OPINION

Governing the Inevitable: Legal Priorities for the Development of Political Institutions in the Age of Artificial Intelligence

Sergey Kabyshev,

Kutafin Moscow State Law University (MSAL)
(Moscow, Russian Federation)

https://doi.org/10.21684/2412-2343-2026-13-1-8-16

Received: July 20, 2025
Reviewed: September 3, 2025
Accepted: November 26, 2025

Abstract. This article examines systemic constitutional challenges arising from the diffusion of artificial intelligence (AI) into the political sphere. It argues that AI is reshaping democratic institutions by generating risks of substituting public deliberation with opaque algorithmic processes, fostering algorithmic discrimination, enabling information micro-manipulation, and concentrating power in the hands of technology holders. Particular attention is devoted to the threat posed to the socio-humanistic paradigm in the context of a normative choice between classical humanism and transhumanism. As a response to these challenges, the article proposes a framework of eight legal principles for AI regulation, including state-level strategic governance, the "human-in-the-loop" principle, anthropological primacy, digital equality, and managed transparency. Within the electoral context, the analysis highlights specific risks such as microtargeting, "dark advertising," deepfakes, and automated bots, which undermine electoral integrity and facilitate manipulation of voters' will. The article concludes that ensuring the sovereignty and legitimacy of political institutions in the digital age requires the development of national AI models and robust legal regulation, including mandatory algorithmic audits and the prohibition of manipulative technologies.

Keywords: artificial intelligence; constitutional challenges; political institutions; AI legal regulation; elections and manipulation; socio-humanistic paradigm.

To cite: Kabyshev, S. (2026). Governing the inevitable: Legal priorities for the development of political institutions in the age of artificial intelligence. *BRICS Law Journal*, *13*(1), 8–16.

Artificial intelligence has now evolved into an infrastructural technology that not only permeates all key domains of state activity, the economy, and social life, but also defines a new political era. Access to such technologies is becoming an increasingly decisive factor, which, together with unprecedented informational capabilities—including a vast potential to enhance the depth and quality of knowledge for strategic governance and political decision-making—can significantly influence shifts in voter preferences[1] and provoke serious deviations from constitutional principles with potentially destructive consequences.

The constitutional challenges posed by artificial intelligence technologies are systemic, synergistic, and existential in their nature. These challenges are outlined here only in a schematic and non-exhaustive manner, among others:

• the substitution of democratic institutions of popular will-formation, grounded in public deliberation, by governance through recommendations generated by opaque algorithms that analyze big data, coupled with a blind faith in the capacity of such algorithms to resolve the most complex social and political problems. This dynamic leads to the oversimplification of public issues, the erosion of democratic debate over values and developmental goals, and transforms algorithms from mere instruments into, in a certain sense, "sources" of power derived from a unique form of "expert" legitimacy. This problem is further exacerbated by the risk of societal fragmentation under the influence of recommendation algorithms governing social media, which divide society into isolated groups inhabiting different "realities," thereby calling into question social cohesion and the unity of the people as the bearer of sovereignty and the sole (and indivisible) source of power;

• multidimensional technology-driven inequality (where radically divergent conditions for the exercise of rights and freedoms, depending on access to technologies, constitute a new vector of social stratification and create preconditions for the de facto marginalization of citizens not integrated into the digital environment), as well as algorithmic discrimination (differential treatment based on biased data);

• the micro-manipulation of information flows through the suppression of certain topics and the amplification of others by means of algorithms and the simulation

[1] Lin, H., et al. (2025). Persuading voters using human–artificial intelligence dialogues. *Nature, 648*, 394–401; Hackenburg, K., et al. (2025). The levers of political persuasion with conversational artificial intelligence. *Science, 390*(6777).

of support or protest (via automated bots and AI agents), as well as the creation of highly persuasive "fakes," thereby undermining the very foundations of the free formation of beliefs and informed choice;

• the displacement and concentration of power resources—grounded in access to big data, computational capacity, and advanced analytical capabilities—in the hands of algorithm owners, which may disrupt the balance within the system of public authority, in particular by substantially weakening the institutional capacity of legislative and judicial bodies.

At the same time, the overarching meta-threat that looms over all others is the challenge to the socio-humanistic paradigm, which is of fundamental significance not only for the development of constitutionalism and law as such, but for civilization itself—a civilization that, within this paradigm, unfolds around and for the human being. V.D. Zorkin, exposing the very core of the problems of socio-legal development in the context of a new technological reality, calls for a focus on the worldview dimension of the issue: upon what value-conceptual foundation is it permissible and justified to construct legal regimes governing the diffusion of advanced technologies—on the basis of transhumanism, which, by linking human perfection, including its ethical dimension, to the level of technological development, not only strives for a synthesis of human and machine but ultimately humanizes and normatively elevates the machine; or on the basis of classical humanism, which embodies traditional spiritual and moral values and rejects the separation of personality from conscience, as well as the embrace of soulless moral relativism and diluted responsibility.

It is precisely within this dilemma, according to Zorkin, that the principal and fundamental challenge of artificial intelligence to humanity and its legal order resides. Law, under these conditions, risks losing its humanistic essence and moral foundation, potentially degenerating into a system of speculative and fictive algorithms that reduce the individual to a set of reflexes and behavioral responses, thereby "liberating" the person from their natural constraints and allowing them to acquire—or "choose"—any artificial forms.[2] This transhumanist paradigm entails a dual consequence: the "humanization" of the machine and the depersonalization (algorithmization and dehumanization) of the human being.

Artificial intelligence technologies, especially when combined with bio- and psychometrics and neurotechnologies, may call into question the very autonomy of the human person and their inherent dignity. What is at stake is the reduction of the legal and political understanding of the human being as a free, rational, and integral subject to a predictable "profile," a move that dehumanizes social relations and the political sphere. At the same time, technological systems of monitoring and control that are evolving toward total pervasiveness risk transforming the public

[2] Zorkin, V. D. (2024). *Lectures on law and the state*. Constitutional Court of the Russian Federation. (In Russian).

space into a panopticon, in which the individual is placed in the position of an object of constant evaluation and social scoring, rather than that of a free subject.

In seeking responses to the fundamental challenges associated with the diffusion of artificial intelligence in the political sphere, it is essential to rely on a measured and balanced general approach to the construction of legal regimes for the relevant technologies, avoiding both alarmism and complacent overconfidence. In this context, the author proposes—by way of discussion—an original perspective on a set of core principles that may serve as a methodological framework for ensuring the sustainable development of artificial intelligence technologies grounded in the protection of the interests of the individual, society, and the state within the digital domain.

1. The Principle of State Strategic Governance and the Assurance of Technological Integrity in the Field of Artificial Intelligence. Recognizing artificial intelligence as a strategic vector of development, the Russian Federation must exercise a comprehensive set of sovereign prerogatives and measures aimed at the formation and protection of a complete national technological cycle—from fundamental research and human capital development to the creation of hardware components, software, data resources, and markets. This includes the proactive shaping of regulatory, economic, and infrastructural conditions conducive to the accelerated development and safe deployment of artificial intelligence, as well as the attainment and maintenance of leadership positions through the creation of competitive, secure, and ethically grounded technologies. This principle also encompasses sovereign control over critically significant artificial intelligence technologies in vital domains such as national security, defense, critical information infrastructure, and energy systems, among others, and the establishment of "national identity filters" designed to ensure that algorithmic operations conform to the principles and values underpinning Russian society and its traditional spiritual and moral ideals.

2. The Principle of Guaranteed Protection of the Injured Party and Enhanced Liability.[3] This principle implies the legislative recognition, at a minimum, of strong artificial intelligence (autonomous systems capable of making decisions without human involvement) as a source of increased danger, which would likely entail a regime of joint and several liability. Such liability should extend, in particular, to the developer (the creator of the algorithm or model), the operator, and the owner (the entity deploying or using the system). It is evident that failures may occur at any stage of the technological chain, while for the injured party it is often technically impossible to establish the precise cause of harm. In this regard, consideration should be given to the establishment of a special state or industry-based insurance or compensation fund, coupled with the possibility of recourse claims against participants in the relevant technological chain. At the same time, it

[3] Sukhanov, E. A. (2025). On civil law problems of digitalization. *Herald of Civil Procedure, 1*, 37–52. (In Russian); Vasilevskaya, L. Yu. (2025). Delict liability for harm caused by artificial intelligence: Problems and development prospects. *Civil Law, 4*, 2–5. (In Russian).

is essential to avoid maximalist regulatory approaches that could prompt a "flight" of high-technology companies to jurisdictions with more lenient regulation or lead to excessively conservative algorithmic design—for example, the refusal of diagnostic systems to identify rare conditions.

3. The "Human-in-the-Loop" Principle in Critical Decision-Making. This principle requires the mandatory retention of final decision-making authority by a human being in socially significant matters affecting life, health, and personal liberty, the allocation of substantial public resources, and the use of coercive force.

4. The Principle of Anthropological Primacy. Any artificial intelligence technologies must be designed on the premise that the human being—his or her autonomy, dignity, freedom, and well-being—constitutes its purpose, measure, and ultimate constraint. Technology has no intrinsic value outside its service to these ends. This principle stands in opposition to a narrow technocratic paradigm in which efficiency, speed, and economic gain enabled by algorithms are elevated to ends in themselves. Accordingly, the design and interfaces of such technologies should be oriented toward augmenting human capabilities—figuratively speaking, toward co-piloting rather than autopiloting. They should enhance situational awareness, relieve humans of routine tasks, and process large-scale data, while preserving human leadership in meaning-making, creativity, and responsibility. This, in turn, presupposes design solutions that preclude the illusion of machine infallibility: algorithmic outputs should be accompanied by explanations of the key contributing factors, visualizations of uncertainty, and opportunities for expert evaluation of the principles and methods underlying decision-making. At the same time, the development and deployment of artificial intelligence technologies that exploit personal and behavioral data for covert influence on users—by leveraging known cognitive biases in order to undermine the autonomy of will, induce decisions contrary to the user's evident interests, or operate in circumstances where the user is in a particularly vulnerable position known to the system—must not be permitted.

5. The Principle of Digital Equality and the Prohibition of Discrimination. This principle entails, *inter alia*, the mandatory conduct of regular independent audits to detect and assess bias in artificial intelligence technologies used for decision-making in public and socially significant domains, as well as a prohibition on the deployment of technologies that exhibit unexplained and irremediable disparities in treatment.

6. The Principle of Respect for Human Dignity and Non-Interference with the Private Sphere. This principle requires a prohibition on mass psychographic profiling aimed at covert behavioral manipulation, limitations on the use of affective technologies (including emotion recognition) and social scoring in the public sphere, and the preservation of a space for "non-digitized" life.

7. The Principle of Risk-Based Managed Transparency. For high-risk systems, this principle requires ensuring the individual's right to obtain a substantive explanation

of an algorithmic decision that has affected his or her rights and interests, presented in a form comprehensible to a human user. It also imposes an obligation on developers to provide regulatory authorities with access to algorithms and data for the purposes of oversight and verification.

8. The Principle of Technological Neutrality and Anticipatory Regulation. This principle calls for a regulatory focus on the properties and risks of systems rather than on specific technologies—which are inherently dynamic and prone to rapid obsolescence—and for the establishment of special legal regimes that enable the testing of innovations under real yet controlled conditions.

Within this framework, particular attention must be paid to a number of significant risks affecting political and legal relations, including the electoral process, which require appropriate reflection in legal regulation. It should be emphasized that the scope and influence of artificial intelligence applications are steadily expanding. As demonstrated by analyses of international practice, in many instances such technologies are employed with the intent to harm political competitors or to interfere with the organization and conduct of elections, while the provenance of the relevant content often remains technically untraceable or legally indeterminable.[4]

Since the functioning of artificial intelligence is determined by algorithms that are developed and trained within a specific socio-cultural and historical context, the use of foreign-developed AI models in the political sphere may pose risks to Russia's overarching national identity. Such models may shape perceptions, attitudes, and patterns of action (or inaction) on the basis of values and interpretative frameworks that do not correspond to Russia's political and legal traditions, ideals, and social realities. Expert studies indicate that many existing artificial intelligence models tend to reproduce preference structures aligned with the value systems prevalent in the United States and Western Europe, particularly with respect to individual rights, gender policy, and conceptions of social justice. A discernible tendency has been identified for these models to treat Western normative frameworks as universal and globally applicable. Moreover, specific empirical studies demonstrate that, in situations involving moral choice, artificial intelligence technologies frequently rely on Western philosophical traditions—most notably utilitarianism and liberal ethical theory.[5]

In this regard, the introduction of artificial intelligence into the political sphere should be oriented toward the development of domestic AI models and the establishment of a system of trusted technologies, which constitutes a critical prerequisite for safeguarding sovereignty. Nationally developed models make it

---

[4] Ustinovich, E. S. (2024). Generative artificial intelligence in the 2024 electoral processes worldwide: Disinformation campaigns and online trolls. *Social Policy and Social Partnership, 3*. (In Russian).

[5] Münker, S. (2025). *Cultural bias in large language models: Evaluating AI agents through moral questionnaires*. arXiv. https://arxiv.org/html/2507.10073v1; Peters, U., & Carman, M. (2024). Cultural bias in explainable AI research: A systematic analysis. *Journal of Artificial Intelligence Research, 79*, 971–1000; Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9), 346.

possible to account for the domestic context, the specificities of Russian law, the priorities of state policy, and the structural features of society.

Artificial intelligence is capable of fundamentally transforming the very notion of freedom of choice, with potentially corrosive effects on free, conscious, and voluntary political will-formation. In particular, microtargeting technologies based on the analysis of personal data enable the creation of highly personalized forms of political persuasion that construct an artificial informational reality for the voter—so-called "semantic hallucinations"—and effectively steer motivation and beliefs. The use of so-called dark advertising, which is visible only to a specific user or narrowly selected audience, remains inaccessible to public scrutiny. It involves the dynamic modification of content, whereby algorithms test thousands of variants of headlines, images, and texts to identify the most emotionally resonant options. This practice transforms political communication into a form of covert and unilateral influence.

For similar purposes, technologies commonly described as "bot armies" may be employed, whereby artificial intelligence controls networks of fake accounts that simulate public opinion, amplify scandals, and create the appearance of mass support or condemnation, thereby manipulating the public agenda. Such systematic influence, aimed at inducing a predetermined choice, can in effect perform functions analogous to electoral campaigning; however, given its diffuse nature and extensive informational scope, it is extremely difficult to trace and to identify as such.

At the same time, contemporary technologies make it possible to generate fabricated statements by candidates, compromising materials, or distorted reports from polling stations that are virtually indistinguishable from authentic content. Under such conditions, truth is often unable to keep pace with falsehood. Such materials may be deployed to discredit political opponents, suppress voter turnout, or provoke social tension.

In this respect, although not specific to electoral processes as such, regulatory approaches governing the operation of information content platforms in China are of particular relevance for the development of the political sphere. These rules provide, *inter alia*, that where a platform employs personalized algorithmic recommendation technologies for the dissemination of information, it must construct recommendation models based on predefined restrictions and prohibitions. These include bans on the dissemination of rumors; obligations to prevent the spread of information that relies on exaggerated or misleading headlines where the content substantially fails to correspond to the title; content designed to inflame scandals, conflicts, or misconduct; inappropriate commentary on natural disasters; and the promotion of vulgar or kitsch content (Arts. 6, 7, and 12 of the Provisions on the Governance of the Online Information Content Ecosystem, 2019[6]). The Provisions on the Administration

[6] Cyberspace Administration of China. (2019). *Provisions on the governance of the online information content ecosystem*. https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm. (In Chinese).

of Algorithmic Recommendation Services for Internet Information Services (2021) explicitly prohibit providers of algorithmic recommendation services from using algorithms to register fake accounts, manipulate user accounts, or generate false reactions such as "likes," comments, or reposts; from interfering with information presentation through practices such as blocking content, creating excessive recommendations, manipulating trending lists or search results; and from controlling search queries or samples, influencing online public opinion, or evading supervision and regulation (Art. 14).[7]

The diffusion of artificial intelligence also raises the issue of a de facto technological qualification threshold in elections, which poses a direct threat to the principles of electoral equality and fairness. Financial and technological constraints result in unequal access to AI capabilities that confer unjust advantages. At the same time, the emergence of fully digital candidate avatars or hyper-realistic replicas of real politicians—capable of being "fine-tuned" to the preferences of virtually any audience—becomes conceivable. This blurs the boundary between a real individual with authentic convictions and a controllable software construct, thereby misleading voters. As a consequence, an entirely new landscape of political interaction is taking shape.

Ultimately, it is necessary to speak of a systemic and comprehensive constitutional threat: the erosion of electoral integrity and good faith, which may result in a large-scale crisis of the political and legal epistemology of elections—namely, a situation in which it becomes impossible to trust what one sees or hears. Irrespective of the specific type of electoral system employed, elections are inextricably linked to the human being as the bearer of values, moral–political orientations, and convictions. As follows from the jurisprudence of the Constitutional Court of the Russian Federation, a rule-of-law democracy, in order to remain sustainable, requires effective legal mechanisms capable of protecting it from abuses of public authority, the legitimacy of which is largely grounded in public trust. For this reason, heightened requirements may be imposed with respect to the reputation of persons holding public office, so as to prevent the emergence of doubts among citizens regarding their moral and ethical qualities (see, for example, Judgment of October 10, 2013 No. 20-P).

In the context of the expanding use of artificial intelligence, however, elections risk being transformed into a concealed "war of algorithms," in which victory accrues not to those who persuade more convincingly, but to those who manipulate more effectively, possess greater resources, and more skillfully circumvent moral constraints.

Ensuring the legitimacy of elections under these new conditions requires not only technological solutions (such as deepfake detection tools), but also robust legal regulation aimed at achieving algorithmic transparency, the "visibility" of synthesized content through mandatory labeling and archiving, the prohibition of

---

[7] Cyberspace Administration of China. (2021). *Provisions on the management of algorithmic recommendation services*. https://www.chinalawtranslate.com/en/algorithms/. (In Chinese).

covert manipulative technologies, and the enhancement of voters' digital literacy. It is therefore justified to raise the issue of introducing mandatory audits of algorithms used in the electoral process, so that any AI-based technology employed during elections is subject to prior assessment, in particular with regard to explainability (the ability to understand the logic of decision-making) and the absence of discriminatory biases. In light of existing foreign experience (including South Korea and certain US states, such as California and Texas), the question of prohibiting—and potentially criminalizing—the malicious use of deepfakes in elections likewise warrants serious consideration.

References

Hackenburg, K., et al. (2025). The levers of political persuasion with conversational artificial intelligence. *Science, 390*(6777). https://doi.org/10.1126/science.aea3884

Lin, H., et al. (2025). Persuading voters using human–artificial intelligence dialogues. *Nature, 648*, 394–401. https://doi.org/10.1038/s41586-025-09771-9

Münker, S. (2025). *Cultural bias in large language models: Evaluating AI agents through moral questionnaires*. arXiv. https://arxiv.org/html/2507.10073v1

Peters, U., & Carman, M. (2024). Cultural bias in explainable AI research: A systematic analysis. *Journal of Artificial Intelligence Research, 79*, 971–1000. https://doi.org/10.1613/jair.1.14888

Sukhanov, E. A. (2025). On civil law problems of digitalization. *Herald of Civil Procedure, 1*, 37–52. (In Russian). https://doi.org/10.24031/2226-0781-2025-15-1-37-52

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9), 346. https://doi.org/10.1093/pnasnexus/pgae346

Ustinovich, E. S. (2024). Generative artificial intelligence in the 2024 electoral processes worldwide: Disinformation campaigns and online trolls. *Social Policy and Social Partnership, 3*. https://doi.org/10.33920/pol-01-2403-03. (In Russian).

Vasilevskaya, L. Yu. (2025). Delict liability for harm caused by artificial intelligence: Problems and development prospects. *Civil Law, 4*, 2–5. https://doi.org/10.18572/2070-2140-2025-4-2-5. (In Russian).

Zorkin, V. D. (2024). *Lectures on law and the state*. Constitutional Court of the Russian Federation. (In Russian).

Information about the author

Sergey Kabyshev (Moscow, Russian Federation) – Associate Professor; Chairman, Committee on Science and Higher Education of the State Duma of the Federal Assembly of the Russian Federation; Professor, Department of Constitutional and Municipal Law, Kutafin Moscow State Law University (MSAL) (9 Sadovaya-Kudrinskaya St., Moscow, 125993, Russian Federation; e-mail: svkabyshev@mail.ru).