Secure Systems Design and Technology Development. Veliko Tarnovo, Bulgaria
Proceedings of the International Scientific Conference, 2026. Volume I, 479-485

A Conceptual Model for Risk Management in Hybrid Threats and Cognitive Attacks

Milko Stefanov
Rakovski National Defense College
Sofia, Bulgaria
m.j.stefanov@rndc.bg

***Abstract*— This report presents an integrated conceptual model for risk management, adapted to the specifics of hybrid threats and cognitive warfare. In the conditions of a changed strategic environment, where the objects of attack are not only physical assets, but also human consciousness and social cohesion, traditional risk assessment models prove to be insufficient. The report analyzes the essence of cognitive attack at the biological, psychological and social levels, considering the influence of technologies such as artificial intelligence and algorithmic sabotage. Qualitative risk indicators are presented through the OODA (Observe, Orient, Decide, Act) loop and NATO's "The House" model. The central place is occupied by the proposed conceptual model for quantitative transformation of cognitive risk, based on institutional vulnerability, media amplification and cognitive resilience of society.**

Keywords*— *hybrid threats, cognitive warfare, risk management, cognitive resilience.

I. Introduction

The modern strategic environment marks a transition from conventional conflicts to multidimensional hybrid strategies operating below the threshold of armed conflict. At the center of this paradigm is cognitive warfare, aimed at paralyzing the will to resist and destroying public consensus by manipulating perceptions. For countries like Bulgaria, located in a geopolitically sensitive region, it is critically important to have an adequate framework for analyzing these asymmetric impacts. The purpose of the development is to define and substantiate an integrated conceptual model for risk management, which allows the transformation of qualitative indicators of cognitive attacks into quantitative values for the needs of strategic planning of national security. The working hypothesis is that the construction of effective national resilience against hybrid threats is possible only by moving from traditional physical metrics to a complex assessment of cognitive risk, uniting institutional readiness, the technological influence of the information environment and the psychological immunity of society.

II. Materials and Methods

To achieve the objectives of the report, a multidisciplinary approach was used, including:

Theoretical analysis and synthesis - study of NATO doctrinal documents and concepts of strategic competitors.

Comparative analysis - comparison between traditional information operations and the essence of cognitive warfare.

Modeling - use of the OODA (Observe, Orient, Decide, Act) cycle to identify qualitative risk indicators and application of the DISARM framework for threat taxonomy.

Assessment methodology - combining the expert assessment method (using a 5-point scale) with statistical processing of sociological data to calculate cognitive resilience.

Systemic approach - analysis of the national security architecture of Bulgaria through the prism of the "three lines of defense" model.

III. Results and Discussion

A. Cognitive warfare – essence and forms of manifestation

Unlike traditional information operations (InfoOps) or psychological operations (PsyOps), which are usually limited in time and aim for short-term penetration of specific messages, cognitive warfare seeks deep, lasting, and often irreversible changes in the cognitive patterns, behavioral dispositions, risk perception, and civic cohesion of the adversary.

Online ISSN 3044-7771
https://doi.org/10.68302/std2026.vol1.46

The NATO Chief Scientist's report on cognitive warfare [1] distinguishes three main levels of influence that hybrid state and non-state actors systematically exploit. The biological level targets the nervous system as the physical and focal basis of thought, emotion and behaviour. Offensive actions here include the induction of physiological stress, sensory overload and the potential use of neurotechnologies to disrupt sleep, concentration and the basic combat capability of personnel. At the psychological level, the focus is on influencing cognitive appraisal, emotional framing and the exploitation of existing human vulnerabilities and cognitive biases. Here, attackers exploit phenomena such as in-group/out-group polarization, confirmation bias and emotional priming to shape individual and collective attitudes, beliefs and judgements in their favour. The social level is the macro level of cognitive warfare, where the goal is to destroy social cohesion. Attacks are aimed at undermining trust in democratic authorities, delegitimizing institutions, eroding alliance commitments, and dividing society by instilling chronic doubt and cynicism.

The forms of manifestation of cognitive attacks are extremely diverse, adaptive and strongly determined by the speed of technological development. Artificial intelligence (AI) acts as an accelerator and multiplier of the power of cognitive attack. It allows the creation of the so-called synthetic credibility on an industrial scale. This is achieved by generating deep fakes, creating synthetic experts with fake academic profiles and distributing forged evidence that is visually and logically indistinguishable from reality for the average user. Algorithmic sabotage is another advanced form of manifestation. In it, the hybrid adversary does not limit itself to creating disinformation, but actively interferes with information architectures and flows. This includes "poisoning" machine learning databases (data poisoning / LLM poisoning), which introduces adversarial noise into the algorithms that societies, intelligence and military structures rely on for situational awareness [2]. The goal is to compromise the integrity of command and control systems by generating false recommendations or distorting the decision-making process.

The experience and doctrinal documents of NATO's strategic competitors demonstrate the systematic and resource-rich integration of these methods. The Russian Federation uses a complex combination of communication technologies, cyberattacks and perception manipulation operations to build cognitive traps. They aim to provoke either strategic overreaction or complete paralysis and indecision among decision-makers in the Alliance [3]. The Russian approach to hybrid warfare actively exploits the discrepancy in reaction thresholds of democratic societies. The People's Republic of China conceptualizes cognitive warfare through its Three Warfares strategy, which involves the use of public opinion, psychological operations and lawfare to achieve political and military victories without the need for open kinetic confrontation [4]. As technology advances, this strategy has evolved into the concept of intelligent warfare [5].

The European Centre for Countering Hybrid Threats (EU Hybrid CoE) conceptualizes these processes in the civilian sector as cognitive intrusion. It is defined as the deliberate manipulation of collective mental processes that relies on nurturing a culture of disinformation and inflaming negative emotions of anger to circumvent normal channels of democratic political expression and provoke political violence [6].

B. Qualitative indicators in assessing the risk of cognitive attacks

Assessing risk in the cognitive domain requires moving away from traditional metrics, such as media coverage, impressions, or share volume, to more in-depth metrics that focus on the quality, speed, and integrity of decision-making. To identify which quality metrics actually impact risk, the academic and military communities, and NATO research teams in particular, have developed several specialized theoretical frameworks [7].

One of the most applicable and comprehensive theoretical models for analyzing cognitive vulnerabilities is the interactive framework based on the OODA loop. This framework views cognitive warfare not as a one-way media campaign, but as an adaptive, continuous race between attacker and defender to gain advantage in the decision-making cycle. The model identifies qualitative indicators over time horizons – acute effects that lead to immediate degradation within minutes or days, and chronic effects that accumulate over months and years [8].

Risk assessment through the OODA loop allows the identification of specific offensive effects and corresponding countermeasure indicators at each decision-making phase:

Observe – In this initial phase, the attacker seeks information saturation, noise generation, and selective deception to disrupt shared situational awareness. Key risk indicators here include the ability of institutions to correctly assess the credibility of information, the degree to which accurate observation is maintained, and susceptibility to information overload.

Orient – Here, efforts are directed toward emotional framing, exploitation of prejudice, and erosion of institutional trust through the creation of ambiguity. Resilience is measured by the capacity for collective meaning-making of events and the effectiveness of calibrating trust between official channels and unauthorized sources.

Decide – The attack aims to induce strategic hesitation, confusion, undue overconfidence, or complete paralysis of will. Risk indicators include decision latency (delay compared to the norm), frequency of canceled orders, and inadequate changes in confidence thresholds when making choices.

Act – In the final phase, the adversary provokes premature, inappropriate, or intentionally passive actions that serve its interests. Indicators of breakthrough are the degree of deviation from the initial operational concept and the missed time windows for tactical or political execution. This transition from qualitative to quantitative dimensions is essential for transforming the model into a working tool for national security.

In parallel to the OODA model, NATO's Science and Technology Organisation (STO) developed "The House Model" for mitigating and responding to cognitive warfare. This model maps human cognition through seven key domains that function as metrics for vulnerability and resilience. The model consists of three vertical Pillars and four horizontal intersecting Bars:

- Pillars of vulnerability – cognitive neuroscience, cognitive and behavioural science, social and cultural science. The risk increases proportionally to the extent to which an adversary can penetrate these scientific domains and turn them against the defender.

- Horizontal multipliers – situational awareness and sensemaking, cognitive effects, modus operandi, and technological force multipliers [9].

Any registered weakness or breach in these pillars represents a primary qualitative indicator that the relevant organization or society is at high risk of a successful cognitive hack.

For operational intelligence purposes, a NATO research team integrated the General Morphological Analysis (GMA) and the Disinformation Risk Analysis and Management Framework (DISARM) to create a structure of Indicators and Warnings (I&W). The DISARM framework, based on the logic of the MITRE ATT&CK cyber model, catalogs specific tactics and techniques that serve as qualitative diagnostic indicators. The DISARM Red Framework identifies indicators in the planning, execution, and evaluation stages of a campaign. The main qualitative indicators here are the creation of clusters of inauthentic social media profiles, the generation of micro-targeted clickbait, the systematic exploitation of historical or local traumatic narratives, and temporal anomalies in cyberspace [7]. The main challenge facing classical risk management, however, lies in the fact that these listed indicators – decision latency, institutional trust, emotional reactivity, social cohesion – are extremely qualitative, fuzzy, subjective and difficult to measure with conventional methods. To be useful in the defense planning process, they must undergo a precise quantitative transformation.

C. A conceptual model for risk management in hybrid threats and cognitive attacks

The national security architecture of the Republic of Bulgaria is a complex system of institutions, regulations, strategic documents and operational procedures, the purpose of which is to protect national interests. It is built on the principles of democratic control, the rule of law and the separation of powers, but at the same time it is faced with the imperative need to increase its flexibility and speed of response. A fundamental element in the management of national security in the Republic of Bulgaria is the cycle of strategic planning, monitoring and accountability. This cycle is most clearly implemented through the annual report on the state of national security, which is prepared by the executive branch and submitted for consideration and approval by the legislative body. This document performs several critical functions simultaneously – it serves as a diagnostic tool for identifying existing vulnerabilities, functions as a platform for accountability for actions taken to date and represents a roadmap for future initiatives. The report aims to present a detailed description of the measures that state-level institutions plan to take in the coming period in order to proactively counter the diagnosed challenges and exploit emerging opportunities [10]. This planning requires an extremely high level of coordination between the different ministries and agencies. The executive branch in Bulgaria is responsible for the operational counteraction to malicious influence, which requires a continuous analysis of the effectiveness of national security policy. Despite the existence of this structured planning cycle, a practical analysis of institutional interaction reveals certain structural deficiencies. Fragmentation of efforts is often observed, in which different departments approach hybrid threats through the prism of their narrow specialization, without taking into account the full scale of the cognitive impact. For example, technical agencies focus exclusively on the integrity of networks, while information directorates analyze the media environment, but the synthesis between these two data streams is often not fast enough to prevent the deployment of a complex cognitive attack. In the context of risk management and ensuring resilience against hybrid attacks, Bulgarian state institutions adapt and apply the conceptual framework of the three lines of defense [11]. This framework is designed to provide a structured approach to risk management through a clear allocation of responsibilities:

- First line – operational managers, who own the processes and are directly responsible for managing risks in their daily activities.

- Second line – oversight and risk management functions, which set policies, standards and carry out monitoring.

- Third line – independent internal and external audit, which provides an objective assessment of the effectiveness of the first two lines.

Despite its theoretical robustness, the implementation of this framework in the Bulgarian state administration often encounters three common problems identified in contemporary security architecture research. These problems include the lack of clear communication between lines, the blurring of responsibilities in the event of complex, multi-domain incidents, and the insufficient expertise of the first line to recognize subtle cognitive or social engineering attacks. These structural weaknesses create blind spots in the security architecture through which foreign malicious influence can penetrate and spread before being detected by monitoring systems. In order for the Bulgarian security architecture to be effective in the context of cognitive impacts, it is necessary to analyze its current critical vulnerabilities. Despite the declared intentions in the annual reports on the state of national security, the system continues to demonstrate weaknesses that make it susceptible to hybrid impacts. These vulnerabilities can be conceptualized through the academic formula of cognitive risk

$$R_c = \frac{V_i . A_m}{C_r} \quad (1)$$

where:

Rc – cognitive risk;

Vi – institutional vulnerability;

Am – algorithmic and media amplification of the attack;

Cr – cognitive resilience of society.

The cognitive risk formula is a conceptual model. It does not use a specific mathematical law with precise numerical dimensions, but is an adaptation of the classical academic equation for assessing risk in security and management, applied specifically to the mechanisms of cognitive warfare. In Bulgarian practice, in internal audit and risk management methodologies in state administration, the basic formula for quantitative assessment is most often used:

$$Risk = \text{Probability} \times \text{Impact} \quad (2)$$

In more detailed national security models, this concept is often expanded to the logic:

$$Risk = \frac{\text{Vulnerability} \times Threat}{\text{Threat mitigation capacity}} \quad (3)$$

Thus, the formula serves as a theoretical model that translates standard risk management principles into the language of modern information and psychological threats. Since this is a qualitative model, the values for individual variables are not calculated mathematically, but are obtained through a qualitative strategic analysis of processes in the state and society.

Institutional vulnerability (Vi) is assessed based on the existing structural deficiencies in the state architecture. It is high when problems such as poor implementation of the three lines of defense framework in the administration and lack of capacity for preventive identification and neutralization of latent vulnerabilities in digital systems are observed.

Algorithmic and media amplification of the attack (Am) reflects the role of the information environment as a catalyst for the threat. Its value increases when the media ecosystem uncritically reproduces the narratives of the attacking actors, leading to distorted perceptions of systemic vulnerabilities, increased public anxiety, and erosion of institutional trust.

Cognitive resilience of society (Cr) is measured by the ability of society to resist psychological operations. Resilience is low when certain groups are highly susceptible to manipulations such as linguistic hacking, in which emotionally charged keywords are injected to activate historical traumas and incite civil disobedience.

This model illustrates the strategic logic that in order to reduce the overall cognitive risk to national security, Bulgarian institutions must simultaneously minimize their own structural weaknesses and the influence of the disinformation environment, while simultaneously building high psychological immunity in society.

To transform qualitative assessment into quantitative, the most adequate results would be obtained by a combined approach that integrates the method of expert assessments and sociological surveys. The method of expert assessments is considered one of the most up-to-date and reliable methods for deriving quantitative values based on primary qualitative information. In the practice of Bulgarian institutions, this method is often applied through the use of a 5-point rating scale (from 1 - "very low", to 5 - "very high"), through which specialists assess the effectiveness and interaction in a given system. Institutional vulnerability (Vi) can be assessed by independent auditors and security experts, who can give a numerical assessment of structural deficiencies. Media amplification (Am) can be assessed by open source analysts (OSINT) and communication experts, who can score the scale of dissemination of a given disinformation. To increase accuracy, the expert assessment method can be successfully combined with the conduct of simulation models and exercises that test how employees respond to crises, allowing for a more objective assessment of their actions. Surveys and statistical processing are applicable to the calculation of the denominator – the cognitive resilience of society (Cr), where expert assessments are not sufficient. Here, the application of nationally representative surveys is absolutely necessary. Through statistical processing of citizens' responses, accurate numerical coefficients can be derived, reflecting the levels of public trust in institutions, susceptibility to fake news and emotional reactivity to hybrid narratives. A similar approach to converting qualitative indicators into numerical values has already been established in the Bulgarian administration in the management of other types of risk, in which the level is calculated by multiplying the numerical assessment of the severity of the impact by the probability of occurrence. The resulting result falls within predefined numerical limits for low, medium or high risk levels. Similarly, by creating a standardized matrix based on expert assessments (scale from 1 to 5) and sociological data, the cognitive risk formula can be transformed from a conceptual model into a fully functional mathematical tool for the executive branch.

For the purpose of practical testing and to ensure the compatibility of the model with established practices in internal audit, the variables are assessed using the expert assessment method, combined with open source intelligence (OSINT) and representative sociological surveys. Table I shows how a standardized 5-point scale (from 1 to 5) is used. For the variables Vi and Am, a score of 1 means the least pronounced factor (best case scenario) and 5 means the most pronounced factor (worst case scenario). Conversely, for the variable Cr (resilience), a score of 1 means the least pronounced factor (worst case scenario) and a score of 5 means the most pronounced factor (best case scenario).

TABLE I: FIVE-POINT SCALE FOR DETERMINING VARIABLE VALUES IN RISK ASSESSMENT

Rating	Vi	Am	Cr
1	Optimal interdepartmental coordination,	Isolated and fragmented cases of	A total collapse of the social contract,

Rating	Vi	Am	Cr
	available AI tools for real-time detection, an effective strategy.	disinformation, relying only on organic reach, complete lack of AI.	organized mass protests generated entirely on the basis of disinformation.
2	Availability of basic procedures and action plans, but with a recorded slow response; partial coordination.	Localized campaigns, using traditional troll farms with easily recognizable handwriting.	Systemic distrust of authority, extremely high political polarization, dominant public apathy
3	Presence of serious institutional blind spots, fragmented implementation of controls.	Synchronized campaigns operating across multiple platforms simultaneously, successful micro-targeting.	A divided society, tangible presence of historical traumas, susceptibility to conspiracy theories.
4	Deep structural fragmentation, lack of an up-to-date national strategy, dysfunctional regulators.	Large-scale and resource-secured use of GenAI (deepfakes), automated saturation of the info-space.	Stable trust in key government bodies, moderate public reaction to emotional manipulation.
5	Complete institutional paralysis, inability to make decisions, penetration of foreign influence.	AI-Enhanced Reflexive Control (AIRC), complete takeover of the national information flow.	Extremely high public trust in institutions and the media, strong social consensus on strategic topics.

The calculation of the ultimate risk (Rc) generates a numerical value ranging from 0.2 to 25, as shown in Table II. This scale allows the categorisation of risk into four strategic quadrants that dictate the required response.

TABLE II: RISK LEVEL AND RESPONSE

Scope (Rc)	Risk level	Strategic response	Main subjects
0,2-0,4	Low	Standard monitoring and data collection.	First line of defense.
4,1-10	Moderate	Preventive communication campaigns and enhanced surveillance.	Compliance functions.
10,1-18	High	Interdepartmental coordination and active countermeasures.	Independent audit and specialized units.
18,1-25	Critical	Emergency restructuring and protection of national security.	State leadership and Security Council.

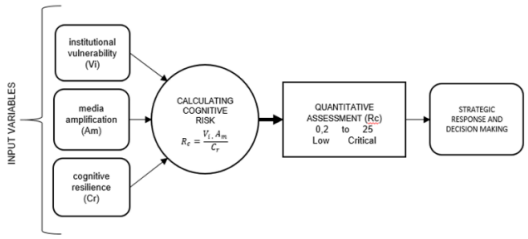

Fig. 1. Schematic representation of the cognitive risk (Rc) assessment model.

Fig. 1 presents the logical and mathematical relationship between the input data, the calculation process and the strategic result. The main interrelationships it highlights are the directly proportional dependence of cognitive risk (Rc) on institutional vulnerability (Vi) and media amplification (Am), which work in synchrony. If institutions are weak and the media campaign is large-scale, the risk increases exponentially. An inversely proportional dependence on cognitive resilience (Cr) is presented as a regulator. Even with high vulnerability and a strong attack, a high level of resilience of society can drastically reduce the final risk. Subjective categories, such as media influence or public fear, are transformed into a specific numerical value, which is not an end in itself, but a basis for decision-making.

In the proposed conceptual model, the values of the variables are not summed, but interact multiplicatively, which reflects the asymmetric nature of cognitive warfare. When we define a value of 4 (high) for Am, this means that the information ecosystem uncritically reproduces foreign narratives that distort perceptions of systemic vulnerabilities. Advanced technologies such as AI are used to generate synthetic credibility and deep fakes on an industrial scale. Algorithmic selection of social networks creates “echo chambers” that accelerate the spread of disinformation. When this amplification is combined with 2 (low) for Cr, a defense collapse effect is observed. Since Cr is in the denominator of formula (1), low resilience sharply increases the ultimate risk. At a value of 2, society is highly susceptible to linguistic hacking, in which emotionally charged keywords activate historical traumas and incite civil disobedience. The combination of massive pressure (value 4) and weak immunity (value 2) leads to cognitive intrusion, which bypasses normal channels of democratic expression and provokes political violence.

In order to convert qualitative responses from sociological surveys into quantitative coefficients for formula (1), it is necessary to apply a statistical normalization process. This ensures that the data from citizens are compatible with the expert assessment (scale 1–5) for Vi and Am. Questions from nationally representative surveys measuring trust in institutions, the ability to recognize fake news, and levels of social polarization are used. Each response is assigned a weight. For example, in a question about trust in the media, the response “I completely trust” receives the highest sustainability weight, and “I do not trust at all” the lowest.

The percentage values (P) are transformed into an index on a scale of 1–5 by linear interpolation:

$$Iindex = 1 + \left(\frac{P - Pmin}{Pmax - Pmin}\right) * 4 \quad (4)$$

where:

Iindex – normalized value;

P – current value of the indicator;

Pmin – minimum value of the indicator;

Pmax – maximum value of the indicator.

The results are distributed in a matrix where:

1.00 – 1.80: Very low resilience.

1.81 – 2.60: Low resilience.

2.61 – 3.40: Moderate resilience.

3.41 – 4.20: High resilience.

4.21 – 5.00: Very high resilience.

In order to test the functionality of the conceptual model in operational conditions, the campaign against the adoption of the euro and the full economic integration of the Republic of Bulgaria into the eurozone during the period 2024–2025 was chosen as the object of the approbation [12]. This example is relevant for the purposes of the study, as it provides an illustration of the convergence between the three levels of cognitive warfare – biological, psychological and social. It also demonstrates the use of technological force multipliers, such as generative artificial intelligence and deep fakes, to manipulate mass consciousness in a member state of the European Union. Bulgaria's accession to the eurozone, initially planned for the beginning of 2024 and subsequently postponed to a later stage, has been defined as the top strategic priority of the state. During the observed period, Bulgaria is in a state of permanent political turbulence and institutional instability. Holding seven parliamentary elections in just four years has generated mass electoral apathy, deep disillusionment with the political elite, and chronic fatigue in society [13]. Disinformation researchers and analysts categorically warn that because of these factors, Bulgaria has one of the most permissive and susceptible information environments for undemocratic malicious manipulation in all of Europe [14].

The approbation process requires a quantitative assessment of the three independent variables based on the collected empirical data, expert analyses and sociological surveys, using the defined 5-point scale.

Institutional Vulnerability Score (Vi) = 4 (high). Despite the repeatedly declared intentions in the annual national security reports and the existence of an official communication strategy of the Ministry of Finance and the Bulgarian National Bank on the euro, institutional vulnerability remains extremely high. The score of 4 is justified by the following systemic deficits. The Bulgarian state has not yet adopted a comprehensive national strategy to counter disinformation. This indicates a lack of strategic leadership at the highest level. There has been administrative and political delay in the appointment and empowerment of a national Digital Services Coordinator (DSC). This institution is the critical element for imposing regulations and sanctions on large technology platforms under the EU Digital Services Act (DSA). Without it, the state is powerless to demand the removal of coordinated hybrid networks [15]. Although the Coordination Council plans a large-scale information campaign in three phases (preparatory, substantive and de facto), its approach remains largely traditional and reactive. It is focused on conventionally informing about the benefits and refuting speculation, but institutions categorically lack the technological capacity and procedures to detect and neutralize AI-Enhanced Reflexive Control (AIRC) in real time.

Algorithmic and Media Amplification (Am) Score = 5 (critical). The distribution environment in Bulgaria has evolved from a simple communication space into a perfect catalyst for the hybrid threat. The score of 5 reflects the maximum degree of technological exploitation, justified by the proven implementation of artificial intelligence for industrial generation of micro-targeted content and textual deepfakes in Bulgarian, which successfully bypass international platform detectors [16]. The specificity of social media algorithms, which systematically prioritize content that evokes strong negative emotions, leads to immediate and viral sharing of disinformation long before any official fact-checking is possible. There is the adaptive tactic of using influencers and proxies, who do not directly create fake news, but repackage disinformation narratives in accessible formats and legitimize them to their followers [17].

Cognitive Resilience (Cr) score = 1 (very low). Cognitive resilience is assessed as extremely weak and deficient. Sociological research by the Ministry of Finance itself reveals a society that is deeply polarized and uncertain on the most important strategic issue. Despite extensive communication, while 51% of citizens support the adoption of the euro, as many as 45% are categorically against it [12]. This fragile balance provides ideal ground for hybrid exploitation. There is a systematic, deep-rooted distrust not only of the political class, but also of traditional public media. Bulgarian National Television and Bulgarian National Radio have registered a decline in trust levels in recent years, accompanied by accusations of politicization and deliberate concealment of documentaries about Russian disinformation [15]. Emotional reactivity is extremely high. Justified fear of impoverishment is combined with irrational fears of loss of identity, making the population susceptible to organized provocations. These fears materialize into real physical actions, including radical protests and attacks on diplomatic missions [18].

Applying the values of the scores for Vi, Am and Cr in formula (1), we obtain a score of 20 out of a maximum possible 25. This positions the risk categorically in the upper echelon of the "Critical Risk" quadrant. This far exceeds the tolerance threshold of conventional internal control and national security systems. The result proves that the current approach of institutions, based solely on the distribution of press releases, information brochures and traditional PR, is doomed to failure when faced with AI-

driven hybrid attacks. In order to reduce the overall risk below the critical threshold, the state faces insurmountable barriers in two of the three directions. It cannot rely on a quick and magical change in the mentality and historical burden of society, nor does it have the geopolitical power to completely turn off the algorithms of global technology platforms. Therefore, the only mathematically possible and operationally feasible way out is to focus the main effort on a drastic, uncompromising reduction of institutional vulnerability. This imperative requires a radical change in the actions of institutions, which can only be realized through a complete revision of the "three lines of defense" model.

IV. Conclusions

The proposed conceptual model for cognitive risk management allows for the transition from a purely descriptive to a predictive approach in countering hybrid threats. The integration of quantitative indicators (Vi, Am, Cy) transforms subjective perceptions of security into measurable quantities, which is critical for making timely strategic decisions. The analysis of the case study with the campaign against the euro confirms that the main weakness is not only in the intensity of the attack, but in the low cognitive resilience and institutional vulnerability. Future development of the model should focus on automating monitoring through artificial intelligence, which will allow for real-time cognitive risk assessment and more effective calibration of strategic communications.

REFERENCES

[1] J. Giordano, "Cognitive Warfare 2026: NATO's Chief Scientist Report as Sentinel Call for Operational Readiness," Jan. 6, 2026. [Online]. Available: https://inss.ndu.edu/Research-and-Commentary/View-Publications/Article/4371195/cognitive-warfare-2026-natos-chief-scientist-report-as-sentinel-call-for-operat/. [Accessed: Mar. 19, 2026].

[2] R. D. Whiteside, "Narrative Manipulation, Malinfluence Operations, and Cognitive Warfare Through Large Language Model Poisoning with Adversarial Noise," Oct. 1, 2025. [Online]. Available: https://mipb.ikn.army.mil/issues/jul-dec-2025/narrative-manipulation-malinfluence-operations-and-cognitive-warfare-through-large-language-model-poisoning-with-adversarial-noise/. [Accessed: Mar. 19, 2026].

[3] B. Catena, O. Ditrych, and N. Kovalčíková, "Smoke and mirrors: Building EU resilience against manipulation through cognitive security," Oct. 7, 2025. [Online]. Available: https://www.iss.europa.eu/publications/briefs/smoke-and-mirrors-building-eu-resilience-against-manipulation-through-cognitive. [Accessed: Mar. 19, 2026].

[4] J. Baughman, "How China Wins the Cognitive Domain," Jan. 2023. [Online]. Available: https://www.airuniversity.af.edu/Portals/10/CASI/documents/Research/CASI%20Articles/2023-01-23%20How%20China%20Wins%20the%20Cognitive%20Domain.pdf. [Accessed: Feb. 12, 2026].

[5] K. Takagi, "The Future of China's Cognitive Warfare: Lessons from the War in Ukraine," *Hudson Institute*, Jun. 22, 2022. [Online]. Available: https://www.hudson.org/national-security-defense/the-future-of-china-s-cognitive-warfare-lessons-from-the-war-in-ukraine. [Accessed: Feb. 12, 2026].

[6] M. Lebrun, "Hybrid CoE Strategic Analysis 33: Anticipating cognitive intrusions: Framing the phenomenon," *European Centre of Excellence for Countering Hybrid Threats*, Jul. 2023. [Online]. Available: https://www.hybridcoe.fi/publications/hybrid-coe-strategic-analysis-33-anticipating-cognitive-intrusions-framing-the-phenomenon/. [Accessed: Mar. 19, 2026].

[7] R. Burda, P. Buvarp, D. Brown, and M. Fitzpatrick, "Cognitive Warfare Indicators and Warnings Identification Framework," *NATO STO Meeting Proceedings*, Apr. 15, 2025. [Online]. Available: https://publications.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-HFM-377/MP-HFM-377-09.pdf. [Accessed: Mar. 19, 2026].

[8] B. Rushing, W. Hersch, and S. Xu, "Cognitive Warfare: Definition, Framework, and Case Study," Mar. 6, 2026. [Online]. Available: https://arxiv.org/pdf/2603.05222. [Accessed: Mar. 19, 2026].

[9] J. M. Blatny and Y. R. Masakowski, "Mitigating and responding to Cognitive Warfare," Jul. 4, 2023. [Online]. Available: https://www.google.com/search?q=https://publications.sto.nato.int/publications/STO%2520Technical%2520Reports/STO-TR-HFM-ET-356/$$TR-HFM-ET-356 ALL.pdf](https://publications.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-HFM-ET-356/$$ TR-HFM-ET-356-ALL.pdf). [Accessed: Mar. 19, 2026].

[10] Council of Ministers of the Republic of Bulgaria, "Report on the State of National Security of the Republic of Bulgaria in 2024," 2024. [Online]. Available: https://www.parliament.bg/bills/51/51-602-00-7.pdf. [Accessed: Mar. 9, 2026].

[11] The Institute of Internal Auditors, "The IIA's Three Lines Model," Jul. 2020. [Online]. Available: https://www.theiia.org/globalassets/site/about-us/advocacy/three-lines-model-updated.pdf. [Accessed: Mar. 19, 2026].

[12] E. Maclachlan, "Bulgaria prepares to join eurozone amid fears of Russian-backed disinformation," *The Guardian*, Oct. 28, 2025. [Online]. Available: https://www.theguardian.com/world/2025/dec/28/bulgaria-europe-eurozone-russia-disinformation-economy. [Accessed: May 8, 2026].

[13] R. Margova, "The disinformation landscape in Bulgaria," *EU DisinfoLab*, Nov. 17, 2025. [Online]. Available: https://www.disinfo.eu/publications/disinformation-landscape-in-bulgaria/. [Accessed: May 8, 2026].

[14] E. Nilsson-Julien, "Bulgaria prepares for disinformation ahead of snap elections," *Euronews*, Apr. 16, 2026. [Online]. Available: https://www.euronews.com/my-europe/2026/04/16/bulgaria-prepares-for-disinformation-ahead-of-snap-elections. [Accessed: May 8, 2026].

[15] V. Petrova, "BULGARIA: FRAGILE MEDIA FREEDOM PROGRESS AT RISK OF BACKSLIDING WITHOUT URGENT REFORM," *International Press Institute (IPI)*, Oct. 2025. [Online]. Available: https://ipi.media/wp-content/uploads/2025/10/Bulgaria-Report-2025.pdf. [Accessed: May 8, 2026].

[16] J. Udo-Udo Jacob, "Manufacturing Dissent: The Kremlin's Digital Playbook for Destabilizing Bulgaria," *AUBG Today*, Jun. 19, 2025. [Online]. Available: https://www.aubg.edu/aubg-today/manufacturing-dissent-the-kremlins-digital-playbook-for-destabilizing-bulgaria/. [Accessed: May 8, 2026].

[17] S. Nikolaeva, "Mapping Disinformation Narratives in Bulgaria: Second Working Group Report," *American University in Bulgaria (AUBG)*, May 2025. [Online]. Available: https://www.aubg.edu/wp-content/uploads/2025/07/Bulgaria-Working-Group-Report_May-2025_Mapping-Disinformation-Narratives-in-Bulgaria.pdf. [Accessed: May 8, 2026].

[18] T. Benakis, "Parliament endorses Bulgaria's adoption of the Euro amid a harsh disinformation campaign," *European Interest*, Jun. 9, 2025. [Online]. Available: https://www.europeaninterest.eu/parliament-endorses-bulgarias-adoption-of-the-euro-amid-a-harsh-disinformation-campaign/. [Accessed: May 8, 2026].