Original Scientific Paper/Original naučni rad UDC/UDK: 004.8:004.738.5.056(497-15)
Paper Submitted/Rad primljen:31.12.2025.
Paper Accepted/Rad prihvaćen:10.01.2026.
DOI: 10.5937/SJEM2601119S

Улога вештачке интелигенције у јачању хибридних претњи на Западном Балкану

Marko Savković[1], Igor Novaković[2]

[1]ISAC Fund, Belgrade, Serbia. E-mail: marko.savkovic@isac-fund.org
[2]ISAC Fund, Belgrade, Serbia. E-mail: igor.novakovic@isac-fund.org

Сажетак: Вештачка интелигенција (ВИ) све више утиче на безбедносни пејзаж Западног Балкана, региона који је обележен крхким политичким окружењем и оспораваном управом. У овом раду се истражује како ВИ појачава хибридне претње као што су сајбер напади, олакшава мешање у изборни процес и омогућава дестабилизацију од стране државних и недржавних актера. Путем машинског учења и аутоматизованих система, злонамерни актери могу да креирају кампање дезинформација, искористе рањивости у дигиталној инфраструктури и манипулишу јавно мњење. Истраживање поставља поменуте појаве у шири контекст хибридних претњи, наглашавајући међудејство између технолошких иновација и системских слабости демократских институција. Анализом недавних случајева и нових трендова, у раду се тврди да безбедносни изазови које покреће ВИ на Западном Балкану нису само техничке природе, већ дубоко политички, те захтевају координисане одговоре који комбинују мере сајбер безбедности, регулаторне оквире и јачање отпорности на друштвеном и институционалном нивоу.

Кључне речи: вештачка интелигенција (ВИ), сајбер безбедност, изборне интерференције, хибридне претње, кампање дезинформација, Западни Балкан, политичка стабилност

AI's Role in Amplifying Hybrid Threats in the Western Balkans

Abstract: Artificial Intelligence (AI) is increasingly shaping the security landscape of the Western Balkans, a region marked by fragile political environments and contested governance. This paper examines how AI amplifies hybrid threats such as cyberattacks, facilitates election interference, and enables destabilization by both state and non-state actors. Through machine learning and automated systems, malicious actors can create disinformation campaigns, exploit vulnerabilities in digital infrastructure, and manipulate public opinion. The study places these developments within the broader context of hybrid threats, highlighting the interplay between technological innovation and systemic weaknesses in democratic institutions. By analyzing recent cases and emerging trends, the paper argues that AI-driven security challenges in the Western Balkans are not merely technical but deeply political, requiring coordinated responses that combine cybersecurity measures, regulatory frameworks, and resilience-building at societal and institutional levels.

Keywords: artificial intelligence (AI), cybersecurity, election interference, hybrid threats, disinformation campaigns, Western Balkans, political stability

1. Introduction

Hybrid threats are reshaping Europe's security landscape. They are blurring lines between war and peace, internal and external security, and state and non-state action (Hoffman, 2007; NATO, 2014). Strategies employed combine cyber operations, disinformation, political influence, economic pressure, and legal or institutional exploitation all while staying below the threshold of open conflict (EU Commission, 2016; Kofman & Rojansky, 2015). By expanding the reach and effectiveness of these tactics, AI has become their key enabler (Brundage et al., 2018).

Numerous and recent examples show how through machine learning hostile actors produce and spread disinformation quickly, tailor their messages, exploit digital weaknesses of intended targets, and manipulate public debate (Buchanan, 2020; Nemitz, 2018). In effect, trust, legitimacy and democratic integrity of one society

institutions are eroded (Risse, 2020). Yet within the policy and scholarly discussion, what we understand as deeply political nature of AI-enabled hybrid threats remains overlooked.

This new dynamic is clearly present in the Western Balkans, a region with weak democratic institutions, contested governance, polarization, and uneven progress toward Euro-Atlantic integration (Bieber, 2018; Keil & Perry, 2015). At the same time, societies and their respective governance systems have been digitalizing rapidly, often without strong regulation or safeguards. This has heightened exposure to manipulation (Freedom House, 2023) and has created favorable conditions for state and non-state actors to influence politics by using AI tools.

Within this paper we examine how AI has amplified hybrid threats in the Western Balkans, focusing on cyber operations, election interference, and disinformation. In our understanding AI is a "force multiplier" that, rather than creating new types of hybrid activity, exploits existing vulnerabilities (Brundage et al., 2018). It increases the speed, reach, precision, and deniability of political interference. Thus, rather than being merely technical, threats are made fundamentally political (Arkan, 2025).

The paper contributes to wider security debates by linking hybrid-threat theory (Hoffman, 2007) with emerging research on AI, highlighting Western Balkan-specific conditions, and calling for responses that combine cybersecurity, regulation, institutional reform, and societal resilience (Nemitz, 2018; Risse, 2020).

Methodologically, it uses a conceptual policy-analysis approach (Browne et al., 2019), drawing on cases and observable trends rather than full empirical testing - appropriate given fast-moving AI technologies and the opaque nature of hybrid operations (Buchanan, 2020). The aim is to identify how technological innovation interacts with political vulnerability.

2. Conceptual framework

Rather than viewing AI as a new or separate threat, this analysis places it within existing hybrid strategies, highlighting how AI strengthens political, informational, and cyber tactics already present in modern conflict. For instance, Romansky et al. (2024) argue that emerging hybrid threat strategies increasingly rely on "exploiting economic dependencies and manipulating societal polarization to undermine state resilience". Furthermore, malicious actors have weaponized digitalization and distorted information environments.

As for hybrid threats, we accept their common definition as "coordinated activities that combine military and non-military tools, involve both state and non-state actors, and use a mix of overt and covert methods to pursue strategic aims" (EU Commission, 2016; NATO, 2014). As such, we find them in the "grey zone between peace and war", exploiting legal and institutional gaps to achieve political effects, but without escalation (Mazarr, 2015). "Core goal" is not in territorial control but shaping perceptions, weakening institutions, and eroding social cohesion (Kofman & Rojansky, 2015), gradually, over time.

Hybrid threats have three main features. They are multidimensional, combining cyber operations, information manipulation, and economic pressure. They rely on plausible deniability, complicating attribution (who has done it?) and collective action (what can be done about it?). And they are highly context-specific, exploiting media fragmentation, polarization, and limited institutional capacity (Hoffman, 2007; Risse, 2020). These traits make them especially effective in regions with weak democratic and institutional consolidation or contested statehood (as in the case of Kosovo), such as Western Balkans.

Hybrid tactics are not new: propaganda, subversion, and covert influence have long been part of international politics. They "simply reflect long-standing methods of influence and coercion seen throughout earlier conflicts" (NATO, 2024). What distinguishes contemporary hybrid threats is the integration of digital technologies, which greatly increases the speed, reach, and coordination of such activities (Buchanan, 2020). As a result, modern strategies increasingly target the informational dimensions of security, challenging traditional, state-centric and military-focused ideas of threat and defense (Nemitz, 2018).

3. Artificial intelligence as a force multiplier in hybrid threats

AI is not a single technology but a set of methods. Machine learning, natural language processing, and automated decision-making can support many kinds of hybrid activity (Brundage et al., 2018). Which is why in this paper AI is not understood as a standalone threat, but a "force multiplier" (Hynes et al., 2025) that expands the scale, effectiveness, and adaptability of operations.

AI strengthens hybrid tactics in three main ways. First, it enables large-scale automation. Tasks such as content generation, data analysis, or cyber reconnaissance, previously requiring extensive human labor, can now be carried

out quickly and cheaply (Buchanan, 2020). In disinformation campaigns, AI systems generate tailored content, boost selected narratives, and fine-tune messages based on user reactions, lowering barriers to sustained influence operations (Ryan-Mosley, 2023). Second, AI increases speed and adaptability. Machine learning tools process big datasets in real time, detect emerging trends, and shift tactics quickly. In political or electoral contexts, this allows hybrid actors to exploit crises or uncertainty and shape public debate before fact-checkers or institutions can respond (Nemitz, 2018). Third, AI improves targeting and personalization. By analyzing behavior and social networks, AI can identify specific groups and deliver messages that match their identities or grievances. This makes hybrid operations more effective and less visible, especially in polarized societies (Risse, 2020).

AI does not cause instability by itself; its impact depends on political choices, institutional strength, regulation, and overall societal resilience (Brundage et al., 2018). Strong institutions can limit its effects, while weak governance allows AI-enabled hybrid threats to deepen existing vulnerabilities.

4. The Western Balkans as a hybrid threat environment

A core vulnerability in the Western Balkans is the fragility of democratic institutions (Nix & Tamisiea, 2025). Although democratic frameworks exist, their effectiveness, independence, and public trust remain uneven and often contested (Freedom House, 2023). Electoral oversight, media regulation, judicial autonomy, and accountability mechanisms are regularly exposed to political pressure (European Commission, 2025), reducing institutions' capacity to detect or counter covert interference. In such conditions, hybrid threats aimed at weakening institutional legitimacy can have massive effects.

Political polarization further heightens these weaknesses. The region continues to face deep ideological, ethnic, and identity-based divisions rooted in unresolved conflicts and competing historical narratives (Bieber, 2018). These divides create fertile ground for manipulation, allowing hybrid actors to amplify existing grievances, undermine political opponents, and weaken trust in institutions. AI-driven tools intensify this process through highly targeted and emotionally charged messaging that is difficult to detect or attribute.

The region's uneven digital transformation adds another layer of risk. While governments, political actors, media, and citizens have rapidly adopted digital platforms, this shift has not been matched by investments in cybersecurity, regulatory oversight, or digital literacy (Buchanan, 2020; Nemitz, 2018). Digital infrastructures and information ecosystems remain exposed to manipulation and abuse. This gap between fast technological adoption and weak institutional capacity makes the environment particularly conducive to AI-enabled hybrid threats.

The media landscape represents an additional structural vulnerability. Although pluralism formally exists, many outlets operate under economic pressure, political influence, or opaque ownership (Freedom House, 2023). Disinformation spreads easily, especially via social media. AI-generated content, including synthetic text, audio, and imagery, heightens these challenges by increasing the volume and credibility of misleading narratives. For example, An AI- generated fake interview video featuring a Bosnian politician circulated in August 2025, using synthetic speech to make it appear as though he made statements he never actually said. This unlabeled AI- manipulated clip spread across Bosnia and Herzegovina and the wider Western Balkan information space, creating public confusion as fact- checkers later confirmed its artificial origin (IFEX, 2025).

External influence also shapes the region's hybrid-threat environment. The Western Balkans sit at the intersection of European, transatlantic, and global power dynamics. Prolonged uncertainty around Euro-Atlantic integration creates incentives for external actors to rely on indirect, deniable forms of influence (Kofman & Rojansky, 2015). Hybrid strategies are attractive because they allow impact without open confrontation. AI further lowers engagement costs, increases deniability, and enables sustained operations across multiple domains.

Despite the dominant narrative portraying the Western Balkans as a "passive target" in reality it is a space where domestic political actors, media networks, and social groups actively amplify and disseminate foreign disinformation, enabling external actors to exploit existing ethnic and political divisions (Strategic Analysis, 2023). Hybrid threats in the region are not solely externally driven: local media platforms, political parties, and interest groups often themselves participate in spreading manipulated content, blending economic interests, corruption, and partisan competition, thus making the region a producer, not only a consumer, of hybrid activities (EUISS, 2024).

Taken together, these factors make the Western Balkans a highly permissive environment for hybrid threats, with AI acting as an amplifier rather than a root cause. The next section examines why such threats are deeply political.

5. Why AI-driven hybrid threats are political

Debates on AI and security often focus on technical risks, such as system robustness, cybersecurity, or platform governance. While they are important, these concerns overlook the political nature of AI-enabled hybrid threats. Also, it is often human error that makes cyber-attacks possible. As shown throughout this paper, AI acts as a force multiplier for existing influence strategies because it reshapes relations of power, legitimacy, and accountability within democratic systems (Brundage et al., 2018; Nemitz, 2018).

First, AI-driven hybrid threats target political authority and democratic legitimacy. Disinformation erodes trust; election interference aims to delegitimize the process; and cyber operations often signal state weakness (Benkler et al., 2018; Mazarr, 2015; Risse, 2020; Buchanan, 2020). As noted before, AI increases scale, speed, and deniability, but the core effect is political disempowerment: weakening public belief in democratic institutions' ability to mediate conflict.

Second, AI-enabled hybrid tactics are context-dependent. Their impact varies with institutional strength, media structures, and social cleavages. As explained, vulnerabilities in the Western Balkans stem from contested governance, polarized media, and uneven regulation (Bieber, 2018; Freedom House, 2023). AI does not act autonomously; rather, it amplifies existing power asymmetries and governance gaps (Keil & Perry, 2015; Nemitz, 2018).

Third, treating AI-driven hybrid threats as political reframes, or allows for, attribution and accountability. Hybrid tactics thrive on deniability and legal ambiguity (Hoffman, 2007; NATO, 2014). AI heightens these challenges by enabling cross-border, semi-automated operations that evade clear attribution. The core issue becomes one of governance: existing rules on political communication, campaign finance, data protection, and platform responsibility lag behind technological realities (Nemitz, 2018).

Fourth, having a political lens helps us see the cumulative nature of AI-enabled hybrid threats. Their most damaging effects emerge gradually (Benkler et al., 2018; Mazarr, 2015). In the Western Balkans, targeted influence and disruption locks societies into circles of low trust, incentivizing domestic actors to use similar tactics (Bieber, 2018; Risse, 2020).

Finally, recognizing the political character of these threats has direct policy implications. Technical measures, such as cybersecurity upgrades, authentication tools, detection systems, are necessary but insufficient. Lasting mitigation requires institutional and societal strategies: for instance, transparent rules for digital campaigning, independent oversight bodies, media reforms that strengthen professionalism and ownership transparency, civic and media literacy, and regional coordination to prevent regulatory gaps (EU Commission, 2016; Freedom House, 2023). Effective responses depend as much on parliaments, regulators, courts, and newsrooms as on technical infrastructure. All this, however, is not enough without whole of society campaigns – for instance, on digital hygiene.

Viewed this way, the Western Balkans offer a relevant case: AI accelerates hybrid tactics not because technology is determinative, but because political structures are contested and institutions remain fragile. Understanding AI as a political amplifier shifts attention toward governance configurations that shape both vulnerability and resilience.

6. Evidence from the Western Balkans

Once amplified by AI, disinformation will only reinforce existing skepticism toward democratic institutions and media across the region. During elections, protests, or disputes, such narratives that question the credibility of electoral bodies, courts, or independent media circulate widely, without offering coherent alternatives. In North Macedonia's 2018 name-change referendum, for example, coordinated online campaigns amplified claims of foreign manipulation and institutional bias, contributing to low turnout and public distrust. The referendum's 37% turnout was influenced by boycott campaigns and by accusations of Russian attempts to influence the outcome (New Eastern Europe, 2018). The effect across the region is cumulative: declining epistemic trust, normalization of suspicion, and weakening confidence in democratic procedures (Benkler et al., 2018; Risse, 2020). Because it exploits pre-existing distrust, AI will be especially effective here.

Rather than manipulating vote counts, malicious actors seek to undermine the perception of fairness. The online information space surrounding the 2020 parliamentary elections in Montenegro featured coordinated disinformation efforts, including narratives alleging electoral fraud, external influence, and questioning the legitimacy of institutions, which contributed to a climate of distrust and heightened perceptions of procedural

uncertainty (ENEMO, 2020). The long-term impact is the delegitimization of elections as instruments of democratic accountability (Mazarr, 2015). AI tools could enhance such strategies through micro-targeting and rapid adaptation.

Cyber operations in the region follow the same political logic. High-profile incidents, such as the 2022 coordinated cyberattacks on Albania's government services, produced lasting political fallout. In retaliation for Albania's hosting of the Mujahedin- e- Khalq (MEK), an exiled Iranian opposition group, state-linked actors carried out a cyber-attack against country's government that destroyed data and disrupted essential public services, including the e- Albania portal, and leaked Albanian government data, such as emails from senior officials (Foreign, Commonwealth and Development Office, 2022). Automation has the potential to make such incidents more frequent.

In Serbia, a well- documented illustration of the interplay between political contention and information manipulation emerged during the mass anti- government protests that followed the November 2024 collapse of the Novi Sad railway station canopy, widely perceived as emblematic of country's systemic corruption and institutional failure. As student- led demonstrations expanded into the largest civic mobilization in Serbia in decades, government actors and pro- government media deployed a coordinated narrative strategy aimed at delegitimizing the protests, framing them as a "foreign- orchestrated color revolution" rather than a domestic movement for accountability (Đorđević, 2024). This messaging was amplified through synchronized primetime broadcasts, quasi- expert commentary, and extensive use of bot networks disseminating identical pro- government talking points across social media, reflecting a hybrid information environment in which state- aligned outlets and digital assets operate in tandem (Ibid, 2024). Independent monitoring further indicates that these narrative operations were accompanied by intensified coercive measures: the CIVICUS Monitor and BIRN documented increased arrests of protesters, intimidation of student leaders, and smear campaigns portraying civil society actors as "foreign mercenaries," often linked to alleged Western efforts to destabilize Serbia (Baletić, 2024). Parallel analyses by Freedom House and the Reuters Institute observe a marked deterioration in online freedoms during this period, including the targeting of activists with spyware, the dominance of government-aligned media ecosystems, and the strategic marginalization of independent broadcasters, even as students relied heavily on social media platforms to mobilize and circulate uncensored information (Freedom House, 2024; Milivojević, 2025). Serbia's protest cycles are affected by hybrid governance practices, wherein information manipulation, media capture, and coercive policing mutually reinforce one another to contain dissent and reframe it as externally driven subversion rather than legitimate democratic action (Stojanović, 2025).

Another illustrative example concerns the 2023 case involving Željko Mitrović, owner of TV Pink, who publicly promoted what he described as an "AI- based" system capable of generating satirical political content. In practice, however, the system was used to fabricate video and audio representations of opposition politicians, placing statements in their mouths that they had never made. The generated videos were broadcast in Pink's main news programs, provoking significant public and political criticism. President Aleksandar Vučić himself condemned the practice as "unfair" and inappropriate for a democratic society, explicitly stating that artificial intelligence should not be used to falsify individuals' speech or likeness. (Insajder, 2023; Nova.rs, 2023). This event demonstrates how AI- labeled technologies can be operationalized within hybrid media systems to distort political communication. It also illustrates the strategic ambiguity of "AI" as a discursive tool: while Mitrović presented the content as benign satire enabled by advanced technology, its effect was the production of politically consequential deepfakes.

Finally, second example concerns a 2024 incident in which an audio recording surfaced allegedly capturing Damir Zobenica, high- ranking SNS official and Vice- President of the Assembly of Vojvodina, giving detailed instructions to activists on how to provoke incidents during civic protests. When the recording was published by opposition figure Marinika Tepić, President Vučić responded by stating that Zobenica had informed him the audio was produced using artificial intelligence, implicitly suggesting that the recording was a deepfake. (N1, 2024). Independent audio- forensics experts quickly challenged the claim, arguing that the recording exhibited natural speech patterns inconsistent with current AI- generated Serbian- language voice synthesis. Audio engineer Dejan Tomka stated that producing such an authentic- sounding recording using existing AI tools would be "impossible," pointing to emotional cues, breathing, hesitation patterns, and tonal consistency that contemporary models cannot reliably reproduce. (N1, 2024b.) This case exemplifies a different but equally significant political use of AI: invoking the idea of artificial intelligence as a rhetorical shield to deflect allegations.

At the regional level, external influence interacts with internal vulnerability. Prolonged uncertainty around Euro-Atlantic integration (while Albania, Montenegro and North Macedonia are NATO members, Serbia, Bosnia and Herzegovina and Kosovo are not) and geopolitical competition create incentives for interventions that can

later be denied. Region- wide disinformation and influence campaigns have been documented across Bosnia and Herzegovina, Kosovo, North Macedonia, and Serbia; malign narratives circulate seamlessly across borders.

Serbia functions as a central hub for pro- Kremlin messaging, exporting narratives that delegitimize the EU and Western institutions into Montenegro, Bosnia and Herzegovina, and North Macedonia, where they are adapted to local grievances and ethnic divisions (Press Room, 2024). The same patterns emerge in the EUISS regional assessment, which highlights that disinformation about the EU, NATO, and Western actors travels with ease across Western Balkan borders, enabled by shared language ecosystems, high media interoperability, and cross-border political networks (EUISS, 2021).

Serbia's centrality within the regional information environment is structural rather than externally imposed, stemming from a combination of linguistic, media, political, and digital factors that enable narratives originating in Belgrade to circulate widely across neighboring states. The shared linguistic space of the Western Balkans allows Serbian- language media content to flow easily into Montenegro, Bosnia and Herzegovina, and parts of Kosovo and North Macedonia, giving Serbian broadcasters a built- in cross- border audience. This influence is reinforced by the dominance of pro- government outlets, such as Pink and Happy, across regional cable packages, which means that Serbian political narratives routinely spill into adjacent media markets and shape public discourse beyond Serbia's borders (Milivojević, 2025). Additionally, political networks aligned with Belgrade, including parties and officeholders in Republika Srpska, Montenegro, and segments of North Macedonia, help reproduce and legitimize Serbian state- aligned messaging across their own domestic spheres.

These activities are not driven predominantly by Russian (or other foreign) actors; rather, they rely on domestic political elites, aligned media conglomerates, and cross- border networks of local influencers, who deploy Russia- compatible narratives because they serve their own power consolidation strategies. In examining Serbia's contemporary media ecosystem on Facebook, it is notable that the principal pro- government pages do not exhibit operational or organizational ties to Russian state- aligned outlets; their messaging strategies remain largely domestically rooted and aligned with ruling- party communication priorities. By contrast, a parallel ecosystem of grey- zone platforms, such as IN4S and Srbin.info, maintains clear ideological affinities with Kremlin- aligned narratives and frequently amplifies content compatible with Russian geopolitical messaging. These outlets have historically served as vectors for disinformation campaigns and influence operations, and until at least 2022 were often sharply critical of President Vučić, portraying him as insufficiently aligned with Moscow's strategic interests. This dual structure, state- aligned Facebook pages on one side and Russian- connected grey- zone portals on the other, reflects what the ISAC Fund's Vulnerability Index – Serbia identifies as a fragmented information landscape particularly susceptible to malign foreign influence, especially in the domain of Kremlin- linked information operations (ISAC Fund, 2021).

This is the context in which AI enters the fray. In October 2024, the European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE) organized a major wargame in Vienna, Austria, focused specifically on hybrid threats aimed at the Western Balkans. Over four days (7–10 October), national teams, including delegations from Serbia, Montenegro, and North Macedonia, participated in a simulation built around a destructive earthquake scenario. Within the exercise, malign actors used "novel AI tools" to generate fabricated emergency announcements and AI- produced visuals portraying institutional failure, with the goal of destabilizing the regional information environment and eroding public trust in state authorities (Hybrid CoE, 2024).

In Albania, the 2024 disinformation landscape documented by SEE Check showed a marked expansion in the use of AI- generated manipulated visuals spreading across Facebook, Instagram, and anonymous online platforms. The Tirana- based fact- checking organization Faktoje verified more than 1,000 misleading claims during 2024, many of which relied on AI- fabricated or AI- enhanced imagery designed to mimic legitimate news reporting. These synthetic visuals contributed to a significant erosion of public confidence in Albanian media, institutions, and democratic processes (SEE Check, 2025).

These developments occurred alongside high- profile AI- driven political interference elsewhere in Europe, most notably during the 2023 Slovak parliamentary elections, where a deepfake audio recording targeted pro- European candidate Michal Šimečka. Scholars and analysts highlighted the Slovak incident as a test case for the type of AI- driven hybrid operations likely to be replicated in the Western Balkans, given similar low- trust political environments and susceptibility to pro- Russian disinformation (de Nadal & Jančárik, 2024).

Taken together, these cases illustrate how the Western Balkans has already entered a new phase of hybrid- threat exposure in which AI is not hypothetical, but operational. Technical defenses, cybersecurity upgrades, content moderation, detection tools are necessary but insufficient. Durable resilience depends on institutional reform, regulatory clarity, transparent media ownership, stronger professional standards, and societal capacity to resist

manipulation (EU Commission, 2016; Freedom House, 2023). Effective responses must therefore address the political foundations of democratic governance, not just technical vulnerabilities.

7. Policy responses

If AI-driven hybrid threats are treated only as technical problems, responses will remain reactive and fragmented. As shown earlier, AI amplifies hybrid threats by exploiting political, institutional, and societal vulnerabilities - conditions especially visible in the Western Balkans. Effective mitigation therefore requires a governance-centered approach that links cybersecurity with regulatory reform, institutional strengthening, and societal resilience.

Technical defenses remain essential, particularly for public institutions, electoral bodies, and critical infrastructure. Several Western Balkan governments have begun to modernize their cyber frameworks: North Macedonia's National Cybersecurity Strategy (2023) strengthens Computer Emergency Response Team (CERT) functions and mandates incident reporting, while Serbia's 2022 Law on Information Security introduces stricter risk-assessment requirements for operators of essential services (Government of North Macedonia, 2023; CERT Serbia, 2022). AI can improve anomaly detection and automated threat monitoring, but must operate within systems that ensure transparency and accountability (Buchanan, 2020).

Regulatory upgrades are equally necessary. Legal frameworks governing political advertising, data protection, and platform accountability still lag behind the realities of AI-enabled manipulation. Some progress exists - for example, Montenegro's 2023 media-law amendments introduced stronger transparency requirements for online political advertising, while Bosnia and Herzegovina has moved toward GDPR (General Data Protection Regulation) – inspired data-protection standards (Montenegro Ministry of Culture & Media, 2023; BiH Personal Data Protection Agency, 2022).

Yet regulatory reforms must avoid over-securitization. Treating too many social, political, and informational challenges primarily as security threats produces distortions that ultimately weaken, rather than strengthen, democratic resilience. Overly restrictive rules risk being used to control speech or target political opponents, already visible in the region through selective takedown practices and ambiguous "fake news" provisions. Safeguarding fundamental rights and ensuring independent oversight remain central (Nemitz, 2018).

Institutional resilience is just as important as technical security. Independent electoral commissions, media regulators, data-protection authorities, and courts provide the backbone of democratic legitimacy under persistent hybrid pressure. Practical examples include the Central Election Commission of Kosovo, which monitors online political advertising during elections, and North Macedonia's Agency for Audio and Audiovisual Media Services (AAVMS), which conducts transparency audits of broadcasters and online portals (CEC Kosovo, 2021; AAVMS, 2022).

Media ecosystems remain a significant vulnerability. Greater transparency in media ownership, sustainable funding for independent journalism, and support for fact-checking organizations can reduce susceptibility to AI-enhanced manipulation. Initiatives such as the Balkan Fact-Checking Network (BFCN), RCC-supported regional disinformation-monitoring programs, and EU IPA III assistance for investigative journalism illustrate practical steps already underway (Freedom House, 2023; RCC, 2022).

Societal resilience depends heavily on civic and media literacy. Public-education programs, such as Serbia's "Check Before You Share," UNDP Montenegro's media-literacy pilots, or EU-supported digital-literacy schools in North Macedonia, improve citizens' ability to identify manipulation and understand algorithmic dynamics (UNDP Montenegro, 2023). In contexts marked by polarization and disengagement, such initiatives strengthen civic agency and reduce vulnerability to hybrid interference.

8. Regional and European coordination

Governments in the Western Balkans face limits in addressing these challenges through purely national measures. Recent regional initiatives show how information sharing, joint training, and coordinated response can strengthen collective resilience. For example, the Western Balkans Cyber Capacity Centre (WB3C), launched in 2023 by France, Slovenia, and regional partners, provides shared cyber incident reporting channels, analytic workshops, and AI-related training modules for national CERT teams (WB3C, 2023). The Regional Cooperation Council (RCC) has also facilitated cooperation against disinformation through regional exchanges among fact-checking organizations and joint monitoring of cross-border influence campaigns (RCC, 2022). Regular bilateral cybersecurity exercises, such as the Serbia–North Macedonia joint cyber drills held since 2021, simulate

coordinated attacks on government platforms and critical infrastructure, improving detection and response capacity (CERT Serbia, 2022). These mechanisms help mitigate asymmetries in capacity and reduce opportunities for hybrid actors to exploit weaker states or institutions.

At the European level, deeper integration of the Western Balkans into EU security, digital, and regulatory frameworks should provide a critical long-term resilience strategy. The EU Hybrid Fusion Cell enhances situational awareness by collecting, analyzing, and sharing information on hybrid threats among EU members and Western Balkan partners (EU INTCEN, 2016). The region has increasingly participated in training programs organized by the European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE), which supports table-top exercises and scenario planning on hybrid and AI-enabled interference (Hybrid CoE, 2021). The EU–Western Balkans Cybersecurity Blueprint Exercise mirrors the EU's Blue OLEx process, helping partner states practice coordinated regional responses to large-scale cyber incidents (ENISA, 2023).

Regulatory alignment is also advancing. Through the accession process, Western Balkan governments have begun approximating elements of the Digital Services Act (DSA), Digital Markets Act (DMA), and the emerging AI Act, including rules on platform accountability, transparency of political advertising, and algorithmic governance (EU Commission, 2021; 2023). Importantly, EU support extends beyond technical standards: initiatives under IPA III, the EU Rule of Law missions, and media-sector assistance programs provide institutional reform, independent regulatory capacity, and support for professional journalism (EU Parliament, 2022). These combined measures help anchor Western Balkan states within rule-based European norms and strengthen their ability to counter hybrid and AI-enabled threats.

9. Conclusion

This paper has examined how artificial intelligence amplifies hybrid threats in the Western Balkans, arguing that AI-driven security challenges are fundamentally political rather than merely technical. By situating AI within established hybrid-threat frameworks, the analysis shows that AI acts as a force multiplier - expanding the scale, speed, adaptability, and deniability of existing tactics of influence and disruption. Its impact depends not only on technological capabilities, but on the quality of democratic governance, institutional resilience, and societal trust.

The conceptual framework clarifies that AI is not an autonomous source of instability. As shown across disinformation, election interference, and cyber operations, AI does not create new forms of hybrid activity; it intensifies pre-existing political dynamics, especially in contexts marked by polarization, weak oversight, and contested legitimacy. The Western Balkans illustrate how such conditions generate permissive hybrid-threat environments, where AI-enabled tools produce cumulative political effects even when individual incidents appear limited.

Empirically and analytically, the paper contributes to international relations and security studies in three ways. First, it connects hybrid-threat scholarship with emerging work on AI and political security, emphasizing interaction effects rather than technological determinism. Second, it provides a region-specific perspective that treats the Western Balkans not as exceptional, but as a setting where broader European and global dynamics appear in concentrated form. Third, it reinforces a process-oriented understanding of hybrid threats, showing how repeated and adaptive interference gradually erodes democratic legitimacy.

Policy implications follow directly. Technical responses, such as cybersecurity upgrades, detection tools, platform governance are necessary but insufficient. Durable resilience requires strategies that strengthen democratic institutions, clarify regulatory authority over digital political spaces, support more transparent and credible media ecosystems, and invest in societal capacity to recognize and counter manipulation. For the Western Balkans, closer alignment with EU regulatory frameworks and deeper regional cooperation are especially important for reducing asymmetries that hybrid actors exploit.

More broadly, the findings suggest that AI-driven hybrid threats challenge not only security policy, but democratic governance in the digital age. As AI capabilities continue to expand, the gap between technological change and democratic oversight risks widening. Addressing this gap is ultimately political: it depends on reinforcing public accountability, institutional legitimacy, and democratic control over the infrastructures that shape political communication.

References

1. AAVMS. (2022). *Annual transparency report on media services*. Agency for Audio and Audiovisual Media Services.
2. Arkan, Z. (2025). Conclusion: Hybrid threats, shared stories - Narratives of security in NATO and the EU. In *European security and hybrid threats* (pp. 67–69). Springer.
3. Baletić, K. (2024). *Protesters' arrests fuel human rights concerns in Serbia, report says*. Balkan Insight.
4. Benkler, Y. (2023). Disinformation, social media, and the democratic crisis. *Journal of Democracy, 34*(1), 50–64.
5. Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
6. Bieber, F. (2018). Patterns of competitive authoritarianism in the Western Balkans. *East European Politics, 34*(3), 337–354.
7. BiH Personal Data Protection Agency. (2022). *Guidelines on GDPR alignment*.
8. Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy, 3*, e32.
9. Browne, J., Coffey, B., Cook, K., Meiklejohn, S., & Palermo, C. (2019). A guide to policy analysis as a research method. *Health Promotion International, 34*(5), 1032–1044.
10. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., … Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Future of Humanity Institute, University of Oxford.
11. Buchanan, B. (2020). *The hacker and the state: Cyber-attacks and the new normal of geopolitics*. Harvard University Press.
12. CEC Kosovo. (2021). *Election integrity and online campaign monitoring report*. Central Election Commission of Kosovo.
13. CERT Serbia. (2022). *Annual report on cybersecurity exercises in regional cooperation*. Ministry of Information and Telecommunications.
14. CERT Serbia. (2022). *Information security annual report*. Ministry of Information and Telecommunications.
15. de Nadal, L., & Jančárik, P. (2024). Beyond the deepfake hype: AI, democracy, and "the Slovak case". HKS Misinformation Review.
16. Đorđević, T. (2024). *When the regulators are raised: How does the propaganda machine "extinguish" the crisis in Serbia?* Istinomer.
17. ENEMO. (2020). *International Election Observation Mission: Montenegro parliamentary elections, 30 August 2020 – Final report*. European Network of Election Monitoring Organizations.
18. ENISA. (2023). *EU–Western Balkans cybersecurity blueprint exercise report*. European Union Agency for Cybersecurity.
19. European Commission. (2016). *Joint framework on countering hybrid threats: A European Union response* (JOIN (2016) 18 final).
20. European Commission. (2021). *Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act)* (COM (2020) 825 final).
21. European Commission. (2022). *Strengthened Code of Practice on Disinformation*.
22. European Commission. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union*.
23. European Commission. (2024). *Regulation (EU) 2024/900 on the transparency and targeting of political advertising. Official Journal of the European Union*.
24. European Commission. (2025). *Rule of Law Report 2025: Candidate countries - Albania, Montenegro, North Macedonia, and Serbia*.
25. European Parliament. (2022). *IPA III programming document – Governance and rule of law*.
26. EU Institute for Security Studies. (2021). *The Western Balkans and EU–NATO cooperation: How to counter foreign interference and disinformation?*
27. European Union Institute for Security Studies. (2024). *Countering cyber-enabled hybrid interference in the Western Balkans: A scenario-based approach*.
28. EU Intelligence and Situation Centre. (2016). *EU Hybrid Fusion Cell: Mandate and functions*.
29. Foreign, Commonwealth & Development Office. (2022, September 7). *UK condemns Iran for reckless cyber-attack against Albania*. UK Government.
30. Freedom House. (2023). *Nations in Transit 2023: The repressive turn in the Western Balkans*.

31. Government of North Macedonia. (2023). *National Cybersecurity Strategy 2023–2028*.
32. Hoffman, F. G. (2007). *Conflict in the 21st century: The rise of hybrid wars*. Potomac Institute for Policy Studies.
33. Hybrid CoE. (2021). *Training and exercises catalogue*. European Centre of Excellence for Countering Hybrid Threats.
34. Hybrid CoE. (2024). Western Balkans in focus at countering disinformation wargame and conference in Vienna. Hybrid Centre of Excellence for Countering Hybrid Threats.
35. Hynes, P., Bew, R., Williamson, J. R., & Kenyon, M. (2025). *AI as a cybersecurity risk and force multiplier*. National Association of Corporate Directors.
36. IFEX. (2025, August 29). *Detecting fake content online in the Balkans*. Mediacentar Sarajevo.
37. Insajder. (2023). Vučić kaže da veštačka inteligencija ne sme da se koristi na način na koji to čini vlasnik TV Pink. https://www.insajder.net/prenosimo/vucic-kaze-da-vestacka-inteligencija-ne-sme-da-se-koristi-na-nacin-na-koji-to-cini-vlasnik-tv-pink
38. Keil, S., & Perry, V. (Eds.). (2015). *State-building and democratization in the Western Balkans*. Routledge.
39. Kofman, M., & Rojansky, M. (2015). *A closer look at Russia's "hybrid war"*. Kennan Institute, Wilson Center.
40. Mazarr, M. J. (2015). *Mastering the gray zone: Understanding a changing era of conflict*. U.S. Army War College Press.
41. Milivojević, S. (2025). Serbia. In *Digital News Report 2025*. Reuters Institute for the Study of Journalism.
42. Montenegro Ministry of Culture & Media. (2023). *Draft law on media – Amendments*.
43. N1. (2024). Vučić o navodnom snimku Zobenice: Poslao mi je poruku da je to veštačka inteligencija.
44. N1. (2024b). Veštačka inteligencija ili Zobenica "stasom i glasom": Stručnjak tvrdi – nemoguće da AI napravi tako autentičan snimak.
45. NATO. (2014). *Wales Summit Declaration*.
46. NATO. (2024). *Hybrid threats and hybrid warfare: Reference curriculum*. NATO Headquarters.
47. Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A, 376*(2133), 20180089.
48. New Eastern Europe. (2018, October 1). *The Macedonian name change referendum*. New Eastern Europe.
49. Nix, S., & Tamisiea, M. (2025). *Democracy at a crossroads: Rule of law and the case for US engagement in the Balkans*. Atlantic Council.
50. Nova.rs. (2023). "To je nefer, to ne sme da se radi…": Vučić o „novom projektu "Željka Mitrovića.
51. Press Room (DISA). (2024, December 18). *Influence and disinformation campaign monitoring in the Western Balkans (MEDIWEB report)*.
52. Regional Cooperation Council. (2022). *Countering disinformation in the Western Balkans – Regional assessment*.
53. Regional Cooperation Council. (2022). *Countering disinformation in the Western Balkans*.
54. Risse, T. (2020). *Governance without a state? Policies and politics in areas of limited statehood*. Columbia University Press.
55. Romansky, S., Hoenig, A., Meessen, R., & Kruijver, K. (2024). *New technologies, changing strategies: Five trends in the hybrid threat landscape*. The Hague Centre for Strategic Studies.
56. Ryan-Mosley, T. (2023). How generative AI is boosting the spread of disinformation and propaganda. *MIT Technology Review*.
57. SEE Check. (2025). Disinformation report: Albania in 2024. SEE Check / Faktoje. https://seecheck.org/index.php/2025/06/05/disinformation-report-albania-in-2024/
58. Strategic Analysis. (2023). *Hybrid threats in the Western Balkans: State and non-state perspective*. Strategic Analysis Think Tank.
59. UNDP Montenegro. (2023). *Media and digital literacy programme outcomes*. United Nations Development Programme.
60. Western Balkans Cyber Capacity Centre. (2023). *Programme overview*. Government of France & Government of Slovenia.