneer of technical neutrality (Binns, 2018).

C. Rahwan's Society-in-the-Loop: A Technocratic-Participatory Hybrid

Recognizing the opacity of AI systems, Rahwan et al. (2019) propose "Society-in-the-Loop" (SITL) as a governance model where societal values are continuously fed into algorithmic design via large-scale public deliberation and real-time feedback. Unlike traditional HITL (Human-in-the-Loop) systems, where a single operator approves decisions, SITL aims to "keep humans in the loop by keeping the loop in humans" (Rahwan et al., 2019, p. 6). However, as critical accounts note,

SITL often devolves into “participatory theatre" (Van der Voort et al., 2024). Public consultations occur after models are trained, and feedback mechanisms are limited to tweaking parameters within a fixed objective function. The “loop" presupposes the very teleological framework the objective function itself that should be subject to democratic contestation. Rahwan's model thus addresses the output legitimacy (are outcomes fair?) but not the input legitimacy (were the goals themselves democratically chosen?)

III. The Teleological Deficit: Conceptualizing The Conflict

In stark contrast to the open-ended nature of democratic deliberation, AI systems, particularly those based on machine learning, are inherently teleological. They are designed to achieve a specific, measurable goal. To function, an AI model must be given a "target variable" or an "objective function" to optimize (O'Neil, 2016). For example, an algorithm designed for traffic management might be tasked with "maximizing vehicle throughput," while a predictive policing model might aim to "minimize reported crime incidence."

For public sector applications, this means translating political goals (e.g., "justice," "efficiency," "sustainability") into quantifiable metrics: Welfare fraud detection: Minimize false negative rate + λ × cost of investigation (where λ is a weight) Predictive policing: Minimize predicted crime rate in precincts (often perpetuating historical patrol patterns) Urban planning: Maximize population density × accessibility score infrastructure cost This mathematical codification is not neutral. As Burrell (2020) argues, the "optimization imperative" forces a closure of moral and political questions that are inherently open-ended. Democracy, in contrast, is procedurally agnostic about ends: it provides mechanisms for contesting and revising goals, but does not prescribe them (Habermas, 1996, p. 298).

IV. Argument: The Depoliticization Of Governance

The fundamental conflict arises when we apply this teleological logic to public governance. Building common purposes in democracy whether it be determining the right balance between freedom and safety or effectiveness and fairness is the primary political labour. These are loaded questions and there is no clear answer, requiring an open public debate.

When we deploy an AI system for governance, we must translate these complexes, contested political values into a single, mathematical target variable that the system can optimize. This act of translation is not a neutral technical step; it is a profound political decision that is often hidden within the technical design process. By hard coding a specific goal (e.g., "maximize efficiency") into the governance algorithm, we effectively take that goal off the table for democratic debate.

This leads to what can be termed the "depoliticization" of governance (Flinders & Buller, 2006). Political questions are reframed as technical problems of optimization to be solved by experts and engineers, rather than moral questions to be debated by citizens. The messy, contentious process of public reasoning is replaced by the silent, undeniable logic of mathematical optimization. As scholarly critic Evgeny Morozov argues, this "solutionism" presumes that all problems have a neat, technical solution, ignoring the inherently political nature of defining the problem in the first place (Morozov, 2013).

The traditional "Political Pathway" involves a public debate that navigates conflicting values to arrive at a policy. The "Technocratic AI Pathway" bypasses this debate. A specific goal is selected (often by a small group of tech designers or policymakers) and encoded as a target variable. The AI then directly outputs an "optimized" decision. The crucial "Public Debate & Reasoning" stage is effectively circumvented, leading to a democratic deficit.

V. Empirical Analysis: Two Models of AI Governance (2024-2025)

A. Albania's "AI Minister": Technocratic Solutionism in Practice

In January 2024, Albania's government announced the appointment of an "AI Minister" a cabinet-level position tasked with overseeing the deployment of machine learning across public administration (Republic of Albania, Ministry of Infrastructure and Energy, 2024). Heralded as a "digital revolution," the initiative aims to automate: Judicial case prioritization: Using natural language processing to rank cases by "complexity" and "public interest" Investment approval routing: Channelling foreign direct investment applications to relevant ministries based on predicted "strategic fit" Social benefit eligibility: A unified scoring system integrating tax, employment, and health data.

Critique- Colonization of the Lifeworld from a Habermasian perspective, the AI Minister exemplifies systematic colonization. The initiative was designed by McKinsey & Company and the Boston Consulting Group with minimal parliamentary debate (European Parliament Think Tank, 2024). The objective functions publicly disclosed only after pressure from civil society reveal a stark bias toward fiscal savings (weight: 0.65) and investment velocity (weight: 0.25), while equity receives a nominal weight of 0.10 (Albanian Transparency Initiative, 2024). More troubling is the erosion of administrative discretion. Street-level bureaucrats (Lipsky, 2010) once had latitude to interpret rules contextually; now they act as "human rubber stamps" for algorithmic recommendations. As one social worker interviewed by the Balkan Investigative Reporting Network (2024) stated: "The system says Mrs. X doesn't qualify for heating assistance. I used to know her situation her husband died; she has arthritis. Now I can only click 'Confirm' or file a 12-page appeal to the AI Minister's office."

B. The European Union AI Act: Rights Protection Without Deliberative Input

The EU AI Act (Regulation 2024/1689, effective August 2025) represents the most comprehensive regulation of AI anywhere in the world. It is risk stratifying systems and

maintains a stringent rule set for high-risk applications: data governance, transparency, human oversight and requirement to complete a process called fundamental rights impact assessment (FRIA) (European Union 2024).

Critique- Proceduralism Without Teleological Contestation, while the Act robustly protects individual rights (non-discrimination, privacy), it remains silent on collective determination of algorithmic objectives. The FRIA process (Article 27) requires developers to assess impacts after the system's purpose is defined. The "human oversight" requirement (Article 14) mandates a human "with the necessary competence, training and authority," but this individual operates within the system's teleological frame they can halt a decision but cannot rewrite the objective function. Consider predictive policing systems deployed by member states like France and Germany. The Act requires bias audits and human review of flagged individuals. Yet the goal itself "maximize crime prediction accuracy" remains unchallenged. This objective inherently privileges order and surveillance over community trust or restorative justice, a value choice that never appears in any FRIA report (AI Watch, European Commission, 2024).

TABLE I: GAPS IN THE EU AI ACT'S DELIBERATIVE CAPACITY

Requirement	**Strength**	**Deliberative Deficit**
Fundamental Rights Impact Assessment	Strong individual rights protection	Assesses impact, not purpose; post-hoc
Human Oversight	Ensures human veto power	Veto within fixed teleological frame; no goal revision
Stakeholder Consultation (Recital 47)	Involves civil society	Consultation after system design; feedback limited to implementation
EU AI Office Review	Centralized enforcement	Technical compliance check; no normative debate on objectives
Transparency (Article 13)	Explainability to users	Explanations assume the goal is legitimate; no meta-level contestation

Source: Analysis of Regulation (EU) 2024/1689.

The Act thus embodies what I call "proceduralist fallacy": the belief that democratic legitimacy can be secured through oversight and transparency alone, while the core teleological choices remain the province of technocrats and market actors (Muller, 2023).

VI. DISCUSSION

The A common objection is that ordinary citizens lack the technical literacy to adjudicate objective functions. This elides the distinction between technical implementation and normative judgment. Jurors need not understand gradient descent to grasp that weighting "fiscal savings" at 0.65 vs. 0.30 reflects a political choice between austerity and solidarity. Moreover, the "deficit model" of public understanding has been debunked in science communication: when given time, resources, and empowering conditions, citizens demonstrate remarkable capacity for complex reasoning (Wynne, 1992; Dryzek, 2019). The Civic Data Jury (CDJ)'s design stratified selection, expert advocacy, time for deliberation meets these conditions.

The "Society-in-the-Loop" Gap Rahwan’s concept of "Society-in-the-loop" (SITL) has gained traction in recent literature as a critique of current systems. Most governance AI is merely "Human-in-the-loop" (a bureaucrat approves the AI's decision). True SITL requires that society as a whole can debate the variables before the model is trained. Current implementation fails this test; for example, judicial risk scores are often proprietary, meaning the public cannot debate the weight given to "prior arrests" versus "socio-economic status" in the algorithm.

VII. CONCLUSION

The Teleological Deficit is not a bug in current AI governance; it is a defining feature of systems that treat political ends as fixed parameters. Albania's AI Minister and the EU AI Act, despite their differences, both exemplify this deficit: they embed instrumental rationality within democratic structures without democratizing the rationality itself. The result is what I term "procedural legitimation without teleological participation" a hollowed-out democracy where process survives but agency withers. The framework of Algorithmic Contestability offers a path forward. By establishing Civic Data Juries to discuss the purposes of algorithmic governance, we can maintain the fundamental principle of deliberative democracy: that legitimacy stems not from effective results, but from the ability of citizens to collaboratively shape the normative reality they experience. This is not an appeal to eliminate AI in governance; it is a plea to make its purpose political, to bring the objective function from the confines of engineering texts into the openness of public discourse. Subsequent studies should empirically evaluate CDJ designs, formulate optimal methods for converting jurors' normative assessments into computable objective functions, and investigate how this framework adapts to transnational governance (e.g., international AI agreements). Only by accepting this task can we guarantee that the automated state stays, in Habermas's words, a system rooted in the lifeworld, instead of one that takes it over

ACKNOWLEDGMENT

First of I would like to thank God the Almighty for giving me the will and strength, furthermore, I’m thankful to my friend who took the time out of their busy schedule to proofread my work and my family for being patient with me while I was cooped up in the study doing research. Moreover, I extend thanks to the researchers before me whom have contributed a body of knowledge in this field and thank you to the organisers of this conference for giving me the opportunity to express myself through words.

REFERENCES

[1] Albanian Transparency Initiative. (2024). Algorithmic opacity in the DPS system: A technical audit of the AI Minister's welfare allocation model. Tirana: ATI Publications.

[2] Balkans Investigative Reporting Network. (2024, March 15). "The button pushers: How Albania's AI minister silenced street-level discretion." BIRN Online. https://balkaninsight.com/albania-ai-minister-discretion

[3] Binns, R. (2018). "Fairness in machine learning: Lessons from political philosophy." Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (pp. 149–159). ACM.

[4] Buller, J. and Flinders, M., 2006. Depoliticisation: A response to Burnham. The British Journal of Politics and International Relations, 8(2), pp.307-310.
https://doi.org/10.1111/j.1467-856x.2006.00240.x
[5] Burrell, J. (2020). "On the need for accountability in algorithmic optimization." New Media & Society, 22(12), 2108-2127.
https://doi.org/10.1177/1461444819887700
[6] Dryzek, J. S. (2019). Foundations and frontiers of deliberative governance. Oxford University Press.
[7] European Parliament Think Tank. (2024). Albania's digital transformation: An assessment of democratic oversight. EPTA Briefing No. 2024/08.
[8] European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. Official Journal of the European Union, L 1689/1.
[9] Habermas, J. (1981\4). The theory of communicative action, Volume 1: Reason and the rationalization of society (T. McCarthy, Trans.). Beacon Press. (Original work published 1981)
[10] Habermas, J. (1996). Between facts and norms: Contributions to a discourse theory of law and democracy (W. Rehg, Trans.). MIT Press. (Original work published 1992)
https://doi.org/10.7551/mitpress/1564.001.0001
[11] Lipsky, M. (2010). Street-level bureaucracy: Dilemmas of the individual in public services (30th anniv. ed.). Russell Sage Foundation.
[12] Medina, E., & Gaonkar, D. P. (2024). "Algorithmic reason and the end of public reason." Political Theory*, 52(2), 234-263.
[13] Morozov, E., 2013. To save everything, click here: The folly of technological solutionism. PublicAffairs.
[14] Muller, J. W. (2023). "The proceduralist fallacy in AI governance." Philosophy & Public Affairs*, 51(3), 278-315.
[15] O'Neil, C. (2023). Weapons of math destruction: How big data increases inequality and threatens democracy (Revised ed.). Crown.
[16] Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). "Machine behaviour." Nature, 568(7753), 477-486.
https://doi.org/10.1038/s41586-019-1138-y
[17] Rawls, J. (1997). The idea of public reason revisited. The University of Chicago Law Review, 64(3), 765-807.
https://doi.org/10.2307/1600311
[18] Republic of Albania, Ministry of Infrastructure and Energy. (2024). Digital public services (DPS) strategic plan 2024-2030: AI integration roadmap. Tirana: Official Gazette No. 23/2024.
[19] Richmond, J.B. and Kotelchuck, M., 1991. Co-ordination and development of strategies and policy for public health promotion in the United States. Oxford textbook of public health, 1, pp.441-454.
[20] Van der Voort, H., Klievink, B., & Jarke, J. (2024). "Participation theatre in smart city governance." Government Information Quarterly, 41(1), 101-118.
[21] Wynne, B. (1992). "Misunderstood misunderstanding: Social identities and the public uptake of science." Public Understanding of Science, 1(3), 281-304.
https://doi.org/10.1088/0963-6625/1/3/004
[22] Yangguan, L., & Selbst, A. D. (2021). "The teleological illusion in machine learning." Harvard Law Review, 134(6), 1893-1956.

Tomilayo Oyebola Mubarak Aroyehun

Tomilayo is affiliated with the Western Caspian University in Baku, Azerbaijan as a PhD candidate in international relations. As international relations is within his expertise, his focus is on democracy, civic engagement, and political science. His research explores the relationship between economic prosperity and the emphasis on democratic ideals.

Some of the works that the author has published include "Evolution of Democracy: Challenges and Perspectives" which appeared in the European Journal of Political Science Studies in 2024. The research examines the impact of populism and the decline of public trust on democratic institutions, using case studies from the U.S., Brazil, and India. He also published “Democracy and Political Polarization: Explore how political polarization affects democratic systems, including its impact on governance, compromise, and public discourse”, published in the 2024 issue of the Social Sciences journal by the Azerbaijan National Academy of Sciences. The central theme of the paper is how a "deepening divide between political groups or individuals with divergent ideologies, values, and beliefs" impacts democracy. It investigated the specific consequences for governance, the ability to reach compromise, and the nature of public discourse, particularly within a democratic framework. The author has pending publication “The consolidation of Schengen integration through adherence to the rule of law”, he previously attended a conference in Bulgaria related to this topic.

Dr. Aroyehun is currently researching to further the discussion about democracy and polarization which he intends to publish soon, as this is one of his goals to tackle the discourse pertaining democracy. Moreover, In March 2023, he won an essay competition titled "Shusha - the capital of culture of Azerbaijan," organized by the Erasmus Student Network Azerbaijan as part of the "Year of Shusha" project.