UDC 327.56:004.8
DOI: 10.5937/nint54-64046
Review scientific paper

НАЦИОНАЛНИ ИНТЕРЕС
NATIONAL INTEREST
Година XXII, vol. 54
Број 2/2026
стр. 31-52

Aleksandar Pavić*
Military Academy, University of Defense in Belgrade

ALGORITHMIC DETERRENCE: A NEW STRATEGIC LOGIC IN THE ERA OF ARTIFICIAL INTELLIGENCE

Resume

The development of artificial intelligence represents a turning point in the modern understanding of deterrence strategy, shifting its focus from material power to information, perception, and the algorithmic management of actor behaviour. The classic concept of deterrence, as advocated by Thomas Schelling and Bernard Brodie, based on rational cost-benefit assessment, loses its predictive power in conditions of accelerated decision-making, autonomous systems, and algorithmic opacity. Instead of a stable "balance of fear", a state of dynamic instability arises, in which speed, information asymmetry and cognitive effects become key factors of strategic equilibrium. The paper's starting hypothesis is that artificial intelligence does not nullify deterrence but transforms it through a new cognitive and informational dimension: "algorithmic deterrence". Defence is no longer aimed exclusively at inflicting unacceptable damage, but at managing the adversary's perception, expectations, and behaviour in real time. Autonomous systems, machine learning, and predictive models enable the development of "intelligent deterrence" based on tailored strategic signals and algorithmic analysis of behaviour. Special attention is paid to the concepts of cognitive and information deterrence as modern

* Email address: aleksandar.pavic@mod.gov.rs; ORCID: 0009-0005-6737-0094.

forms of strategic communication. In this context, the paper argues that smaller states, such as Serbia, can develop asymmetric models of intelligent deterrence in the cyber and cognitive domains within the concept of total defence and digital sovereignty. In conclusion, the revitalisation of deterrence does not imply a return to classical models, but their transformation into a system of dynamic risk, information and perception management in the era of artificial intelligence.

Keywords: algorithmic deterrence, strategic logic, artificial intelligence, digital era.

INTRODUCTION

The development of artificial intelligence is rapidly transforming the modern security environment and challenging the fundamental assumptions of strategic thinking. In conditions where algorithmic systems analyse data, predict outcomes, and make decisions at a speed that exceeds human cognitive capacities, classic concepts based on rationality, predictability, and clear communication of intent are increasingly difficult to explain the dynamics of deterrence (Payne 2021). What was perceived as a relatively stable "balance of fear" during the Cold War is increasingly transforming into a state of dynamic instability, marked by speed, information asymmetry, and algorithmic opacity.

Classical deterrence theory, founded in the works of Thomas Schelling and Bernard Brodie, assumed rational actors capable of assessing costs and benefits, understanding signals, and controlling the escalation of conflict (Schelling 1960; Brodie 1959). Deterrence is understood in this context as the "art of risk manipulation," in which a credible threat deters aggression. However, in the era of artificial intelligence, rationality increasingly relies on algorithmic calculations of probabilities and scenarios, altering the conditions for deterrence's credibility and predictability (Horowitz 2010; Geist and Lohn 2018).

Contemporary literature indicates that artificial intelligence significantly influences the character of warfare, command, and decision-making, but it remains unclear how these processes transform the strategy of deterrence itself as a form of risk management (Allen and Chan 2017; Johnson 2021). In particular, the impact of autonomous

and semi-autonomous systems on threat credibility, perception management, and stability under conditions of epistemological uncertainty is insufficiently developed. It is precisely in this theoretical gap that the need to conceptualise algorithmic deterrence as a new strategic logic arises.

The paper's starting hypothesis is that artificial intelligence does not nullify the concept of deterrence but transforms it through a new cognitive and informational dimension. In this context, deterrence is no longer focused solely on the threat of causing unacceptable damage, but on managing the adversary's perceptions, expectations, and behaviour in real time. This transformation leads to the emergence of algorithmic deterrence, in which technological systems take on an active role in shaping security stability. Its essence lies not in replacing man with machine, but in creating a mutual relationship in which algorithms act as an extension of human rationality, but with a different pace of decision-making and a specific epistemology.

Methodologically, the work is based on a qualitative analysis of relevant theoretical and strategic literature in international relations, security studies, and military strategy, with a comparative interpretation of classical and contemporary concepts of deterrence. A conceptual modelling of algorithmic deterrence as an analytical framework, as well as an illustrative consideration of the implications for smaller states in the contemporary security environment, complements the analysis.

In a broader context, algorithmic deterrence is associated with the development of cognitive and informational deterrence, in which the control of information flows, narratives and perceptions becomes a key resource of power (Mazarr 2018). This dimension is particularly important for smaller states, which often lack the resources to engage in symmetric deterrence against great powers but can develop asymmetric and adaptive approaches within the concepts of total defence and digital sovereignty.

This paper aims to offer an analytical framework for understanding algorithmic deterrence as a new phase in the evolution of strategy. The paper first considers the development and limitations of classical deterrence theory, then analyses the impact of artificial intelligence on rationality, credibility, and risk management, and finally defines the key dimensions of algorithmic deterrence. In the final section, the analysis focuses on the implications for smaller states, with

particular attention to the Republic of Serbia as an example of strategic adaptation in the algorithmic era.

THEORETICAL AND HISTORICAL FRAMEWORK OF THE DETERRENCE STRATEGY

The concept of deterrence emerged within classical strategic thought in response to the changing nature of war in the nuclear era. Instead of physical conflict and battlefield victory, the essence of strategy became the prevention of war through psychological and political manipulation backed by the threat of force (Freedman 2004). In this context, deterrence was defined as "the use of potential, rather than actual, force" to control an adversary's behaviour (Brodie 1959, 21).

Classical theorists, such as Schelling and Brodie, laid the foundation for what would later be called the "rationalist paradigm of deterrence." Schelling observed that strategy in the nuclear age had become "the art of communication," in which the key resource was the adversary's belief in the credibility of the threat (Schelling 1960). Deterrence, therefore, depended not only on military superiority but also on the ability of the actor to convince the opposing side of its readiness to escalate, even if the outcome would be catastrophic for both sides.

Brody had already observed in 1946 "the purpose of weapons is no longer in their use, but in their possession" (Brodie 1959, 23). Thus, the logic of nuclear deterrence fundamentally changed the relationship between politics and strategy. Force was no longer an instrument of victory, but an instrument of behavioural control. Since then, deterrence has become a central category of modern strategic theory and a key concept in maintaining global stability during the Cold War.

Deterrence theory rests on three fundamental pillars: threat credibility, communication of intent, and actor rationality (Jervis 1989). Credibility is a prerequisite for any successful deterrence; the adversary must believe that the threat will be carried out if he crosses the red line. Communication, in turn, allows that threat to be understood, while rationality ensures that both sides will act predictably in accordance with the logic of costs and benefits.

However, this model contains several inherent paradoxes. The first is the paradox of credibility: how can a threat be credible if its

implementation leads to mutual destruction? The second is the paradox of rationality: what happens when one side acts irrationally, or when bounded rationality influences risk assessment (Barros 2010)? The third is the paradox of communication: how is intent conveyed amid strategic ambiguity, propaganda, and disinformation?

These paradoxes have become the starting point for expanding the concept of deterrence beyond the nuclear context. With the emergence of new technologies and new domains, namely cyber, space, and information, deterrence has transformed into a networked and cognitive phenomenon (Libicki 2009). Thus, a whole series of new models emerged from "cyber deterrence", through "deterrence by resilience", to the concept of "cognitive deterrence" (Mazarr 2018).

The Cold War model of deterrence was based on a binary logic of power and balance. Its symbol was the concept of "Mutually Assured Destruction" (MAD), which guaranteed stability through the fear of total catastrophe. However, the end of the Cold War, the digital revolution and the spread of new security challenges have led to the relativisation of this logic. Contemporary deterrence is less about the physical dimension of force and more about perceptual and cognitive control.

In the post-Cold War world, states and non-state actors use information operations, media narratives and cognitive techniques to influence the behaviour of adversaries without the overt use of force (Nye 2011). As authors Pavić and Beriša point out, in the modern environment, information becomes an instrument of strategic control and stability (Pavić and Beriša 2025). Deterrence has thus become multidimensional: military, economic, technological, cognitive, and cultural. Rather than threatening physical attack, it now seeks to create a perception of uncertainty, which restrains the enemy through doubt in one's own capabilities.

Cognitive deterrence uses control over information to shape decisions. According to Mazarr, the goal is "to restrain the adversary by managing his perception of risk" (Mazarr 2018). Here, deterrence is no longer based on threats, but on manipulating reality. This opens the way for the next phase, algorithmic deterrence, in which perceptions and decisions are increasingly mediated by technology.

The development of artificial intelligence changes not only the instruments of strategy, but also its internal logic, which is precisely what constitutes the essence of the transformation. Algorithms capable

of prediction, analysis and independent decision-making introduce a new type of actor into the deterrence process. Autonomous systems introduce an unimaginable speed into the dynamics of escalation, making classic mechanisms of communication and control increasingly unreliable (United Nations General Assembly [UNGA] 2024; Horowitz 2020).

Historically, every technological revolution in weapons has brought a new form of strategic logic: the steam engine – industrial warfare, nuclear energy – global deterrence, digital technology – information warfare. Artificial intelligence, as the "fourth industrial revolution" (Schwab 2017), introduces the logic of automated strategy, in which decision-making processes are delegated to algorithms that operate in real time across multi-domain systems.

This also changes the basic assumptions about control. While classical deterrence was based on the stability of decisions, algorithmic deterrence introduces the dynamics of constant learning and adaptation. Instead of a static balance of fear, a dynamic balance of data emerges, where the key is the system's ability to recognise and simulate the adversary's intentions in real time. This is no longer deterrence in the classical sense, but a form of predictive prevention, in which the algorithm learns from continuous analysis of the environment and seeks to act as a preventive element between potential threats and their deterrent effects before they become real. A concept defined in this way could be called pre-emptive deterrence.

Although deterrence theory was one of the most successful concepts of the 20th century, it shows limitations in the 21st century. First, it assumes stable actors with clearly defined interests and rationality. Today's security environment is characterised by multi-actor complexity: states, corporations, algorithmic platforms, and autonomous systems operate in the same space. Second, the classical model assumes temporal continuity and predictability. Algorithmic speed, however, abolishes this predictability, creating the possibility of unexpected escalations (Payne 2021). Third, traditional deterrence relies on transparent communication. Algorithmic systems, in contrast, operate under conditions of epistemological opacity (the black-box problem), which calls into question actors' ability to understand one another. If even programmers cannot fully explain the behaviour of their systems, how can an adversary be expected to interpret their signals correctly?

Therefore, classic concepts such as "credibility", "escalation" and "stability" require reinterpretation. A new theoretical framework must include the dynamics of algorithmic learning, the level of simulation, and the issue of algorithmic transparency. This is where the field of algorithmic deterrence begins as a new paradigm in strategic theory.

Algorithmic deterrence represents a combination of three evolutionary layers of strategic thought: nuclear, informational and cognitive. It is not a complete rejection of old theories, but their reinterpretation through a technological prism. Just as nuclear deterrence is based on the fear of physical destruction, and cognitive deterrence is based on the fear of losing control over perception, algorithmic deterrence is based on the fear of losing control over an autonomous system.

Its essence lies in the control of predictive power: whoever controls the data controls the perception of reality, and therefore the possibility of deterrence. Thus, algorithmic deterrence is not just a military or technological category, but also a new epistemological doctrine, a way of understanding and producing stability in a world governed by data.

THE CONCEPT OF ALGORITHMIC DETERRENCE: STRUCTURE, MECHANISMS AND DIMENSIONS

Algorithmic deterrence in perspective may represent a new theoretical and practical model in the evolution of strategic thought. Artificial intelligence is not just a new instrument of strategy; it is changing the meaning of strategic thinking. It is based on the assumption that, in the era of artificial intelligence, decisions about the use of force, risk assessment, and crisis management increasingly depend on algorithmically generated analyses, simulations, and recommendations. In this sense, algorithmic deterrence is not only a technological phenomenon but also an epistemological transformation, in which knowledge, perception, and power are combined into a single system of behaviour management.

In the classical model, deterrence functions as a communication of threat: one actor seeks to convince another to refrain from action, because the cost will be too high (Schelling 1960). In the algorithmic model, machines that analyse, predict, and respond in real time mediate communication. The threat is no longer just an explicit message, but

also a demonstrative simulation of power, a networked presence that operates through data, not words.

Therefore, the essence of algorithmic deterrence lies not in intent but in the system's structure. It operates not through the announcement of a threat, but through the visible ability to predict and control. When an adversary knows that his actions are constantly monitored, analysed, and simulated in an algorithmic environment, the mere awareness of this fact becomes a deterrent.

The first dimension of algorithmic deterrence is cognitive. While classical deterrence was based on human rationality, the ability of actors to logically assess consequences, algorithmic deterrence introduces machine rationality, that is, the ability for algorithms to process large amounts of data and create behavioural models inaccessible to the human mind.

In the process, AI-based systems become intermediaries between human decision-making and reality. They not only interpret data but also shape perceptions of the security situation. As Coker points out, in the era of autonomous systems, war is no longer a clash of wills, but a clash of algorithms that construct reality (Coker 2019). This is precisely the cognitive nature of the new deterrence: it operates at the level of risk perception, not at the level of direct threat.

The cognitive power of algorithms stems from their ability to "predict the unexpected" and detect patterns of behaviour that escape the human mind. This predictive function serves as the basis for preemptive deterrence, in which systems respond before the adversary consciously makes a decision. In this way, deterrence becomes an "intelligent ecosystem" that functions through continuous learning and adaptation.

The second dimension is operational, namely, how algorithmic deterrence operates within specific security and military systems. It relies on a combination of three levels of technological integration:

- Sensory level – data collection through sensor networks, satellites, unmanned systems and cyber interfaces.
- Analytical level – use of artificial intelligence to analyse behavioural patterns and simulate scenarios.
- Decision-making – transfer of analysis results into the strategic decision-making process, either human or autonomous.

This architecture enables a continuous cycle of situational awareness in which each new piece of information changes the dynamics of the threat. In this context, deterrence is no longer a linear process, but an adaptive decision-making system.

The theoretical positions presented above, implemented in systems such as Project Maven (United States), Skynet (PRC) or Prometheus AI (Russian Federation), show that deterrence becomes a function of a state's ability to interpret reality more quickly and accurately. The number of tanks or missiles no longer measures power, but by the quality of algorithms and the amount of data at its disposal.

Such a model leads to the emergence of what can be called algorithmic dominance. A situation in which superiority in predicting the behaviour of the adversary is transformed into strategic superiority. Instead of states seeking to "persuade" the adversary, they "outsmart" it. This is a fundamental paradigm shift: from persuasion to simulation.

The third, most profound dimension of algorithmic deterrence is epistemological and concerns who controls knowledge and how knowledge is transformed into power. Algorithms not only analyse data, but also create epistemological frameworks through which reality is interpreted. They choose what will be visible, what will be labelled as a threat, and what will be labelled as insignificant noise. Boratov implies that artificial intelligence is not just a technological advancement, but also a strategic upheaval. Traditional models built on human rationality, clear signalling, and predictable escalation are all inadequate in the face of autonomous systems and algorithmic warfare. According to him, the security dilemma is no longer just about capabilities, but also about understanding (Boratov 2025).

The problem is that most modern systems based on artificial intelligence are a kind of so-called „black box". The logic in such decision-making systems is not fully understandable even to the programmers themselves (Burrell 2016). In such an environment, the traditional concept of threat credibility becomes difficult to sustain. If actors cannot explain how and why the system makes certain decisions, how can they expect the adversary to believe it? Accordingly, the epistemological dimension of algorithmic deterrence implies two opposing tendencies:

- Opacity as a source of fear and uncertainty, which can act as a deterrent.
- Transparency as a prerequisite for credibility and control.

Between these two extremes lies contemporary strategic logic: the smarter the system, the less understandable it is. In addition, the less understandable it is, the harder it is to control, but potentially more effective in deterrence. This dialectic makes algorithmic deterrence a new form of epistemological power. Algorithmic deterrence operates through several key mechanisms:

- Perceptual dominance – creating awareness among the adversary that he is under constant surveillance and that his behaviour is predicted in advance.
- Simulation demonstration – using simulations, artificial intelligence analysis and media narratives to demonstrate superiority in information architecture.
- Algorithmic unpredictability – deliberately limiting transparency to produce psychological uncertainty.
- Integrated adaptation – the continuous evolution of a system that "learns" from the adversary's reactions and thereby adjusts its deterrence model.

These mechanisms operate simultaneously and dynamically. Instead of the classical logic of stability, algorithmic deterrence relies on controlled instability – maintaining a level of uncertainty that prevents escalation but maintains the perception of power. This is what Horowitz and Scharre describe as "the dynamic equilibrium of fear in the age of algorithms" (Horowitz 2020; Scharre 2018).

Like all previous models, this one contains internal paradoxes. The first is the paradox of autonomy: the more autonomous a system is, the less under human control it is. The second is the paradox of speed: a split-second algorithmic reaction can prevent an attack, but can also trigger an unwanted escalation. The third is the paradox of trust: actors must trust a system that they themselves do not fully understand.

These paradoxes introduce a new type of strategic dilemma, the dilemma of machine rationality. While classical deterrence relied on human ability to assess risk, the stability of the system now relies on algorithms' ability to stay within acceptable bounds of behaviour. This means that strategic stability is no longer a political issue, but also a technological and ethical one.

Algorithmic deterrence represents the intersection of knowledge and capabilities of technology, strategy, and perception, as shown in Table 1.

Table 1. Dimensions of algorithmic deterrence

Dimension	Key concept	Basic function
Cognitive	Predictability	Shaping perceptions of risk
Operational	Adaptation	Simulation and management by behaviour
Epistemological	Opacity	Creation of psychological effects and distractions

Source: Author's processing.

The combination of these elements creates a new form of power that is not based on force or threat, but on the architecture of knowledge. Algorithmic deterrence thus becomes a paradigm in which power is defined by the ability to predict, control, and reprogram behaviour in conditions of global digital addiction.

GEOPOLITICAL AND SECURITY IMPLICATIONS OF ALGORITHMIC DETERRENCE

The development of artificial intelligence has accelerated the change in the structure of global power, introducing a new axis of competition that is no longer based exclusively on military or economic capabilities, but on algorithmic superiority, that is, the ability of a state to develop, control and integrate intelligent systems into all domains of strategic action. Algorithmic deterrence in this context becomes a means of projecting power in the digital space, where security increasingly relies on controlling data, infrastructure, and the cognitive environment. Instead of the classical balance of power, the international system is gradually entering an era of algorithmic balance, in which stability is based on the dynamic interaction of artificial intelligence, information flows and human surveillance. According to Pavić and Beriša, artificial intelligence and its rapid development are seen as one of the mechanisms that will significantly influence the instruments and mechanisms that demonstrate a country's technological and technical component of development in the future. That will certainly be used as an element of deterrence or of active action to advance national

interests, alongside other factors (Pavić and Beriša 2024), which supports the theory of algorithmic deterrence.

For major powers, especially the United States and the People's Republic of China, algorithmic deterrence is becoming a central element of national security strategy. In the American case, artificial intelligence is integrated into the concept of multi-domain operations and advanced decision-making (decision advantage), where automated systems are used to detect, assess, and respond to threats more quickly. The key goal is to shorten the time cycle from detection to decision, enabling faster action and creating a deterrent effect without the need for force.

China, on the other hand, sees artificial intelligence as the basis of 21st-century strategic rivalry, within the framework of a vision of "intelligent warfare", where algorithms act as "power multipliers" in all domains, from surveillance and information analysis to cognitive influence and information operations. Its approach seeks to integrate military, industrial, and civilian AI infrastructure into a single system of state power. In this way, algorithmic deterrence in the Chinese model becomes an instrument of total integration of knowledge and power, where deterrence is achieved not through threat, but through dominance in the field of technological dependence.

In both cases, a shift from material to structural power is visible (Strange 1988), in which controlling the architecture of the technological ecosystem becomes a new form of strategic influence. Great powers seek not only to gain technological advantage, but also to set standards and protocols that define global norms for the use of artificial intelligence in security contexts. Thus, algorithmic deterrence is transformed into an instrument of geopolitical norms, a means by which the rules of the game in the age of artificial intelligence are shaped.

For smaller states, including Serbia and similar countries with modest technological capabilities, algorithmic deterrence poses a dual challenge: technological dependency and strategic transparency. Without their own capacities to develop sovereign AI-based platforms, these states risk becoming objects of other people's algorithmic structures, in which data, infrastructure, and analytics are controlled from the outside. Such a situation leads to the emergence of algorithmic asymmetry, in which power does not come from armed force but from the ability to manage the digital ecosystem. However, smaller states are

not without room for manoeuvre. The concept of intelligent deterrence implies a flexible strategy based on knowledge, transparency, and cooperation with trusted partners. Instead of racing for global leadership in AI development, they can build their own capacities in areas such as cybersecurity, data protection, and cognitive security systems. This would create a strategy of adaptive resilience, in which algorithms are used as instruments of controlled predictability, rather than as substitutes for human strategic rationality.

The global system is entering a phase of technological fragmentation. The American and Chinese technological spheres increasingly function as separate ecosystems of standards, infrastructure, and data. Algorithmic deterrence in such an environment becomes a mechanism for maintaining the balance between these spheres. Instead of the classic nuclear balance, which was based on the symmetry of destruction, a digital balance of predictability is now emerging, where stability depends on whether and to what extent one sphere can understand and predict the other's behaviour.

This so-called "algorithmic balance" is ambiguous. On the one hand, it can act as a stabilising factor, reducing the likelihood of unintended escalation through systemic interdependence. On the other hand, the opacity of algorithms and their tendency to self-adapt can lead to instability and unpredictability, leaving actors unable to control the dynamics of their own systems. In such conditions, deterrence is no longer based on fear of punishment, but on fear of losing control.

Algorithmic stability is becoming the new goal of strategic deterrence in the era of artificial intelligence. Authors Jensen, Atalan, and Macias argue that artificial intelligence will play both stabilising and destabilising roles in the future (Jensen, Atalan and Macias 2024). It involves the controlled integration of technology into political and military systems to maintain human superiority in critical decision-making. Instead of threat-based deterrence, the focus is on managing the perception of technological control, and a state that can demonstrate responsible use of artificial intelligence simultaneously builds credibility, trust, and stability. This new form of balance of power is based not on the quantity of weapons, but on the quality of algorithmic transparency and the ability to use technology without escalation. In this sense, algorithmic deterrence represents a fundamental shift in the strategic logic of international relations: from the classic "threat deterrence" to "predictability deterrence." It opens up new space for

smaller states that can become carriers of stability in an algorithmically fragmented world, provided they develop their own models of resilience and digital sovereignty.

SERBIA AND STRATEGIC ADAPTATION IN THE AGE OF ALGORITHM DETERRENCE

In the context of accelerated technological transformation and global competition in artificial intelligence, Serbia faces the challenge of defining its strategic position in a system increasingly governed by the algorithmic logic of power. As a medium-sized state with limited resources but also relatively well-developed institutional capacities in digital infrastructure, Serbia has the opportunity to develop a model of adaptive strategic resilience that combines traditional principles of deterrence with new forms of cognitive and algorithmic security. According to Danilović, smaller states survive not by reciprocally opposing each other, but by embedding their security in systems of predictability and perception (Danilović 2002). The author's claim is perfect for the argument about intelligent resilience, which can also be applied to the example of Serbia.

Serbia is not in a position to compete with large states in developing autonomous platforms based on artificial intelligence technologies. Still, it has the potential to position itself as a regional centre of algorithmic sovereignty and security innovation. For this reason, according to Blagojević, the position of smaller states in the great power system depends on their ability to direct limited resources rationally (Blagojević 2022). The Government of the Republic of Serbia has so far adopted two strategies for the development of artificial intelligence. The first covered the period from 2020 to 2025 and set the institutional framework for the application of AI in the public sector, industry, and security. While the second, the Strategy for the Development of Artificial Intelligence in the Republic of Serbia for the period 2025–2030, ensured that Serbia, as the first in the Southeast European region, made a breakthrough in managing the development of artificial intelligence and positioned itself as a regional leader. Which was also recognised through its leadership of the Global Partnership for Artificial Intelligence (Vlada RS, 011-13287/2024-2). Although this strategy is not directly related to national security, it represents a starting point for building a domestic concept of data-driven deterrence,

a model in which advantage is measured not by the quantity of weapons but by control over digital resources and analytical processes.

A key challenge for Serbia is its reliance on foreign technological platforms and the lack of domestic infrastructure for developing artificial intelligence. Although there are significant infrastructure facilities and projects, they are still not integrated into a single system that functions as a whole. In such a context, strategic adaptation implies developing a hybrid approach that combines limited technological autonomy with smart international positioning. This means that Serbia must use its geopolitical hub as an instrument for balancing between different technological spheres from the East and the West, while at the same time preserving digital independence in critical domains such as data security, infrastructure systems and cognitive security.

In an algorithmic environment, deterrence no longer depends solely on material capacities but also on society's perceptual structures. Similarly, according to Blagojević, the modern strategic concept of extended deterrence has expanded not only to conventional deterrence, but also to the application of economic, technical-technological, informational, cultural and other instruments and means, which has essentially enabled even "small and medium-sized" states to strive to apply them effectively in regional frameworks (Blagojević 2025, 52). Serbia, as a country with a complex media and information space, is in a high-risk zone for cognitive operations and disinformation campaigns that can undermine trust in institutions. Algorithmic deterrence in the Serbian context must therefore also include information defence, i.e., building society's resilience to manipulation, as well as the development of tools to detect and analyse cognitive threats using artificial intelligence.

This implies systemic cooperation among security structures, academia, and the information and telecommunications sector to establish a potential national algorithmic centre that would unite data analysis, crisis management, and strategic communications. Such a centre would not only be a technological but also an epistemological instrument. It could be a tool for integrating knowledge, risk assessment, and decision-making in complex information environments.

In a world where deterrence is based on the predictability of algorithms, smaller states like Serbia must develop their own algorithmic identity, a set of principles, protocols, and values that define

the way a state uses, regulates, and communicates technology. Based on this, four strategic recommendations can be identified:

- Building algorithmic sovereignty – Establishing national data centres and developing domestic models of artificial intelligence for security purposes, with transparent regulation of the use of algorithms in state administration and the defence sector.
- Integrating artificial intelligence into the national security system – By applying predictive analytical tools for risk assessment, disinformation detection and early warning of cognitive threats.
- Regional cooperation and strategic diplomacy – Creating initiatives in the field of digital security and ethical application of artificial intelligence in Southeast Europe, which would allow Serbia to act as a mediator between different technological blocs.
- Education and culture of digital responsibility – Developing educational programs and public awareness on the security aspects of artificial intelligence, to reduce the vulnerability of society to algorithmic manipulations and increase the resilience of the cognitive space.

Serbia could develop its own concept of intelligent deterrence, a model that combines traditional strategic approaches with technological innovation and cognitive resilience. This concept would be based on the idea that deterrence need not be driven by fear, but by trust in a system that is predictable, transparent, and adaptive. Instead of a race for technological dominance, the goal would be for Serbia to demonstrate responsibility, control and stability in the use of artificial intelligence, thereby becoming a factor of predictability in the region. Such a strategy would also represent a practical application of the concept of algorithmic deterrence in a national context, transforming technology into an instrument of stability rather than dependence. Serbia, in this sense, has the potential to become an example of algorithmic statehood in which knowledge, technology and strategic culture intertwine into a unique form of intelligent defence.

CONCLUSION

The development of artificial intelligence and digital technologies has profoundly transformed the notion of security and the nature of deterrence. Instead of the classic dichotomy between offensive

and defensive capabilities, modern security systems increasingly rely on algorithmic power dynamics, in which controlling data, perceptions, and cognitive flows becomes as important as possessing material force. In this context, the concept of total defence takes on a new meaning: it is no longer just the mobilisation of society in the physical sense, but the integration of a nation's entire information, technological, and cognitive potential. As Freedman notes, strategy has always been the imposition of the reason of power; artificial intelligence now calls into question who or what uses that reason (Freedman 2015), further supporting the idea that algorithmic deterrence and its logic must be an active factor in the strategic thought of the 21st century. This principle speaks to the necessity of the interaction between man and algorithm in strategy, and, at the same time, in the strategy of deterrence.

This work shows that the algorithmic logic of resilience is a key pillar of the new strategic paradigm. It is not based on fear of the enemy, but on the system's ability to absorb, analyse, and transform the crisis into an advantage. Resilience, in this sense, becomes intelligent, dynamic, adaptive and predictive. Instead of static deterrence, the modern model implies a continuous evolution of strategic capacities, based on learning algorithms, prediction and cognitive control of information.

For Serbia, as a country with a complex geopolitical environment and limited resources, this process is not only a technological but also a cultural challenge. Serbia's strategic approaches, grounded in the idea of resistance and its historical experience of asymmetric defence, provide fertile ground for the development of the concept of intelligent total defence. It is precisely this concept that allows national security to be defined not only by armed capabilities but also by society's ability to understand, filter, and manage information amid algorithmic instability.

Serbia's key strategic advantage in the 21st century lies in its ability to integrate human and technological intelligence into a single system of deterrence and resilience. This involves building national institutions that combine military, academic, and technological capabilities, developing its own artificial intelligence models, and ethically regulating their application. Such a strategy would not only enhance the state's security but also position Serbia as an active factor in shaping the region's digital security norms.

Ultimately, it can be concluded that the future of defence is intelligent, not just violent. States that develop the ability to predict,

adapt, and become cognitively resilient become new centres of stability in a world of fluid crises. Serbia, if it manages to synthesise its historical experience, technological development, and strategic culture into a coherent doctrine of algorithmic resilience, can become a model of a small state that balances not between forces but between the human mind and artificial intelligence, thereby building its own algorithm of security and the future. States without their own artificial intelligence infrastructure must build a concept of digital self-sufficiency as a prerequisite for strategic autonomy.

REFERENCES

Allen, Gregory C, and Chan Taniel. 2017. *Artificial Intelligence and National Security.* Washington, DC: Center for a New American Security (CNAS).

Barros, Gustavo. 2010. "Herbert A. Simon and the concept of rationality: Boundaries and procedures." *Brazilian Journal of Political Economy* 30 (3): 455–472. doi: 10.1590/S0101-315 72010000300006.

Blagojević, Veljko. 2022. *Moć i sila-Srbija i vojni faktor u međunarodnoj politici.* Beograd: Institut za strategijska istraživanja.

Blagojević, Veljko. 2025. "Srbija i proširenо odvraćanje: iskustva i putokazi." *Politika nacionalne bezbednosti* 28 (1): 41–57. doi: 10.5937/pnb28-56266.

Boratov, Sardor. 2025. "AI, Autonomy, and the New Security Dilemma: Deterrence, Escalation, and Strategic Ambiguity in the Algorithmic Age." *SSRN*. doi: 10.2139/ssrn.5452734.

Brodie, Bernard. 1959. *Strategy in the missile age.* Princeton University Press.

Burrell, Jenna. 2016. "How the machine 'thinks': Understanding opacity in machine learning algorithms." *Big Data & Society* 3 (1): 1–12. doi: 10.1177/2053951715622512.

Coker, Christopher. 2019. "Artificial Intelligence and the Future of War." *Scandinavian Journal of Military Studies* 2 (1): 55–60. doi: 10.31374/sjms.26.

Danilović, Vesna. 2002. *When the Stakes Are High: Deterrence and Conflict among Major Powers*. Michigan: University of Michigan Press.

Freedman, Lawrence. 2004. *Deterrence.* Wiley.

Freedman, Lawrence. 2015. *Strategy: A History.* Oxford: Oxford University Press.

Geist, Edward, and Andrew J Lohn. 2018. "How might artificial intelligence affect the risk of nuclear war?" *Rand Corporation*. https://www.rand.org/content/dam/rand/pubs/perspectives/PE200/PE296/RAND_PE296.pdf.

Horowitz, Michael C. 2010. *The diffusion of military power: Causes and consequences for international politics*. New Jersey: Princeton University Press.

Horowitz, Michael. 2020. "A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence". *NSI.* Last accessed 9 November 2025. https://www.nsiteam.com/speaker-series-event/2020-03-18.

Jensen, Benjamin, Yasir Atalan, and Jose M Macias. 2024. *Algorithmic Stability: How AI Could Shape the Future of Deterrence.* Washington, DC: Center for Strategic and International Studies - CISIS.

Jervis, Robert. 1989. *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon*. Ithaca: Cornell University Press.

Johnson, James S. 2021. "Artificial intelligence and the future of warfare: The USA, China, and strategic stability." *Journal of Strategic Studies* 46 (3): 749–751. doi: 10.1080/01402390.2022.2104255.

Libicki, Martin C. 2009. *Cyberdeterrence and cyberwar.* Santa Monica: RAND Corporation.

Mazarr, Michael J. 2018. *Understanding Deterrence.* Santa Monica: RAND Corporation.

Nye, Joseph S. 2011. *The Future of Power.* New York: Public Affairs.

Pavić, Aleksandar, and Hatidža Beriša. 2024. "Artificial intelligence and national security strategy development: Challenges and perspectives." *Nacionalni interes - časopis za nacionalna i državna pitanja* 48 (2): 53–75. doi: 10.5937/nint48-49980.

Pavić, Aleksandar, and Hatidža Beriša. 2025. "Strategic communications: Theory and development." *Nacionalni interes* 51 (2): 107–129. doi: 10.5937/nint51-58670.

Payne, Kenneth. 2021. *I, warbot: The dawn of artificially intelligent conflict*. Oxford University Press.

Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. New York: W. W. Norton.

Schelling, Thomas C. 1960. *The Strategy of Conflict.* Cambridge: Harvard UP.

Schelling, Thomas C. 1966. *Arms and Influence.* New Haven: Yale University Press.

Schwab, Klaus. 2017. *The Fourth Industrial Revolution.* Crown Currency.

Strange, Susan. 1988. "The persistent myth of lost hegemony: reply to Milner and Snyder." *International Organization* 42 (4): 751–752. doi: 10.1017/S0020818300034056.

United Nations General Assembly [UNGA]. 2024. "UN Doc. A/79/88, Lethal Autonomous Weapons Systems." https://docs.un.org/en/A/79/88.

Vlada Republike Srbije [Vlada RS], Strategija razvoja veštačke inteligencije u Republici Srbiji za period 2025–2030. godina, 05 broj 011-13287/2024-2*,* 10. januar 2025.

Александар Павић*

Војна академија, Универзитет одбране у Београду

АЛГОРИТАМСКО ОДВРАЋАЊЕ: НОВА СТРАТЕГИЈСКА ЛОГИКА У ЕРИ ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ

Сажетак

Рад истражује како развој вештачке интелигенције трансформише класичну стратегију одвраћања. Полазна хипотеза је да ВИ не поништава одвраћање, већ га претвара у динамички, когнитивно и информационо заснован систем у којем се управља перцепцијом, очекивањима и понашањем противника у реалном времену. Уместо традиционалног „баланса страха", настаје алгоритамско одвраћање, засновано на предиктивној моћи алгоритама, аутономним системима и контроли информација. Циљ рада је да понуди аналитички оквир за разумевање овог новог облика стратегије, дефинишући његове кључне димензије и механизме, и да оцени импликације за велике и мање државе, укључујући Србију. Рад истражује како алгоритамско одвраћање функционише у когнитивном, оперативном и епистемолошком смислу: од обликовања перцепције ризика и адаптације у реалном времену, до контроле знања као извора моћи. Методолошки, рад се заснива на квалитативној анализи теоријске и стратешке литературе из области међународних односа, безбедности и војне стратегије, као и на компаративном тумачењу класичних и савремених концепата одвраћања. Концептуално моделовање омогућава илустрацију како алгоритми делују као продужетак људске рационалности у сложеном мултидоменском окружењу. Достигнућа рада показују да алгоритамско одвраћање не замењује класичне моделе, већ их трансформише у систем управљања ризиком, информацијама и перцепцијом. Велике силе користе га за постизање надмоћи у предикцији и контролисању глобалних токова података, док мање државе, попут Србије, могу развити асиметричне стратегије у сајбер и когнитивном домену како би очувале дигитални суверенитет и

* Имејл адреса: aleksandar.pavic@mod.gov.rs; ORCID: 0009-0005-6737-0094.

утицај. Рад такође истиче унутрашње парадоксе алгоритамског одвраћања, попут дилеме транспарентности и непредвидивости, и указује на потребу за новим етичким, технолошким и политичким рамовима у управљању овом стратегијском логиком. Закључно, рад доприноси разумевању еволуције одвраћања у дигиталној ери, пружа концептуални оквир за даљу анализу алгоритамског управљања ризиком и перцепцијом, и омогућава практичне смернице за државе које желе да се адаптирају у свету у коме технологија постаје централни фактор стратешке моћи и стабилности.

Кључне речи: алгоритамско одвраћање, стратегијска логика, вештачка интелигенција, дигитална ера.

* Овај рад је примљен 12. јануара 2026. године, а прихваћен на састанку Редакције 31. марта 2026. године.