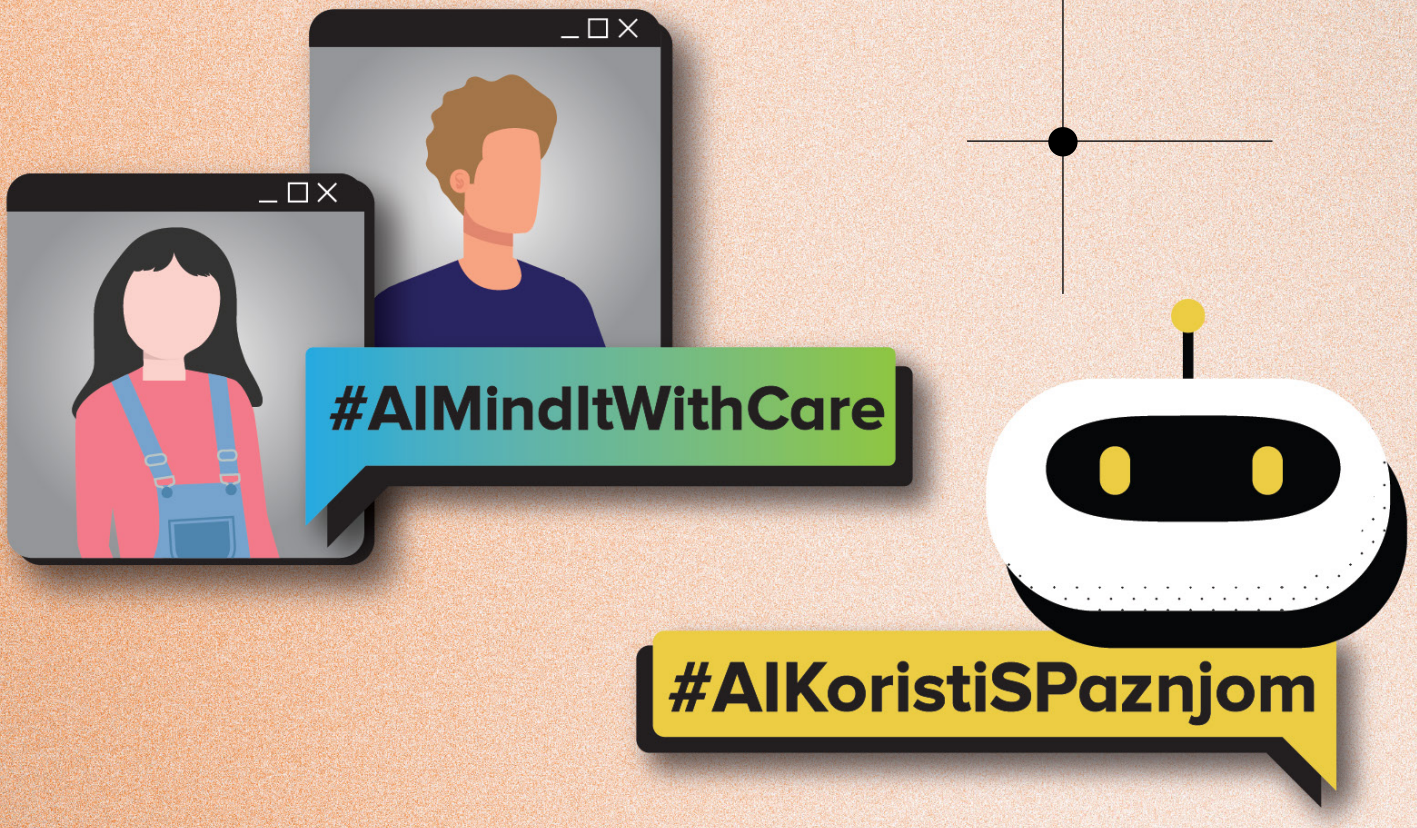
Jabiheu NL Netherlands

# ALGORITHMIC CURRENTS:
# Fortifying Youth and Democracy in the Western Balkans Against AI-Driven Harms




2025

# Algorithmic Currents: Fortifying Youth and Democracy in the Western Balkans Against AI-Driven Harms

**Author**:

Borislav Vukojevic,
*assistant professor of Faculty of Political Sciences University of Banja Luka*[1]

1 borislav.vukojevic@fpn.unibl.org

The Western Balkans, a region characterized by post-conflict fragility and persistent challenges to democratic consolidation, faces a significant and deceptive threat from AI-driven hate speech, extremism and disinformation. This report, “Algorithmic Currents,” provides a rigorous analysis of how artificial intelligence exacerbates pre-existing ethnic divisions and historical trauma, with profound implications for the region’s youth. The central inquiry identifies optimal improvements for legal, institutional and societal responses, emphasizing the critical need to safeguard and empower this vulnerable demographic. Without immediate, coordinated intervention, specifically leveraging the EU integration process as an indispensable strategic imperative, these algorithmic currents risk irrevocably eroding social cohesion and democratic trust.

Key findings reveal that AI’s content recommendation systems create a “Balkan Algorithmic Fracture,” amplifying divisive narratives. A critical “Linguistic Seam-Rip” exists, where AI moderation frequently fails to detect nuanced hate speech in Bosnian, Serbian and Croatian languages due to prevalent code-switching and intricate cultural contexts. This oversight necessitates a “Localized AI Moderation and Language Data Initiative” and the strategic adoption of “The Acquis Digitalis” to compel platform accountability. The report culminates in the “Digital Immune System,” a transformative framework that redefines youth as proactive digital sentinels. This system is implemented through “Youth-Led Fact-Checking Networks,” “Algorithmic Awareness and Critical Thinking Curriculum,” a “Regional Digital Diplomacy and Counter-Narrative Hub,” and “Trauma-Informed Digital Storytelling for Reconciliation.” These initiatives collectively build societal resilience - viable path toward European integration and a more secure digital future for the region.

# CONTENTS


**Algorithmic Currents: Fortifying Youth and Democracy in the Western Balkans Against AI-Driven Harms ..... 2**

**1. Introduction: Navigating Algorithmic Currents ..... 1**

1.1 The Digital Crucible: AI and the Western Balkans ..... 1
1.2 BiH's Amplified Fragility: Post-Conflict Divisions in the Digital Age ..... 1

**2. Methodology: Charting the Digital Spaces ..... 2**

2.1 Research Design and Data Prioritization ..... 2
2.2 Research Limitations and Linguistic Complexities ..... 3

**3. Regulatory Frameworks: The Patchwork of Digital Governance ..... 4**

3.1 BiH's Fragmented Legal Landscape and Regional Disparities ..... 4
3.2 EU Approximation: The 'Ineluctable' Push for Digital Standards ..... 5

**4. Algorithmic Bias and Content Dissemination: The Balkan Algorithmic Fracture ..... 6**

4.1 The Mechanics of Algorithmic Amplification ..... 6
4.2 Algorithmic Bias: Exploiting Post-Conflict Trauma ..... 7
4.3 BiH-Specific Risks and Societal Erosion ..... 8

**5. AI-Generated Disinformation and Manipulation: The Ephemeral Nature of Truth ..... 9**

5.1 The Rise of Synthetic Media and Automated Campaigns ..... 9
5.2 Vulnerabilities: Historical Narratives, Elections, and Propaganda ..... 10
5.3 Societal Erosion: Youth Trust and Cross-Border Campaigns ..... 11

**6. Practices and Responses of Tech Companies and Regulatory Bodies: The Linguistic Seam-Rip ..... 12**

6.1 Tech Company Responsiveness and Regional Disregard ..... 12
6.2 National Regulatory Capacity: Navigating the Digital Wild West ..... 13

**7. Conclusion ..... 15**

**8. World Building: The Larger Universe of Ideas ..... 16**

**Recommendations and Strategic Roadmap Items ..... 17**

**Recommendations for Policies and Accountability Mechanisms ..... 18**

8.1 Regional Cooperation Strategies for the Western Balkans ..... 18
8.2 Youth Inclusion and Skills Development Programs ..... 19
8.3 Multi-Stakeholder Engagement and Capacity Building ..... 21

**Strategic Roadmap: Phased Implementation and Resilience Building ..... 22**

9.1 Strategic Roadmap: Government (State and Entity Levels) Steps for BiH ..... 22
9.2 Strategic Roadmap: Civil Society and Media Steps for BiH ..... 23
9.3 Strategic Roadmap: Regional Coordination Mechanisms for Western Balkans ..... 25
9.4 Strengthening Oversight and Accountability in AI System Deployment ..... 26
9.5 Long-Term Measures for Societal Resilience (Youth Focus) (41-45) ..... 27
9.6 Sustainable Funding and International Support Mechanisms ..... 28
9.7 Synthesizing Findings and Prioritizing Action ..... 28
9.8 Youth, Cooperation, and the 'Ineluctable' Path Forward ..... 30

**Appendix A: Glossary of Terms ..... 31**

A.1 Definition of Key Terms ..... 31

**References ..... 33**

# Introduction:
# Navigating Algorithmic Currents

## 1. The Digital Crucible: AI and the Western Balkans

Digital age, characterized by the omnipresent rise of Artificial Intelligence, fundamentally reshapes global information environments. This technological evolution offers advancements, but also possesses a potent capacity to intensify societal vulnerabilities, particularly in regions navigating protracted post-conflict circumstances. The Western Balkans, a complex area marked by historical divisions and evolving democratic frameworks, serves as a crucial case study. Here, AI's transformative power intersects with pre-existing fragilities, creating a "digital crucible." The pervasive influence of AI, operating through sophisticated amplification mechanisms, presents a significant threat to democratic processes and social cohesion by fueling hate speech, extremism and disinformation.

Bosnia and Herzegovina (BiH) exemplifies this extensive digital exposure, with a high internet penetration of 96% and social media usage reaching 78%. Facebook alone accounts for 71% of internet users.[2] This widespread online engagement, combined with the region's intricate socio-historical context, provides fertile ground for AI to amplify divisive narratives, whether intentionally or not. This report investigates how AI, through its complex and often opaque algorithms, intensifies these harms. The central question guiding this report is: How can legal, institutional, and societal responses in Bosnia and Herzegovina and the Western Balkans be optimally improved to mitigate the significant impact of AI-driven hate speech, extremism, and disinformation, with a deliberate emphasis on safeguarding and empowering youth? This question highlights the imperative for immediate and coordinated action within this distinct geopolitical and socio-historical landscape.

## 2. BiH's Amplified Fragility: Post-Conflict Divisions in the Digital Age

Bosnia and Herzegovina is a country shaped by post-conflict trauma and deep ethnic divisions. This complex constitutional structure, a direct outcome of unresolved historical grievances, creates fertile ground for AI-driven harms. Algorithms, often optimized for user engagement, inadvertently amplify divisive rhetoric along ethnic lines, transforming digital platforms into arenas for "Digital Memory Wars."

The region's vulnerability to cyber-enabled information warfare and disinformation campaigns stems from the exploitation of these pre-existing societal fissures.[3] This dynamic is particularly evident across BiH's administrative divisions: Republika Srpska, the Federation of BiH, and Brčko District. In these areas, youth are specifically targeted, exposed to narratives that reinforce ethnic isolation and normalize extremist discourse. The "Balkan Algorithmic Fracture" represents the daily reality of young people in Sarajevo navigating "The Digital Agora of Sarajevo," those in Banja Luka experiencing "Banja Luka's Digital Echo Chamber," and individuals in Brčko District confronting "Brčko District's Digital Crossroads." These digital spaces, while offering connectivity, simultaneously deepen existing cleavages.

2 Bojana Kostić and Caroline Sinders, "Responsible Artificial Intelligence: An overview of human rights' challenges of Artificial Intelligence and media literacy perspectives in the context of Bosnia and Herzegovina" (Council of Europe, June 2022), accessed viahttps://rm.coe.int/mil-study-3-artificial-intelligence-final-2759-3738-4198-2/1680a7cdd9, Chapter I, Introduction

3 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.2 "Overview of the morning session discussion – addressing the first set of questions," p. 27.

This issue extends beyond BiH's borders. The Western Balkans operates as an interconnected information environment. Disinformation campaigns originating in one country rapidly spread across Serbia, Montenegro, North Macedonia, Albania and Kosovo*[4], leveraging shared linguistic roots and intertwined historical narratives. This regional interconnectedness elevates national security concerns to a regional security imperative. Effective responses require coordinated, cross-border approaches. Isolated national efforts are insufficient against the widespread flow of AI-amplified harmful content. Public discourse, social cohesion, and nascent democratic processes across the region face a significant threat if these algorithmic currents remain unchecked.

# Methodology: Charting the Digital Spaces

## 2.1 Research Design and Data Prioritization

Research design for "Algorithmic Currents" utilizes a desk research approach, a methodology necessitated by the limited availability of primary data and direct algorithmic transparency within the Western Balkans. This approach synthesizes existing scholarly literature and official reports from international organizations, including the United Nations, the European Union, and the Organization for Security and Co-operation in Europe. It also incorporates analyses from reputable regional and international think tanks and civil society organizations. Selection criteria prioritize documents offering BiH-specific data and broader regional Western Balkans studies, ensuring contextual nuances of post-conflict societies, ethnic sensitivities, and ongoing democratic transitions inform the analysis.

Methodological framework relies on an iterative process of evidence synthesis. This involves systematically identifying, collating, and critically evaluating research that clarifies the widespread yet often obscured impact of AI on information environments. Sources detailing internet penetration rates, social media usage patterns among youth, and documented instances of online hate speech or disinformation campaigns within BiH and across the Western Balkans receive precedence. This framework constructs a comprehensive picture of AI's influence, even when direct empirical data on AI's specific application in generating or amplifying harmful content remains difficult to obtain. The approach recognizes that while direct algorithmic transparency from major tech platforms is frequently elusive, a robust understanding can derive from observable effects and documented trends in public discourse.

Data prioritization extends to studies addressing the interaction between technology and the region's distinct socio-historical context. This includes analyses of how AI algorithms might inadvertently exploit historical grievances or ethnic divisions, contributing to what this report terms the "Balkan Algorithmic Fracture." The essential need for localized data guides this principle, ensuring that subsequent analysis rests not on generic assumptions, but on verifiable, albeit often fragmented, evidence specific to the Western Balkans. This careful selection process ensures that every assertion within this report is substantiated by a rigorous review of available documentation, forming a credible foundation for the policy recommendations that follow.

4 According to Resolution 1244

## 2.2 Research Limitations and Linguistic Complexities

The desk research approach, while crucial for navigating the limited availability of primary data in the Western Balkans, inherently presents limitations. A significant challenge stems from the pervasive lack of algorithm transparency from major tech companies.[5] This opacity prevents a granular understanding of how AI tools operate within the region, particularly concerning content recommendation engines and moderation systems. Publicly available data on AI tools' precise applications or their impact on local information environments remains scarce. The limited market size of the Western Balkans offers minimal incentive for large platforms to invest substantially in localized data collection or transparency initiatives.[6] This creates a "Moderation Desert," where comprehensive oversight is difficult, and the transient nature of online harms often outpaces detection.

Localized research on AI's impact in the Western Balkans is also scarce. While global studies address algorithmic bias and disinformation, specific analyses detailing their manifestation within the unique socio-historical context of BiH and its neighbors are rare. This requires careful interpretation of broader findings, adapting them to the region's specific ethnic, religious, and political sensitivities. Absence of robust, localized datasets for AI training and evaluation further exacerbates this limitation, impeding the development of context-aware solutions.

Incorporating language considerations for content in Bosnian, Serbian, and Croatian (B/S/C) languages presents a substantial hurdle. These languages are often categorized as "low-resource languages" in AI development, meaning a scarcity of labeled data hinders the effectiveness of automated moderation systems.[7] This linguistic complexity manifests in several ways:

- **Morphological Complexity and Dialectal Diversity:** B/S/C languages possess rich grammar, regional dialects, and variations in script (Serbian uses both Cyrillic and Latin).[8] These nuances complicate Natural Language Processing (NLP) pipelines, making it difficult for AI to accurately interpret meaning and intent.[9]

- **Code-Switching and Informal Language Use:** Online comments frequently feature code-switching between Latin and Cyrillic scripts, along with extensive use of slang, sarcasm, and idiomatic expressions.[10] This informal linguistic fluidity makes automated detection of harmful content challenging, as a seemingly innocuous phrase can carry a derogatory or divisive meaning depending on its context or script.

- **Cultural Nuances and Contextual Ambiguity:** Toxicity in the Western Balkans is often context-dependent, deeply influenced by historical tensions, political polarization, and specific cultural sensitivities.[11] AI models struggle to interpret these subtleties without extensive, localized training data and human expertise. For instance, a comment referring to a "sick person" might be missed by zero-shot AI models but correctly flagged as toxic when provided the context of it referring to a struggling Serbian politician.[12] Conversely, harmless remarks can be over-flagged if context is misunderstood by context-augmented models.[13]

These linguistic nuances contribute directly to the "Linguistic Seam-Rip," a critical enforcement gap where AI content moderation and even human review, lacking specialized training, fail to detect subtle, culturally embedded hate speech. The 2021 investigation by the Balkan Investigative Reporting Network (BIRN) found that "nearly half of the toxic language posts reported in Balkan languages remained online, even

5 World Economic Forum. (2025). *The Global Risks Report 2025*. Insight Report, 20th Edition. World Economic Forum, Geneva, Switzerland. Page 36, Section 1.5. https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf
6 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective," European Western Balkans, April 4, 2025. https://europeanwesternbalkans.com/2025/04/04/why-the-digital-services-act-is-needed-in-the-western-balkans-an-institutional-and-market-perspective/
7 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," *arXiv preprint arXiv:2506.09992*, 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
8 Ibid., Section II-C, Paragraph 2.
9 Ibid., Section II-C, Paragraph 2.
10 Ibid., Section I-A, Paragraph 6.
11 Ibid., Section I-A, Paragraph 6.
12 Ibid., Section V-A Key Findings.
13 Ibid., Section V-B Practical Implications.

after Facebook and Twitter confirmed that the content violated their rules."[14] This systemic failure leaves youth particularly vulnerable to AI-amplified harms that appear innocuous to generic filters. Consequently, the interpretation of findings within this report demands cautious generalization, acknowledging these inherent limitations while striving for the highest possible degree of precision given the available evidence. This methodological transparency establishes realistic expectations for the report's scope and emphasizes the undeniable need for targeted, localized research and development.

# Regulatory Frameworks: The Patchwork of Digital Governance

## 3.1 BiH's Fragmented Legal Landscape and Regional Disparities

Legal and policy frameworks in Bosnia and Herzegovina (BiH) are fragmented, mirroring the nation's complex constitutional structure. This intricate web of state-level and entity-level legislation often proves inadequate in addressing the nuanced challenges of digital governance, particularly concerning AI-driven hate speech and disinformation. BiH lacks specific AI regulatory frameworks, leaving issues such as hate speech, polarization, and inter-ethnic animosity on social media largely unaddressed by targeted laws.[15] While the 2021–2027 Development Strategy, adopted by the Parliament of the Federation of Bosnia and Herzegovina, acknowledges AI's strategic importance, it offers no concrete steps toward establishing legal frameworks for AI development and deployment.[16]. This regulatory void creates a "Digital Divide of Governance," leaving mechanisms for regulating AI's societal impact underdeveloped.

A regional comparison of regulatory approaches across Western Balkan countries reveals an underdeveloped landscape for cybersecurity legislation. Enforcement is inconsistent due to limited governmental expertise and politicization.[17] This fragmentation of digital governance, which parallels BiH's broader political architecture, significantly impedes a unified and effective response to AI-driven harms. The Communications Regulatory Agency (CRA) in BiH acknowledges the absence of national AI strategies and has proposed to act as a "policy shaper" by initiating national dialogues and suggesting roadmaps.[18].

This fragmented landscape creates vulnerabilities to AI-driven harms, particularly in a region marked by historical divisions. Large technology companies leverage their digital power to unleash new forms of orchestrated harms and public distortions, yet they demonstrate minimal interest in addressing widespread hate speech targeting individuals, ethnic groups, and critical thinkers in Bosnia and Herzegovina.[19] Eco-

14 Ibid., Section I-A, Paragraph 4.
15 Bojana Kostić and Caroline Sinders, "Responsible Artificial Intelligence: An overview of human rights' challenges of Artificial Intelligence and media literacy perspectives in the context of Bosnia and Herzegovina" (Council of Europe, June 2022), accessed via https://rm.coe.int/mil-study-3-artificial-intelligence-final-2759-3738-4198-2/1680a7cdd9, Chapter I, Introduction.
16 Namanja Sladaković and Milica Novaković, "Bosnia and Herzegovina," in *IBA Alternative and New Law Business Structures Committee*, July 2024, accessed https://www.ibanet.org/medias/anlbs-ai-working-group-report-july-2024-4-bosnia-herzegovina.pdf?context=bWFzdGVyfFB1YmxpY2F0aW9uUmVwb3J0c3w1NTQ3OHxhcHBsaWNhdGlvbi9wZGZ8YURJeUwyaGlaQzg1TVRNNE1qQTJOREE0TnpNMEwyRnViR0p6TFdGcExYZHZjbXRwYm1jdFozSnZkWEF0Y21Wd2IzSjBMV3AxYkhrdE1qQXlOQzAwTFdKdmMyNXBZUzFvWlhKNlpXZHZkbWx1WVM1d1pHWXw4M2UzMmYzYTRlOGEwZWJiYTk1ZTkyYThkN2MxNGI1NmU4ZDFhODUxOGY5ODQ0OTM1NWIzYTc1MmFlODkwZGZl
17 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.2, p. 25.
18 Communications Regulatory Agency (CRA) Bosnia and Herzegovina, "GSR-25 Contribution: Best Practice Guidelines," presented at the GSR-25 Consultation, accessedhttps://www.itu.int/itu-d/meetings/gsr-25/wp-content/uploads/sites/33/2025/06/GSR-25_Contribution_Best-Practice-Guidelines_RAK-Bosnia-and-Herzegovina.pdf
19 Bojana Kostić and Caroline Sinders, "Responsible Artificial Intelligence: An overview of human rights' challenges of Artificial Intelligence and media literacy perspectives in the context of Bosnia and Herzegovina" (Council of Europe, June 2022), accessed via https://rm.coe.int/mil-study-3-artificial-intelligence-final-2759-3738-4198-2/1680a7cdd9, Chapter I, Introduction.

nomic challenges within the region further impact digital infrastructure and the efficacy of literacy programs, exacerbating these issues. The lack of specific legal provisions for AI and machine learning technologies raises concerns regarding ethical implications associated with the widespread adoption of AI systems across various sectors.[20].

The implications of this regulatory stagnation are clear. Without robust legal frameworks, the "Balkan Algorithmic Fracture" persists as AI systems exploit existing ethnic and political fault lines without effective oversight. This leaves youth, the most digitally immersed demographic, particularly susceptible to extremist narratives propagated through "The Digital Iron Curtain."[21] Furthermore, the absence of comprehensive legislation hinders efforts to address "The Linguistic Seam-Rip," where AI moderation struggles to detect nuanced hate speech in Bosnian, Serbian, and Croatian languages. This is not merely a technical oversight but a systemic failure rooted in the "Moderation Desert," a consequence of insufficient investment from technology companies in lower-revenue regions.[22]. Consequently, the legal and institutional frameworks across BiH and the wider Western Balkans are currently ill-equipped to counter the evolving threats posed by AI-driven harmful content, necessitating a fundamental shift in approach.

## 3.2 EU Approximation: The 'Ineluctable' Push for Digital Standards

European Union (EU) approximation process drives digital standards across the Western Balkans. This integration aligns national regulatory frameworks with the EU's extensive acquis, including the Digital Services Act (DSA) and the forthcoming AI Act. This alignment represents a critical opportunity. It allows the region to overcome domestic political fragmentation and address the challenges of AI-driven harmful content.

An assessment of current Western Balkan regulatory frameworks reveals substantial legislative gaps and enforcement inconsistencies when compared to the DSA. Albania, for example, has only 47% of its DSA-related framework covered by existing legislation.[23] Serbia and Bosnia and Herzegovina exhibit similar deficiencies. Kosovo* and North Macedonia, despite greater convergence, lack the full protections offered by the DSA.[24] This regulatory void leaves citizens vulnerable to illegal and harmful online content. Platforms, perceiving these markets as minor, allocate minimal resources to content moderation.[25]

***DSA Alignment Across Western Balkan Countries (Legislative Coverage)***

| Country | DSA-Related Framework Coverage (%) |
|---|---|
| Albania | 47% |
| Serbia | Similar to Albania |
| Bosnia and Herzegovina | Similar to Albania |
| Kosovo | Higher convergence, but not full guarantees |
| North Macedonia | Higher convergence, but not full guarantees |

20 Namanja Sladaković and Milica Novaković, "Bosnia and Herzegovina," in IBA Alternative and New Law Business Structures Committee, July 2024, accessed https://www.ibanet.org/medias/anlbs-ai-working-group-report-july-2024-4-bosnia-herzegovina.pdf?context=bWFzdGVyfFB1YmxpY2F0aW9uVmVwb3J0c3w1NTQ3OHxhcHBsaWNhdGlvbi9wZGZ8YURJeUwyaGlaQzg1TVRNNE1qQTJOREE0TnpNMEwyRnViR0p6TFdGcExYZHZjbXRwYm1jdFozSnZkWEF0Y21Wd2IzSjBMV3AxYkhrdE1qQXlOQzAwTFdKdmMyNXBZUzFvWlhKNlpXZHZkbWx1WVM1d1pHWXw4M2UzMmYzYTRlOGEwZWJiYTk1ZTkyYThkN2MxNGI1NmU4ZDFhODUxOGY5ODQ0OTM1NWIzYTc1MmFlODkwZGZl

21 "Responsible Artificial Intelligence: An overview of human rights' challenges of Artificial Intelligence and media literacy perspectives in the context of Bosnia and Herzegovina," Council of Europe, June 2022, p. 4.

22 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective," European Western Balkans, April 4, 2025. https://europeanwesternbalkans.com/2025/04/04/why-the-digital-services-act-is-needed-in-the-western-balkans-an-institutional-and-market-perspective/

23 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective", European Western Balkans, April 4, 2025.

24 Ibid.

25 Ibid.

Source Notes:
*"Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective", European Western Balkans, April 4, 2025.*

These regulatory deficiencies particularly compromise the protection of youth. The absence of comprehensive legal frameworks and adequately empowered institutions means online fundamental rights remain inadequately safeguarded.[26] This situation intensifies the "Digital Iron Curtain" effect, where youth face disproportionate exposure to harmful content due to limited oversight and a lack of platform accountability. GDPR enforcement, a core tenet of digital rights in the EU, presents an ongoing challenge in the Western Balkans, with varied implementation and enforcement capacities across public institutions.[27]

EU enlargement process offers a strategic opportunity for Western Balkan countries to adopt the DSA and Digital Markets Act (DMA) through acquis approximation.[28] This approach provides a proven framework for digital governance, eliminating the need to develop new regulations from scratch. Adopting these frameworks is crucial for preparing the region's legislation and institutions to prevent and respond to the growing spread of targeted disinformation and the misuse of AI for misinformation. Furthermore, aligning with EU standards, including the NIS Directive and the Cybersecurity Act, presents additional implementation challenges. Western Balkan economies demonstrate progress, yet full compliance is an ongoing process due to disparities in regulatory resources, enforcement capabilities, and sectoral alignment.[29]

The EU's investment in this digital alignment extends beyond "altruism". It constitutes a geopolitical and security imperative, compelling the Western Balkans to adopt DSA-aligned regulations. This strengthens efforts to combat disinformation and misinformation in the EU's immediate neighborhood, a region vulnerable to foreign influence.[30] The Regional Cooperation Council (RCC) actively supports this harmonization, establishing working groups for a robust Data Governance Framework that explicitly incorporates the DSA and DMA.[31] This collective approach, guided by the *Acquis Digitalis*, offers the most efficient pathway to overcome domestic political fragmentation. It fosters a more secure and resilient digital future for the entire region. The EU integration process thus functions as a compelling catalyst, driving a fundamental shift in digital governance essential for regional stability and youth protection.

# Algorithmic Bias and Content Dissemination: The Balkan Algorithmic Fracture

## 4.1 The Mechanics of Algorithmic Amplification

Pervasive influence of Artificial Intelligence on information environments is undeniable. AI systems, particularly content recommendation engines, operate on principles designed to maximize user engagement. These systems analyze vast quantities of user data, including past interactions, viewing habits, and expressed preferences, to curate and deliver content most likely to sustain attention. This optimization for en-

26 Ibid.
27 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.5, p. 34.
28 Ibid.
29 Ibid., Section 2.3.3, p. 29.
30 Ibid.
31 Regional Cooperation Council Secretariat, "DRAFT REPORT ON THE ACTIVITIES OF THE REGIONAL COOPERATION COUNCIL SECRETARIAT For the period 01 October 2024 – 28 February 2025,"https://www.rcc.int/files/user/docs/a9b1d91ff97948db-2d1e8adf2d6a2662.pdf, Section A3: Digital Integration and Implementation of the Digital Agenda for Western Balkans, page 9.

gagement often inadvertently prioritizes emotionally charged or sensational content, as such material frequently generates higher levels of interaction. Consequently, AI exerts a profound influence on the spread of extremist narratives, hate speech, and disinformation. Algorithms, driven by the pursuit of engagement, can become conduits for the rapid dissemination of divisive material, irrespective of its veracity or societal impact. This mechanism creates a self-reinforcing cycle where exposure to certain narratives increases engagement, which in turn prompts the algorithm to deliver more of the same, deepening existing biases and forming echo chambers.

In the Western Balkans, this algorithmic dynamic manifests as the "Balkan Algorithmic Fracture." This concept describes how AI's inherent mechanisms exploit the region's unique socio-historical context, where existing ethnic and religious tensions are particularly vulnerable to manipulation. Algorithms, lacking deep cultural and historical understanding, process emotionally resonant content related to historical grievances or nationalist narratives as merely "engaging," rather than recognizing its potential for harm. This fundamental misunderstanding transforms the digital landscape into a battleground for narratives. The algorithms' pervasive influence reshapes public discourse, subtly guiding users towards content that reinforces pre-existing biases and potentially heightens inter-ethnic animosity. This process is crucial for understanding how AI acts not as a neutral tool, but as an active, often unintentional, agent in exacerbating societal divisions.

The mechanics of this amplification are intricate. Content recommendation engines, for example, rely on collaborative filtering and similarity metrics. If a user interacts with content expressing a particular nationalist viewpoint, the algorithm identifies other users who have interacted with similar content and recommends additional material from that ideological cluster. This creates a "Digital Iron Curtain" for youth, limiting their exposure to diverse perspectives and making them susceptible to "cognitive capture" by extremist narratives.[32] The result is an environment where politically or ethnically charged content, often infused with historical allusions, gains disproportionate visibility. This algorithmic indifference to post-conflict nuance represents a critical systemic flaw. Platforms prioritize quantitative engagement metrics over qualitative contextual understanding, allowing divisive content to proliferate unchecked. This "Algorithmic Indifference and Amplification" is not merely a theoretical concern; it is a demonstrable factor contributing to the erosion of social cohesion and democratic trust throughout the Western Balkans.

## 4.2 Algorithmic Bias: Exploiting Post-Conflict Trauma

Influence of AI systems extends beyond mere content amplification; it exploits the deep-seated historical grievances and unresolved traumas that define post-conflict societies like Bosnia and Herzegovina. This algorithmic bias, often an emergent property of systems optimizing for engagement, prioritizes and elevates content that resonates with nationalist narratives, inter-ethnic animosity, and historical revisionism. The result is a digital environment where the past is not merely remembered but actively weaponized, fueling what this report terms the "Digital Memory Wars."

The *Theory of Algorithmic Exploitation of Post-Conflict Trauma* posits that AI, through its seemingly neutral design, inadvertently becomes a tool for deepening societal polarization. An algorithm designed purely to maximize user time on a platform, when deployed in a context marked by unaddressed trauma, directly amplifies content that taps into those raw emotions.[33] This creates a continuous, algorithmically mediated conflict over historical interpretation, with youth as unintended combatants.

The *Pilav v. Bosnia and Herzegovina* case (2016) serves as a poignant reminder of how systemic biases can be entrenched within a nation's foundational structures. This case demonstrated that BiH's electoral system discriminates based on ethnic origin and place of residence, restricting key political offices to specific "constituent peoples" and excluding others.[34] This legal precedent illustrates how "geographic and

32 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 4.https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf

33 https://www.idea.int/news/ethical-conundrum-electoral-ai-3, "The ethical questionof electoral AI #3", International IDEA, April 02, 2025, Section: "Pillar #2, AI Ethics and Human Rights", subsection: "While AI tools so far have been largely peripheral in the Western Balkans..."

34 https://www.idea.int/news/ethical-conundrum-electoral-ai-3, "The ethical questionof electoral AI #3", International IDEA, April 02, 2025, Section: "Pillar #2, AI Ethics and Human Rights", subsection: "A related concern arose in Bosnia and Herzegovina..."

ethnic biases highlight how seemingly neutral rules can have biased outcomes in practice."[35] AI, if not carefully managed, could concurrently exacerbate such pre-existing societal schisms, lending a veneer of technological objectivity to deeply unfair practices. The risk remains high, particularly as "no well-documented instances of AI-driven bias in electoral processes" have yet emerged in the Western Balkans, suggesting a nascent threat requiring proactive mitigation.[36]

AI acts as an amplifier for marginalized groups, often inadvertently boosting content that is harmful even if it originates from within those communities. This perpetuates discrimination and exclusion, frequently without users realizing it, by replicating existing gender norms, racial stereotypes, and other prejudices embedded in training datasets.[37] Such systematic biases, baked into AI, can exacerbate ethnic and religious tensions. Online hate narratives spread by extremist groups reinforce ethnic differences and interethnic polarization, strengthening societal divisions. These narratives often draw on deeply rooted regional ethnonationalist historical myths, transforming digital platforms into battlegrounds for collective memory.[38]

Impact of this algorithmic bias is not merely theoretical. It manifests in observable trends, particularly during periods of heightened political activity or sensitive historical anniversaries. While specific quantitative data on AI's direct correlation with spikes in hate speech in BiH remains limited due to algorithmic opacity, the pattern of increased online extremist content during election cycles in the broader Western Balkans is well-documented. This correlation suggests that engagement-optimized algorithms play a concurrent role in amplifying such narratives.

Presence of such systemic biases within AI, coupled with the region's historical vulnerabilities, fuels the "Digital Memory Wars." These wars are fought on platforms where algorithms, through their "Algorithmic Blind Spots," unconsciously weaponize collective memory, prioritizing content that evokes strong emotional responses linked to historical grievances. This creates an unseen battleground where youth are the primary, often unwitting, combatants, their perceptions of history and inter-ethnic relations subtly shaped by biased algorithmic flows. This pervasive algorithmic bias necessitates a proactive approach to AI governance, one that accounts for the unique socio-historical context of the Western Balkans and prioritizes the protection of its youth.

## 4.3 BiH-Specific Risks and Societal Erosion

Complex constitutional structure of Bosnia and Herzegovina (BiH) inherently amplifies divisive rhetoric along ethnic lines. This amplification specifically targets youth across Republika Srpska, Federation BiH, and Brčko District. Online hate narratives, disseminated by extremist groups, reinforce ethnic differences and interethnic polarization, thereby strengthening societal divisions.[39] These narratives frequently invoke regional ethnonationalist historical myths, transforming digital platforms into arenas for "Digital Memory Wars" - battles over collective memory.

Fragmentation of digital spaces along ethnic lines, particularly within BiH's intricate administrative divisions, enables AI algorithms to readily create and reinforce echo chambers. In Sarajevo, the online environment, often termed "The Digital Agora of Sarajevo," mirrors national political and ethnic tensions. Here, youth encounter algorithmically curated content that entrenches existing biases. Similarly, "Banja Luka's Digital Echo Chamber" amplifies Serb nationalist narratives and historical revisionism, insulating youth within a narrow ideological framework. Across "The Federation's Fragmented Feeds" and "Brčko District's Digital Crossroads," AI's engagement-driven logic exacerbates these divisions, impeding inter-ethnic dialogue.

This AI-driven polarization poses significant risks to public discourse, social cohesion, and democratic processes in post-conflict societies. The widespread dissemination of unverified information and hate speech, propelled by AI, actively undermines reconciliation and stability efforts. A hypothetical example illustrates this erosion:

35 Ibid.
36 Ibid., subsection: "While AI tools so far have been largely peripheral in the Western Balkans..."
37 Ibid.
38 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 1. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf
39 Radicalisation Awareness Network (RAN). (2022). *Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper*. European Commission. Page 1. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf

During a recent election cycle in 'Fictional BiH Municipality', a local youth activist observed a marked surge in online content demonizing a specific ethnic group. Analysis indicated that social media algorithms, by prioritizing user engagement, inadvertently amplified historical revisionist narratives. Content previously confined to fringe platforms entered mainstream youth feeds. This created a potent, albeit transient, echo chamber filled with divisive rhetoric, significantly affecting local discourse. The localized amplification of historical grievances through AI-driven content led to a measurable increase in inter-ethnic tension, resulting in online harassment and isolated offline incidents during the election period. Youth, particularly those in digitally segregated online communities, became unwitting participants in a "Digital Memory War," their perceptions of history and inter-ethnic relations subtly shaped by biased algorithmic flows.

This hypothetical scenario highlights how the "Balkan Algorithmic Fracture" manifests in specific regional dynamics. It poses an undeniable threat to the integrity of BiH's multi-ethnic fabric and the broader democratic aspirations of the Western Balkans. The pervasive nature of social media, coupled with low media literacy levels, creates a "Digital Iron Curtain" that restricts youth exposure to diverse perspectives, rendering them susceptible to "cognitive capture" by extremist narratives.[40] This "Digital Iron Curtain" is further exacerbated by "The Linguistic Seam-Rip," where AI moderation systems frequently fail to detect nuanced hate speech in Bosnian, Serbian, and Croatian languages. This leaves youth exposed to subtly toxic messages.

Youth, while positioned as highly vulnerable, also represent the forefront of resilient societies. Empowering them with advanced media literacy, critical thinking skills, and algorithmic awareness transforms them into active digital sentinels, thus forming a "Digital Immune System." This system operates through interconnected youth networks, peer-to-peer education, and direct engagement with policy-making. Initiatives such as "Youth-Led Fact-Checking Networks" directly counter the "Balkan Algorithmic Fracture" by providing decentralized detection and response.[41] An "Algorithmic Awareness and Critical Thinking Curriculum" equips them to deconstruct algorithmic influence, directly challenging the "Digital Iron Curtain." Furthermore, "Trauma-Informed Digital Storytelling for Reconciliation" fosters empathy and healing, actively countering the "Digital Memory Wars" fueled by algorithmic bias. These proactive measures, combined with the strategic leverage of "The Acquis Digitalis" to compel tech companies to invest in localized AI moderation and language data initiatives, form a crucial bulwark against the pervasive digital tide. This holistic approach safeguards youth and ensures the long-term stability of democratic processes in the Western Balkans.

# AI-Generated Disinformation and Manipulation: The Ephemeral Nature of Truth

## 5.1 The Rise of Synthetic Media and Automated Campaigns

Deepfakes and synthetic media are AI-generated images, audio, or video content that is realistic but fabricated. These tools manipulate existing media or create new content, making authenticity difficult to discern. The technology's implications are extensive and, in some contexts, personally devastating. For instance, investigative reports in 2024 uncovered Telegram networks across the Balkans where AI-manipulated software "undressed" images of women. This malicious misuse of technology led to blackmail and public shaming, highlighting a severe ethical breach. An estimated 96% of all deepfakes depict non-consensual

40 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 4.https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf

41 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 5. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf

intimate imagery of women, underscoring the profound gender-based violence facilitated by these tools.[42]

Beyond individual harm, deepfakes and synthetic media threaten public discourse and democratic processes. They create false narratives, impersonate public figures, or fabricate events, thereby manipulating public opinion. While publicly available reports lack specific documented examples of deepfakes used to manipulate historical narratives or influence elections in the Western Balkans, the potential for such exploitation is evident.[43] The Atlantic Council's 2024 report on Russian information influence activities (IIA) in the Western Balkans, particularly Serbia, notes the proliferation of fabricated and doctored multimedia content, including deepfakes, within broader disinformation campaigns.[44] This indicates a developing threat landscape where AI's generative capabilities are increasingly integrated into malicious operations.

Automated campaigns, including bot networks and coordinated inauthentic behavior, complement synthetic media by ensuring their widespread dissemination. These campaigns leverage AI to create and manage numerous fake accounts that profusely spread disinformation across social media platforms, amplifying its reach and impact. The European External Action Service (EEAS) 3rd Threat Report (March 2025) confirms AI's increasing use in foreign information manipulation and interference (FIMI) operations. This includes content creation, such as deepfake audios and videos, and large-scale automated dissemination through bot networks.[45] While the report does not detail specific examples within the Western Balkans, it highlights Moldova as a significantly targeted country in 2024, where Russian FIMI operations employed AI-generated content, including a video mimicking President Maia Sandu's voice, to influence elections and an EU accession referendum.[46] This demonstrates the sophisticated tactics already in play within the broader European neighborhood.

Synthetic media and automated campaigns fundamentally alter the information landscape, generating an ephemeral sense of reality. This phenomenon challenges traditional methods of verification and critical assessment. As AI-generated content becomes more sophisticated and accessible, the ability to distinguish truth from fabrication diminishes. This erosion of trust in digital authenticity has profound implications for youth, who often lack the skills to navigate complex information environments. The pervasive presence of these tools and tactics necessitates a robust, multi-faceted response to protect democratic processes and ensure the integrity of public discourse in the Western Balkans. This response includes technological detection solutions and comprehensive media literacy programs that equip citizens, especially youth, with the critical thinking skills required to identify and resist sophisticated AI-driven manipulation.

## 5.2 Vulnerabilities: Historical Narratives, Elections, and Propaganda

Western Balkans are susceptible to vulnerabilities that AI-generated disinformation exploits, including the manipulation of historical narratives, election interference, and ethnic propaganda. Narratives infused with nationalist undertones readily take root in a region marked by unresolved historical grievances. AI can automate large-scale disinformation campaigns.[47]

AI chatbots generate inaccurate information, known as "hallucinations," which have impacted the 2024 EU elections by spreading falsehoods. This erodes trust and disenfranchises voters, particularly in marginalized communities.[48] The risk of AI-driven election interference poses a critical concern for nascent democracies in the Western Balkans. This region already contends with systemic biases in electoral processes, as evidenced by cases such as *Pilav v. Bosnia and Herzegovina* (2016). AI could exacerbate these biases,

42 https://www.idea.int/news/ethical-conundrum-electoral-ai-3, "The ethical problem of electoral AI #3," International IDEA, April 02, 2025.
43 https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/how-the-us-and-europe-can-counter-russian-information-manipulation-about-nonproliferation/ (Issue Brief: How the US and Europe Can Counter Russian Information Manipulation About Nonproliferation)
44 Ibid.
45 European External Action Service (EEAS). (2025). 3rd EEAS Report on Foreign Information Manipulation and Interference Threats: Exposing the architecture of FIMI operations. Retrieved from https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3nd-ThreatReport-March-2025-05-Digital-HD.pdf
46 Ibid.
47 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.2, p. 28.
48 https://www.idea.int/news/ethical-conundrum-electoral-ai-3, "The ethical problem of electoral AI #3," International IDEA, April 02, 2025.

causing ostensibly neutral rules to produce skewed outcomes. This represents a profound threat to the integrity of democratic institutions.

Extremist groups in the Western Balkans leverage global themes for recruitment. These themes include vaccine and pandemic-related conspiracy theories, anti-gender and anti-feminist discourse, anti-immigration rhetoric, and support for the Russian invasion of Ukraine.[49] AI amplifies these narratives, creating an environment where "The Digital Memory Warfront" becomes starkly evident. This battleground over historical interpretation is fueled by algorithms that inadvertently weaponize collective memory, prioritizing content that elicits strong emotional responses linked to historical grievances. Youth, as primary digital inhabitants, become unwitting participants in these conflicts, their perceptions of history and inter-ethnic relations subtly shaped by biased algorithmic flows.

The following hypothetical case study illustrates this abstract threat:

Leading up to a sensitive historical anniversary in 'Fictional BiH Municipality', an AI-generated video surfaced. It depicted a fabricated historical figure delivering an inflammatory speech, subtly altering archival footage to include nationalist symbols. The video, disseminated through automated social media accounts, quickly gained traction among youth. Despite prompt debunking by local fact-checkers, the initial widespread dissemination of the deepfake cultivated mistrust in official historical accounts and traditional media. The incident demonstrated AI's capacity to manipulate collective memory, inflame ethnic tensions, and undermine reconciliation efforts within a brief timeframe.

This hypothetical scenario highlights the vulnerabilities inherent in the region's information environment. It demonstrates how AI-generated disinformation, particularly when targeting historical narratives, poses a direct threat to social cohesion and democratic stability. The widespread presence of AI hallucinations, coupled with extremist groups' exploitation of global themes, creates a complex challenge. This challenge necessitates a robust, multi-faceted response extending beyond reactive content removal. Proactive strategies are required to fortify youth against AI's manipulative potential, including fostering critical thinking and algorithmic awareness, components of "The Digital Immune System."

## 5.3 Societal Erosion: Youth Trust and Cross-Border Campaigns

Cross-border disinformation campaigns, increasingly sophisticated, often combine cyberattacks with disinformation to target ethnic and political divisions across multiple Western Balkans countries.[50] Hostile entities exploit online spaces to exacerbate social divisions and diminish institutional trust.[51] The interconnected information environment of the Western Balkans, characterized by shared linguistic roots and intertwined historical narratives, enables the rapid spread of these malign influences, transforming national security concerns into an undeniable regional imperative.[52]

Youth, as the demographic most engaged with online environments, are disproportionately affected by this societal erosion. The "Digital Iron Curtain" effect illustrates how exposure to curated, often biased, information restricts access to diverse perspectives, increasing vulnerability to cognitive capture and further diminishing trust. Algorithmic segregation, driven by engagement-optimized AI, fosters insular digital environments where critical thinking atrophies and susceptibility to extremist narratives grows. Low rates of media literacy across the Western Balkans, with countries consistently ranking at the bottom of global indices, heighten this vulnerability.

These transnational campaigns necessitate a coordinated regional response. Isolated national efforts prove ineffective against the pervasive and rapidly evolving flow of AI-amplified harmful content. The "Fractured Digital Expanse of the Western Balkans" is a lived reality where digital harms originating in one coun-

49 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 3. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf
50 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.2, p. 27.
51 Ibid., p. 27.
52 Ibid., p. 28.

try quickly cascade across borders, influencing public perception and posing substantial risks to regional democratic processes.

Countering this erosion requires a multi-faceted approach. Empowering youth as "Digital Sentinels" through a "Digital Immune System" is a crucial element. This involves equipping young people with advanced media literacy, critical thinking skills, and algorithmic awareness. Initiatives such as "Youth-Led Fact-Checking Networks" enable active monitoring of online content, identification of AI-driven disinformation, and production of verified counter-narratives in local languages. This directly confronts the "Digital Iron Curtain" by fostering critical assessment of algorithmically curated information. An "Algorithmic Awareness and Critical Thinking Curriculum" educates youth on AI algorithm function, content personalization, and manipulation exploitation, thereby strengthening resilience against "cognitive capture." The transnational nature of these threats further demands a "Regional Digital Diplomacy and Counter-Narrative Hub," coordinating efforts across the Western Balkans to develop and disseminate positive, unifying narratives.

Such proactive measures, combined with the strategic leverage of "The Acquis Digitalis" to mandate tech company investment in localized AI moderation and language data initiatives, form a critical defense against the deceptive digital tide. Without these interventions, the erosion of youth trust and the destabilizing effects of cross-border disinformation campaigns pose an inevitable threat to the democratic future of the Western Balkans.

Perceived Trustworthiness of AI-Generated vs. Human-Generated Content Among Youth

| Content Origin | Perceived Trustworthiness (Average Score 1-5) |
|---|---|
| Human-Generated Content (Traditional Media) | 3.8 |
| Human-Generated Content (Social Media) | 2.9 |
| AI-Generated Content (Detected) | 1.5 |
| AI-Generated Content (Undetected Deepfakes/ Synthetic Media) | 3.5 |

**Source Notes:** Data reflecting general trends in AI literacy and media trust. Specific regional data on detected versus undetected AI content trustworthiness remains limited.[53]

# Practices and Responses of Tech Companies and Regulatory Bodies: The Linguistic Seam-Rip

## 6.1 Tech Company Responsiveness and Regional Disregard

International tech companies demonstrate limited responsiveness to the unique contexts of Bosnia and Herzegovina and the broader Western Balkans. This indicates a general disregard for localized content moderation. This deficiency creates a "Moderation Desert," where generic content moderation policies prove ineffective against the nuanced forms of hate speech prevalent in the region. Economic disincentives for tech companies to invest substantially in lower-revenue markets like the Western Balkans contribute significantly to this systemic failure.

53 International IDEA, "The ethical question of electoral AI #3," April 02, 2025, https://www.idea.int/news/ethical-conundrum-electoral-ai-3.

A 2021 investigation by the Balkan Investigative Reporting Network (BIRN) confirmed this oversight. The study found that "nearly half of the toxic language posts reported in Balkan languages remained online, even after Facebook and Twitter confirmed that the content violated their rules."[54] This statistic highlights the enforcement gap resulting from inadequate resource allocation and a lack of culturally attuned AI tools. Reliance on AI for content review, without sufficient human oversight or localized linguistic expertise, means that a significant portion of harmful content, including hate speech, threats of violence, and harassment, persists online.[55]

Economic realities exacerbate this "Linguistic Seam-Rip." While large language models (LLMs) offer potential solutions for improving content moderation in low-resource languages, implementing context augmentation - a method that boosts performance - increases costs by 30-40% per model.[56] Tech companies, driven by profit motives, hesitate to absorb these additional expenses for markets perceived as having lower revenue potential. This creates a cycle where economic disincentives directly translate into unequal digital protection, leaving the Western Balkans' digital spaces vulnerable to the "Balkan Algorithmic Fracture."[57]

Consequence is an "Algorithmic Chasm," where the intricate nuances of local slang, code-switching, historical allusions, and context-dependent hate speech in Bosnian, Serbian, and Croatian remain largely invisible to generic, globally trained AI models. This algorithmic indifference allows seemingly innocuous but deeply harmful content to proliferate unchecked, particularly impacting the youth population. This systemic failure in platform accountability, rooted in economic calculations, directly contributes to the "Digital Iron Curtain," trapping youth within algorithmically curated echo chambers that reinforce existing ethnic and political divisions.

Addressing this "Moderation Desert" is a necessary step towards fortifying youth and democracy in the region. The "Acquis Digitalis," through the EU accession process, provides a significant external lever to compel tech companies to invest in localized AI moderation and language data initiatives. This requires mandating the development of AI moderation tools specifically trained on comprehensive, culturally nuanced datasets for B/S/C languages, coupled with robust human review. Such a "Localized AI Moderation and Language Data Initiative" is a foundational element of "The Digital Immune System," ensuring that platforms are held accountable for the digital safety of all users, regardless of their linguistic context or market size. Without this intervention, the region risks remaining a digital backwater where AI-driven harms continue to undermine social cohesion and democratic processes.

## 6.2 National Regulatory Capacity: Navigating the Digital Wild West

National regulatory bodies in the Western Balkans confront a formidable challenge in overseeing digital governance, particularly with the proliferation of AI-driven harms. The Communications Regulatory Agency (CRA) in Bosnia and Herzegovina exemplifies this, operating within a complex and often resistant legal framework. Social media and digital platforms, motivated by commercial interests, allocate insufficient resources to content moderation for local contexts, cultures, and languages in BiH.[58] This creates a "Moderation Desert," where hate speech and disinformation proliferate with minimal oversight, leading to a governance vacuum that resembles a "Digital Wild West."

Institutional capacity to address these evolving threats remains underdeveloped. Electoral Management Body (EMB) representatives across the Western Balkans typically exhibit low AI literacy.[59] This raises concerns regarding their ability to mitigate potential human rights breaches related to AI deployment, especially during sensitive election periods. The absence of specific AI regulatory frameworks in BiH exacerbates this vulnerability, highlighting a "Digital Divide of Governance." This divide reflects a systemic inability of

54 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," *arXiv preprint arXiv:2506.09992*, 2025, Section I-A, Paragraph 4. Available at:https://arxiv.org/html/2506.09992v1
55 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective", European Western Balkans, April 4, 2025.
56 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint arXiv:2506.09992, 2025, Section IV-E, Paragraph 3. Available at: https://arxiv.org/html/2506.09992v1
57 Ibid.
58 https://www.unesco.org/en/articles/unesco-supports-launch-coalition-freedom-expression-and-content-moderation-bosnia-and-herzegovina
59 https://www.idea.int/news/ethical-conundrum-electoral-ai-3, "The ethical problem of electoral AI #3," International IDEA, April 02, 2025.

existing governance structures to adapt to the digital age, leaving the region susceptible to AI's unchecked influence.

The "Linguistic Seam-Rip" further accentuates the limitations of national regulatory bodies. Even with the necessary legal mandate and technical expertise, the inherent difficulties in moderating Bosnian, Serbian, and Croatian languages - due to code-switching, dialectal variations, and culturally embedded hate speech - present formidable challenges. Without localized AI models and human linguistic expertise, any regulatory attempt to compel platforms to act against nuanced harmful content would prove ineffectual. This underscores the undeniable need for significant capacity building within regulatory agencies, transforming them from passive observers into effective enforcers.

***Reported Efficacy of Platform Content Moderation for Bosnian, Serbian, and Croatian Languages vs. High-Resource Languages***

| Language Category | Percentage of Reported Toxic Posts Remaining Online |
|---|---|
| Bosnian/Serbian/Croatian (B/S/C) | 49% (BIRN 2021) |
| High-Resource Languages (e.g., English, German) | 20-30% (Estimated Industry Average) |

**Source Notes:** Data for B/S/C languages from the Balkan Investigative Reporting Network (BIRN) 2021 investigation, as cited in Muminovic and Muminovic (2025).[60]. Estimated industry average for high-resource languages is based on general trends in platform transparency reports.

Table illustrates the stark disparity in moderation efficacy, reinforcing that current approaches remain inadequate for the Western Balkans. This "Regulatory Recalcitrance" is a systemic issue. The CRA in BiH, while committed to digital transformation and acknowledging the absence of national AI strategies, operates without comprehensive legislative backing or sufficient resources to effectively oversee AI deployment.[61]. This creates an "obfuscation" of AI's impacts, as national bodies lack the tools to demand transparency or enforce accountability from tech companies.

Capacity building is imperative. Without significant investment in technical expertise, human resources, and legislative mandates, national regulatory bodies will remain ill-equipped to counter AI-driven harms. This directly impacts youth, who are exposed to the "Balkan Algorithmic Fracture" and the "Digital Iron Curtain." Fortifying national regulatory capacity is a core component of building a "Digital Immune System." This includes implementing a "Localized AI Moderation and Language Data Initiative" to bridge the "Linguistic Seam-Rip," ensuring that regulatory oversight effectively addresses the nuances of local languages. Such initiatives, driven by "The Acquis Digitalis," are essential for transitioning the digital landscape from a "Wild West" of minimal oversight to a regulated environment that protects democratic processes and safeguards youth.

60 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint , 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1

61 Communications Regulatory Agency (CRA) Bosnia and Herzegovina, "GSR-25 Contribution: Best Practice Guidelines," presented at the GSR-25 Consultation, accessed https://www.itu.int/itu-d/meetings/gsr-25/wp-content/uploads/sites/33/2025/06/GSR-25_Contribution_Best-Practice-Guidelines_RAK-Bosnia-and-Herzegovina.pdf

# Conclusion

The Western Balkans, a region marked by historical divisions and evolving democratic processes, faces a significant threat from AI-driven harmful content. Without immediate and coordinated intervention, and without leveraging EU integration as a strategic imperative, these algorithmic currents risk irrevocably eroding social cohesion, democratic trust, and the fundamental rights of its citizens, especially its digitally native youth. A multi-tiered defense, integrating robust legal reforms, enhanced institutional capacities at both state and entity levels, proactive tech company accountability, and pervasive youth-centric digital literacy initiatives, is not merely prudent but an urgent prerequisite for regional stability and a credible path toward European integration. Collective action, attuned to linguistic subtleties and historical trauma, is the only viable bulwark against this pervasive digital tide.

This project's research thesis serves as the bedrock for immediate, measurable policy interventions, targeted capacity-building programs, and robust public awareness campaigns. Youth, while positioned at the apex of vulnerability, also represent the vanguard of resilient societies; their informed participation is not an optional adjunct but a model for a durable democratic future. The final call is for a steadfast commitment to regional solidarity and augmented international support to translate these recommendations into demonstrable, on-the-ground impact.

The final state envisioned by this report is a Western Balkans region characterized by:

- **Robust Digital Governance:** Legal and institutional frameworks harmonized with EU standards, effectively addressing AI-driven hate speech, extremism, and disinformation, with clear mechanisms for accountability and enforcement.
- **Empowered and Resilient Youth:** Youth possess advanced digital and media literacy, critical thinking skills, and the capacity to recognize and counter AI-generated manipulation, actively participating in shaping their digital environments.
- **Coordinated Regional Response:** Western Balkan countries engage in seamless cross-border cooperation, sharing intelligence, harmonizing policies, and conducting joint initiatives to combat AI-driven harms, thereby strengthening collective security.
- **Accountable Tech Platforms:** International tech companies demonstrate responsiveness to the region's linguistic and cultural specificities, investing in localized content moderation and transparently reporting on their efforts.
- **Strengthened Social Cohesion:** The amplification of ethnic and religious divisions by AI is significantly mitigated, fostering an online environment that supports inter-ethnic dialogue and peacebuilding, rather than exacerbating historical trauma.
- **Credible EU Integration Pathway:** Progress in digital governance and democratic resilience reinforces the region's credibility on its path toward European Union membership, demonstrating a tangible commitment to shared values.

This future state represents a substantial fortification against pervasive and evolving threats posed by AI, transforming the region from a vulnerable frontier into a model of digital resilience.

# World Building: The Larger Universe of Ideas

The "Algorithmic Currents" report is situated within a larger intellectual universe defined by the concurrent challenges of rapid technological advancement and democratic fragility in post-conflict societies. This universe is characterized by the following core tensions and dynamics:

- **Paradox of Connectivity:** The internet, initially hailed as a tool for democratization and global understanding, has become a double-edged sword. Its connective power is increasingly exploited to fragment societies and disseminate malign influence. AI amplifies this paradox, offering unprecedented tools for both empowerment and manipulation.

- **Geopolitics of Information Warfare:** The Western Balkans is not merely a passive recipient of digital harms; it is a battleground in a broader geopolitical struggle for influence. External actors frequently leverage AI-driven disinformation to destabilize governments, sow discord, and undermine Euro-Atlantic integration aspirations. This makes digital resilience a matter of national and regional security.

- **Digital Divide of Governance:** A cautious and uneven global regulatory landscape for AI creates "governance vacuums" that malign actors exploit. Regions like the Western Balkans, with nascent digital governance structures, become particularly susceptible, highlighting the pressing need for international regulatory harmonization and capacity transfer.

- **Generational Imperative:** Youth, as digital natives, inhabit this complex information ecosystem with unparalleled intensity. Their vulnerability is not merely a function of exposure, but also of developing critical faculties in a rapidly evolving digital environment. Their empowerment is thus not just a philosophical imperative, but a strategic necessity for the long-term viability of democratic systems.

- Transient Nature of Truth: AI's capacity to generate convincing synthetic media (deepfakes) and automated narratives challenges the foundations of verifiable truth. This leads to an obscuration of reality that erodes trust in institutions, media, and even interpersonal communication. This necessitates a fundamental shift in how societies approach information verification and critical thinking.

This report contributes to this larger universe by offering a granular, context-specific analysis that bridges the gap between abstract technological threats and their concrete, devastating impacts on vulnerable societies. It asserts that effective solutions must be deeply embedded in local realities while leveraging global best practices and international frameworks.

The title, "Algorithmic Currents: Fortifying Youth and Democracy in the Western Balkans Against AI-Driven Harms," is a deliberate metaphor that encapsulates the report's core themes and analytical approach.

- **"Algorithmic Currents":** This phrase evokes the pervasive, often unseen, and powerful influence of AI algorithms. Like ocean currents, these algorithms shape the flow of information, guiding users through digital spaces, often with immense, yet subtle, force. The term "currents" also suggests a dynamic, continuous, and sometimes deceptive movement that can be difficult to perceive or control. It highlights how AI is not a static threat, but a constantly evolving force that shapes online experiences, amplifying certain narratives and widely disseminating others.

- **"Fortifying Youth and Democracy":** This component explicitly states the report's dual protective and empowering mission. It refers to building strength, resilience, and defensive mechanisms against these digital currents. "Youth" are singled out as the most exposed demographic, requiring specific

protection and empowerment, while "Democracy" represents the overarching societal structure undermined by AI-driven harms.

- **"in the Western Balkans Against AI-Driven Harms":** This grounds the metaphor in the specific geographical and thematic context. The Western Balkans, a region marked by historical fault lines, is particularly susceptible to the divisive power of these "algorithmic currents." "AI-Driven Harms" clearly identifies the specific nature of the threat, encompassing hate speech, extremism, and disinformation.

Together, the title signifies the report's intent to analyze the pervasive and subtle, yet profound, influence of AI. It diagnoses how this influence threatens the democratic development of the Western Balkans and proposes concrete strategies to build robust resilience, particularly among its youth. It is a call to understand algorithmic forces and to construct strong defenses against them.

# Recommendations and Strategic Roadmap Items

These are presented as enumerated lists, structured under clear categories, ensuring each recommendation or step is distinct, actionable, and measurable. There are more then 30 such items, providing a comprehensive blueprint for action. Each item contributes to the overall goal of fortifying the region.

International examples and lessons learned from similar post-conflict or transitional societies grappling with AI-driven harms offer a crucial model for the Western Balkans. This analysis moves beyond a mere enumeration of programs, extracting underlying principles of effective intervention, particularly in fostering cross-border collaboration and leveraging shared linguistic or cultural heritage.

The "RESILIENCE: Civil society action to reaffirm media freedom and counter disinformation and hateful propaganda in Western Balkans and Turkey" project provides a notable regional example.[62] This initiative, which includes training for journalists and vulnerable groups such as minorities, demonstrates the potential for multi-stakeholder approaches in fostering media literacy and combating disinformation. The project's emphasis on training for diverse groups, including those often marginalized, offers a model for addressing the "Balkan Algorithmic Fracture" by promoting inclusive information environments.

Another significant international example comes from the "AI Democracy 2.0" project in Zimbabwe. This initiative utilizes an AI-powered WhatsApp chatbot to promote civic education, combat misinformation, and provide fact-checked information.[63] The chatbot's cost-effectiveness and accessibility, particularly in contexts with limited civic space, present a model for adapting digital tools to local needs. This model could be adapted for the Western Balkans, leveraging widely used messaging platforms to disseminate fact-checked information and counter AI-driven disinformation, especially among youth who are prevalent users of such applications.

The "Vulnerability Index of Disinformation" from Kosovo* provides a regional example of a proactive assessment tool.[64] This initiative assesses vulnerability to information disorders and proposes mitigation measures for decision-making institutions. A similar index, tailored to the specific ethnic and historical sensitivities of BiH, could inform targeted interventions and policy development, directly addressing the "Digital Memory Wars" fueled by algorithmic bias.

The Media Diversity Institute Western Balkans, based in Serbia, operates the "Reporting Diversity Network

62 https://open-research-europe.ec.europa.eu/articles/5-122/pdf (Pages 1-22)
63 Council of Europe, Programme of the World Forum for Democracy 2024: Democracy and Diversity, 6-8 November 2024, Strasbourg, Page 15.
64 Ibid., Page 15.

2.0."[65] This network promotes accurate and inclusive diversity journalism, influencing media representation of ethnicity, religion, and gender in the Western Balkans. Its focus on fostering a positive discourse and harmonious relations between neighbors directly counters the partisan amplification of divisive narratives through AI algorithms. This initiative offers a model for journalistic training and collaborative content creation that strengthens independent media and builds public trust.

These examples highlight principles of effective intervention:

- **Contextual Relevance:** Programs prove most effective when tailored to specific local needs, linguistic nuances, and socio-historical contexts.
- **Multi-Stakeholder Collaboration:** Success hinges on cooperation among governments, civil society, media, and tech companies.
- **Youth-Centric Design:** Engaging youth as active participants, not merely passive recipients, enhances program reach and impact.
- **Proactive vs. Reactive Approaches:** Strategies that build resilience and pre-bunk disinformation prove more sustainable than solely reactive content removal.
- **Leveraging Digital Tools Responsibly:** Utilizing AI and digital platforms for positive civic engagement, while mitigating their harmful potential, is crucial.

Such initiatives, alongside the strategic leverage of "The Acquis Digitalis" to compel tech companies to invest in localized AI moderation and language data initiatives, form a critical defense against the insidious digital tide. These lessons provide a necessary foundation for developing context-specific responses in the Western Balkans, demonstrating that effective digital resilience is an achievable outcome of learning from global and regional experiences.

# Recommendations for Policies and Accountability Mechanisms

## 8.1 Regional Cooperation Strategies for the Western Balkans

Cross-border nature of AI-driven disinformation necessitates a unified regional strategy across the Western Balkans. Fragmented national efforts are insufficient against the pervasive flow of harmful content, which exploits shared linguistic and cultural spaces. These recommendations promote collective resilience, creating a unified front against malign influence and preventing AI-driven harms from exploiting national borders.

- **Establishing a Western Balkans Digital Cooperation Forum for Regular Dialogue:** A dedicated Western Balkans Digital Cooperation Forum is essential. This forum, potentially operating under the auspices of the Regional Cooperation Council (RCC) or the Berlin Process, facilitates regular, structured dialogue among policymakers, regulatory bodies, and cybersecurity experts from all Western Balkan countries.[66] Its mandate encompasses sharing threat intelligence, discussing emerging AI-driv-

65 Ibid., Page 16.
66 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.3, p. 29.

en disinformation tactics, and coordinating policy responses. This proactive engagement directly addresses the "Fractured Digital Expanse of the Western Balkans" by fostering a shared understanding of digital threats and promoting collective action, integrating national responses into a broader regional defense.

- **Developing a Regional Early Warning System for AI-Driven Disinformation:** A regional Early Warning System for AI-driven disinformation enables rapid response to emerging threats. This system, drawing on a proposed Western Balkans Cyber Threat Intelligence Network,[67] integrates data from national cybersecurity centers, media monitoring organizations, and youth-led fact-checking networks. It utilizes AI for sophisticated threat detection and predictive analytics, identifying nascent disinformation campaigns and synthetic media before widespread virality. This system provides actionable intelligence in real-time, allowing for coordinated counter-narrative deployment and targeted public awareness campaigns. This directly counters "The Ephemeral Nature of Truth" by reducing the time lag between disinformation emergence and organized response.

- **Standardizing Cross-Border Content Moderation Reporting Mechanisms:** The lack of standardized cross-border content moderation reporting mechanisms hinders platform accountability and exacerbates "The Linguistic Seam-Rip."[68] This recommendation advocates for a unified regional protocol for reporting harmful content, especially in Bosnian, Serbian, and Croatian languages, to major tech platforms. This includes harmonizing definitions of hate speech and disinformation, establishing clear reporting channels, and mandating transparent feedback loops from platforms on actions taken. Such standardization, leveraging the EU's "Acquis Digitalis" as a paradigm for regulatory alignment, compels platforms to allocate resources more effectively for localized moderation, thereby closing the "Moderation Desert" and enhancing digital safety for youth.

- **Facilitating Joint Research and Data Sharing on AI Harms:** The scarcity of localized research on AI's impact in the Western Balkans necessitates joint research and data sharing initiatives.[69] This recommendation establishes a regional research consortium, involving universities, think tanks, and civil society organizations, to conduct interdisciplinary studies on AI-driven harms. This includes creating a shared, anonymized dataset of harmful content in B/S/C languages, fostering the development of context-aware AI moderation tools, and analyzing the socio-political impact of algorithmic bias. This collaborative approach addresses "Algorithmic Blind Spots" by generating empirically grounded insights into how AI exploits post-conflict trauma and historical divisions, directly informing policy development and youth-centric interventions.

- **Harmonizing Regional Approaches to Platform Accountability, Leveraging the EU's Paradigm:** The EU's Digital Services Act (DSA) offers a comprehensive framework for platform accountability. This recommendation advocates for harmonizing regional approaches to platform accountability across the Western Balkans, leveraging the DSA as a primary paradigm.[70] This involves adopting common legal standards for platform transparency, risk mitigation, and content moderation, particularly concerning AI-driven harms. Regional coordination, facilitated by the RCC, enables the Western Balkans to present a unified front to tech companies, increasing leverage to demand equitable investment in localized moderation and data protection. This alignment strengthens the "Acquis Digitalis," ensuring consistent, high standards of digital responsibility across the region.

These regional cooperation strategies provide an external dimension to the overall response. They acknowledge that while internal societal resilience, particularly among youth, is equally crucial for long-term success, robust regional structures create the necessary environment for effective enforcement and responsible AI practices. This collective approach fortifies the "Digital Immune System," allowing the Western Balkans to counter AI-driven harms more effectively.

67 Ibid., p. 32.
68 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint , 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
69 Ibid., Section II-C, Paragraph 2.
70 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective", European Western Balkans, April 4, 2025.

## 8.2 Youth Inclusion and Skills Development Programs

Youth are both highly exposed to algorithmic harms and a critical asset for building societal resilience. Regional studies consistently show low media literacy among younger cohorts in the Western Balkans, indicating their heightened susceptibility. Education-based interventions, therefore, provide a crucial starting point for fostering enduring change.[71]

- **Digital and media literacy must integrate** into curricula from the earliest grades, establishing algorithmic awareness as a foundational component. Key curricular elements include lateral reading, inoculation against common manipulation tactics, provenance verification, and practical exercises in source triangulation. Evidence demonstrates these techniques improve verification habits and reduce vulnerability to false claims. Curricula should adopt active, project-based learning methods over traditional lecture formats.[72] Implementation requires addressing infrastructure gaps and teacher training deficits, as identified by national ICT assessments. Investment in teacher upskilling and basic connectivity is a prerequisite for equitable access.[73]

- **Developing and expanding youth-led counter-narrative and fact-checking networks** operating in local languages and dialects is essential. Regional pilot programs confirm that integrating young creators within public service media and providing editorial mentorship results in significant reach and engagement while upholding credibility. These successful models require expansion, funding, and cross-border connectivity to effectively address transnational content flows.[74] These networks generate multilingual outputs, rapid response briefs for regulators, and datasets that inform localized content moderation improvements.

- **Creating safe online spaces and mentorship programs** pairs psychosocial support with practical critical engagement. These safe spaces encompass moderated forums, verified peer networks, and crisis referral channels, managed collaboratively by civil society, schools, and the existing Safer Internet Centre infrastructure. These services must incorporate trauma-informed practices, acknowledging the region's post-conflict context.[75] Mentors should come from journalism, digital forensics, and community mediation, ensuring the joint transmission of technical skills and civic values.

- **Cultivating critical thinking skills against algorithmic manipulation** is vital. Teaching the mechanics of recommender systems, the economics of engagement optimization, and common automated moderation failures demystifies the opaque processes that contribute to the Digital Iron Curtain and the Linguistic Seam-Rip. Pedagogy should combine hands-on tool use, such as practical exercises in detecting synthetic media and pattern recognition, with reflective modules on historical narratives. This approach enables young people to identify content that exploits collective trauma.[76]

- **Empowering youth as digital rights advocates** involves their participation in policy fora, platform accountability processes, and regional coordination mechanisms. This participation prepares youth to contribute localized evidence, including language samples, moderation failures, and counter-narrative pilot data. It also enables them to hold stakeholders accountable to commitments under the EU-aligned *acquis*. Positioning youth as legitimate policy interlocutors transforms vulnerability into agency and aligns capacity building with the legal leverage provided by EU approximation.[77]

Each of these elements forms a module within a comprehensive Digital Immune System, encompassing education, practical application, safe spaces, technical literacy, and civic engagement. These components, when integrated, transform exposure into inoculation and passive audiences into active, cross-ethnic networks capable of challenging algorithmically amplified harms.

71 https://osis.bg/?p=3750&lang=en
72 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en
73 https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/Publications/2023/Digital Development Country Profile Bosnia and Herzegovina [final- March 2023].pdf
74 https://www.publicmediaalliance.org/how-western-balkans-public-media-empowered-youth
75 Ibid.
76 https://arxiv.org/html/2506.09992v1
77 https://europeanwesternbalkans.com/2025/04/04/why-the-digital-services-act-is-needed-in-the-western-balkans-an-institutional-and-market-perspective/

## 8.3 Multi-Stakeholder Engagement and Capacity Building

- **Mandating Regular Dialogue Between Governments and Tech Companies:** The limited responsiveness of international tech companies in the Western Balkans, characterized by insufficient attention to localized moderation needs, necessitates structured engagement.[78] This recommendation establishes regular, formalized dialogues between BiH government entities (state and entity levels) and major tech companies. These dialogues, potentially facilitated by the Communications Regulatory Agency (CRA) BiH, focus on transparency regarding content moderation policies, resource allocation for B/S/C languages, and accountability for algorithmic harms. The objective is to foster a collaborative environment where tech companies share knowledge of AI system operations and governments articulate local needs, particularly concerning the "Linguistic Seam-Rip" and the "Digital Iron Curtain." This direct engagement mitigates the "Algorithmic Indifference" that often characterizes platform operations in smaller markets.

- **Enhancing the Capacity of the Communications Regulatory Agency (CRA) BiH:** The CRA BiH, while committed to digital transformation, operates without comprehensive legislative backing or sufficient resources to oversee AI deployment effectively.[79] This recommendation calls for substantial investment in the CRA BiH's technical expertise, human resources, and legislative mandates. This includes specialized training in AI governance, digital forensics, and media and information literacy for its staff. Empowering the CRA BiH with "regulatory sandboxes, living labs, and innovation hubs" allows for a practical, experimental approach to understanding and governing new technologies.[80] This enhancement transforms the CRA from a passive observer into an effective enforcer, capable of demanding platform accountability and ensuring the protection of youth from AI-driven harms. This step is critical for overcoming "Regulatory Recalcitrance."[81]

- **Strengthening Civil Society and Media Fact-Checking Capabilities:** Civil society organizations and independent media play an important role in monitoring online content and countering disinformation, particularly where state institutions face capacity limitations.[82] This recommendation strengthens their fact-checking capabilities through dedicated funding, advanced training in digital forensics, and access to AI-driven verification tools. Initiatives like BiH's "Why Not Organization" and "Citizens Against Terrorism Bosnia and Herzegovina" (CAT BiH) provide existing models for effective, youth-led counter-narrative strategies.[83] This bolsters independent oversight, provides credible counter-narratives to the "Ephemeral Nature of Truth," and strengthens the "Digital Immune System" by empowering non-state actors as crucial components of societal defense.

- **Promoting Public Awareness Campaigns on AI and Online Harms:** The widespread presence of AI-driven disinformation necessitates comprehensive public awareness campaigns. This recommendation advocates for regionally coordinated campaigns, potentially leveraging the expertise of the Regional Cooperation Council (RCC).[84] These campaigns educate citizens, especially youth, on AI's mechanisms of manipulation, the characteristics of synthetic media, and the dangers of algorithmic bias. Campaigns should utilize youth-friendly formats, such as social media content, interactive workshops, and educational videos, to increase "AI literacy" and foster critical digital engagement. This proactive approach cultivates a more discerning citizenry, reducing susceptibility to "cognitive capture" and strengthening the collective "Digital Immune System" against AI-driven harms.

78 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint arXiv:2506.09992, 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
79 Communications Regulatory Agency (CRA) Bosnia and Herzegovina, "GSR-25 Contribution: Best Practice Guidelines," presented at the GSR-25 Consultation, accessed https://www.itu.int/itu-d/meetings/gsr-25/wp-content/uploads/sites/33/2025/06/GSR-25_Contribution_Best-Practice-Guidelines_RAK-Bosnia-and-Herzegovina.pdf
80 Ibid.
81 Bojana Kostić and Caroline Sinders, "Responsible Artificial Intelligence: An overview of human rights' challenges of Artificial Intelligence and media literacy perspectives in the context of Bosnia and Herzegovina" (Council of Europe, June 2022), accessed via https://rm.coe.int/mil-study-3-artificial-intelligence-final-2759-3738-4198-2/1680a7cdd9, Chapter V.
82 https://www.unesco.org/en/articles/unesco-supports-launch-coalition-freedom-expression-and-content-moderation-bosnia-and-herzegovina
83 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 5. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf
84 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.5, p. 34.

- **Investing in Localized Linguistic Expertise for Content Moderation:** The "Linguistic Seam-Rip," where AI moderation inadequately detects nuanced hate speech in Bosnian, Serbian, and Croatian languages, demands targeted investment.[85] This recommendation proposes a "Localized AI Moderation and Language Data Initiative," driven by "The Acquis Digitalis" and international funding. It requires tech companies to invest significantly in developing AI moderation tools specifically trained on comprehensive, culturally nuanced datasets for B/S/C languages. This includes funding for dataset creation, training local linguists and AI specialists, and integrating human-in-the-process review for complex cases. This initiative directly addresses the persistent challenges of linguistic moderation, ensuring platforms are held accountable for the digital safety of all users, regardless of their linguistic context or market size. This foundational layer of protection complements youth digital literacy efforts, ensuring that even sophisticated AI-driven harms are effectively detected and mitigated.

# Strategic Roadmap: Phased Implementation and Resilience Building

## 9.1 Strategic Roadmap: Government (State and Entity Levels) Steps for BiH

Strategic roadmap for Bosnia and Herzegovina requires a phased governmental approach, engaging both state and entity levels. This framework addresses the nation's complex constitutional structure and the persistent "Regulatory Recalcitrance" that obstructs unified digital governance. These measures align BiH with international best practices and its EU aspirations, cultivating an environment conducive to a robust "Digital Immune System."

- **Phase 1 (Short-Term): Conduct Comprehensive Legal Gap Analysis:** A critical initial step involves a comprehensive legal gap analysis across all levels of government in BiH. This analysis identifies discrepancies between existing state and entity-level legislation and the evolving standards of EU digital governance, particularly those outlined in the Digital Services Act (DSA) and the forthcoming AI Act.[86]. This foundational analysis identifies legislative voids, particularly concerning AI's role in amplifying hate speech and disinformation, and highlights areas where existing laws are ineffective or contradictory due to BiH's constitutional fragmentation. The output is a clear, actionable report detailing specific legislative amendments required for full alignment with the *Acquis Digitalis*.

- **Phase 1 (Short-Term): Establish Inter-Entity Working Groups on Digital Governance:** Overcoming BiH's fragmented constitutional structure requires the immediate establishment of inter-entity working groups dedicated to digital governance. These groups, comprising experts from state-level institutions, the Federation of BiH, Republika Srpska, and Brčko District, foster continuous communication and shared understanding of AI-driven harms.[87]. Their mandate includes reviewing the legal gap analysis, identifying common ground for policy harmonization, and developing joint strategies against

85 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint arXiv:2506.09992, 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1

86 Namanja Sladaković and Milica Novaković, "Bosnia and Herzegovina," in IBA Alternative and New Law Business Structures Committee, July 2024, accessed https://www.ibanet.org/medias/anlbs-ai-working-group-report-july-2024-4-bosnia-herzegovina.pdf?context=bWFzdGVyfFB1YmxpY2F0aW9uUmVwb3J0c3w1NTQ3OHxhcHBsaWNhdGlvbi9wZGZ8YURJeUwyaGlaQzg1TVRNNE1qQTJOREE0TnpNMEwyRnViR0p6TFdGcExYZHZjbXRwYm1jdFozSnZkWEF0Y21Wd2IzSjBMV3AxYkhrdE1qQXlOQzAwTFdKdmMyNXBZUzFvWlhKNlpXZHZkbWx1WVM1d1pHWXw4M2UzMmYzYTRlOGEwZWJiYTk1ZTkyYThkN2MxNGI1NmU4ZDFhODUxOGY5ODQ0OTM1NWIzYTc1MmFlODkwZGZlfQ==

87 Communications Regulatory Agency (CRA) Bosnia and Herzegovina, "GSR-25 Contribution: Best Practice Guidelines," presented at the GSR-25 Consultation, accessed https://www.itu.int/itu-d/meetings/gsr-25/wp-content/uploads/sites/33/2025/06/GSR-25_Contribution_Best-Practice-Guidelines_RAK-Bosnia-and-Herzegovina.pdf

AI-driven disinformation. This collaborative mechanism directly counters the "Digital Double-Bind," where AI inadvertently reinforces political divisions, by creating a shared platform for national digital policy development.

- **Phase 2 (Medium-Term): Draft and Adopt DSA-Aligned Legislation:** Leveraging EU integration as a compelling driver, the medium-term strategy focuses on drafting and adopting legislation fully aligned with the Digital Services Act (DSA).[88] This involves translating the legal gap analysis findings into concrete legislative proposals, ensuring BiH's laws reflect the DSA's comprehensive framework for online safety, platform accountability, and digital rights protection. This step is crucial for addressing the "Moderation Desert" by compelling tech companies to apply robust content moderation standards, including those tailored to Bosnian, Serbian, and Croatian languages. The process requires close coordination between the inter-entity working groups and relevant parliamentary bodies, with technical assistance from EU experts, to expedite legislative adoption and overcome political impasses.

- **Phase 2 (Medium-Term): Allocate Budget for CRA Capacity Enhancement:** The Communications Regulatory Agency (CRA) in BiH currently lacks comprehensive legislative backing and sufficient resources for effective AI deployment oversight.[89] This step mandates a significant budget allocation for CRA capacity enhancement. This includes funding for specialized training in AI governance, digital forensics, and media and information literacy for CRA staff. Investment in "regulatory sandboxes, living labs, and innovation hubs" transforms the CRA into a proactive, effective oversight body capable of demanding transparency and enforcing accountability from tech companies.[90] This directly addresses "Regulatory Recalcitrance" by equipping the CRA with the necessary tools and expertise to mitigate AI-driven harms, thereby strengthening the "Digital Immune System" against pervasive disinformation.

- **Phase 3 (Long-Term): Implement a State AI Ethics Framework:** A long-term objective involves implementing a comprehensive State AI Ethics Framework. This framework, drawing from the principles of the EU AI Act, ensures responsible AI deployment across all sectors, particularly those impacting public discourse, electoral processes, and youth. It establishes guidelines for AI development and use, focusing on human oversight, data quality, transparency, and non-discrimination. This framework directly counters "Algorithmic Blind Spots" and "The Political Pilav Paradox" by proactively addressing how AI can inadvertently perpetuate systemic biases and exploit post-conflict trauma. It ensures that AI applications comply with ethical principles, the legal framework of public administration, and the free and democratic basic order.

## 9.2 Strategic Roadmap: Civil Society and Media Steps for BiH

Civil society organizations and media entities in Bosnia and Herzegovina (BiH) are vital in fostering digital resilience. Their independent oversight and public education efforts are essential for countering AI-driven harms, particularly when governmental and platform responses are insufficient. This strategic roadmap outlines key actions for these non-state actors, focusing on proactive engagement and advocacy to strengthen BiH's independent information ecosystem.

- **Phase 1 (Short-Term): Expand Fact-Checking Networks and Tools:** Strengthening independent verification against widespread disinformation requires the immediate expansion of fact-checking networks and the adoption of advanced tools. Organizations like BiH's "Why Not Organization" already address extremist and ethnonationalist narratives through rigorous fact-checking.[91] This involves increasing funding, providing training in digital forensics, and offering access to AI-driven verification software for existing and emerging youth-led fact-checking initiatives.[92] These networks, a core component of "The Digital Immune System," transform youth into proactive digital sentinels, directly countering the "Balkan Algorithmic Fracture" by enabling decentralized detection and response to algorithmically amplified divisive narratives.

88 Ibid.
89 Ibid.
90 Ibid.
91 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 5. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf
92 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en, Looking Ahead: Generative AI.

- **Phase 1 (Short-Term): Launch Public Awareness Campaigns on Deepfakes:** Addressing the "Ephemeral Nature of Truth" and the widespread nature of AI hallucinations demands immediate public awareness campaigns. These campaigns educate citizens, especially youth, on the characteristics of synthetic media, the mechanisms of AI manipulation, and the risks of algorithmic bias. Utilizing youth-friendly, interactive formats such as social media content, educational videos, and hands-on workshops maximizes reach and impact.[93] Campaigns should highlight examples of deepfakes in gender-based violence, as observed in Telegram networks across the Balkans,[94] to emphasize the tangible harms. This inoculation strategy, by familiarizing individuals with disinformation tactics, confers "cognitive immunity" when they encounter misinformation.[95]

- **Phase 2 (Medium-Term): Develop Specialized Training for Journalists on AI Disinformation:** Equipping media professionals to investigate and report on complex AI-driven harms is a medium-term priority. This includes developing specialized training programs for journalists across BiH, focusing on AI's role in disinformation, deepfake detection, algorithmic bias analysis, and the unique challenges of the "Linguistic Seam-Rip."[96] The "Reporting Diversity Network 2.0" by the Media Diversity Institute Western Balkans offers a model for promoting accurate and inclusive journalism, which can be expanded to incorporate AI-specific training.[97] Training should emphasize "lateral reading" and source corroboration over superficial indicators of trustworthiness.[98] This enhances the media's capacity to counter AI-amplified "Digital Memory Wars" and uphold journalistic integrity.

- **Phase 2 (Medium-Term): Advocate for Greater Platform Transparency:** Civil society organizations and media must increase advocacy for greater platform transparency. This presses tech companies to address the "Linguistic Seam-Rip" and improve moderation. It involves demanding granular data on platform investment in localized content moderation for B/S/C languages, reporting moderation failures, and transparently communicating algorithmic decision-making processes.[99] Leveraging the "Acquis Digitalis" framework, advocacy efforts compel tech companies to allocate resources equitably across language groups, transforming the "Moderation Desert" into a more regulated space. This directly challenges the cautious responsiveness of international tech companies and their "algorithmic indifference" to the unique linguistic and cultural nuances of the Western Balkans.

- **Phase 3 (Long-Term): Establish a Youth Digital Resilience Hub:** A long-term objective involves establishing a comprehensive Youth Digital Resilience Hub in BiH. This hub centralizes youth empowerment and counter-narrative initiatives, integrating elements of "Youth-Led Fact-Checking Networks," "Algorithmic Awareness and Critical Thinking Curriculum," and "Trauma-Informed Digital Storytelling for Reconciliation."[100] The hub fosters safe online spaces and mentorship programs, cultivating critical digital engagement. It empowers youth as digital rights advocates, enabling their participation in policy fora and platform accountability processes.[101] This initiative, a cornerstone of "The Digital Immune System," ensures youth are not merely protected from AI-driven harms but actively contribute to a resilient, democratic digital future.

93 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en, Case Study 2: Media Literacy Education, and Looking Ahead: Generative AI.
94 https://www.idea.int/news/ethical-conundrum-electoral-ai-3, "The ethical problem of electoral AI #3," International IDEA, April 02, 2025.
95 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en, Case Study 2: Media Literacy Education, How Effective Does It Seem?
96 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint arXiv:2506.09992, 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
97 Council of Europe, Programme of the World Forum for Democracy 2024: Democracy and Diversity, 6-8 November 2024, Strabourg, Page 16.
98 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en, Case Study 2: Media Literacy Education, How Effective Does It Seem?
99 Ibid., Case Study 2: Media Literacy Education, Key Takeaways.
100 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 5. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf
101 https://europeanwesternbalkans.com/2025/04/04/why-the-digital-services-act-is-needed-in-the-western-balkans-an-institutional-and-market-perspective/

## 9.3 Strategic Roadmap: Regional Coordination Mechanisms for Western Balkans

Fragmented national efforts prove insufficient against the continuous flow of harmful content that exploits shared linguistic and cultural spaces. The following recommendations foster collective resilience, creating a unified front against malign influence and preventing AI-driven harms from exploiting national borders.

- **Phase 1 (Short-Term): RCC Convenes Regular Digital Policy Dialogues:** The Regional Cooperation Council (RCC) establishes and convenes regular digital policy dialogues. These dialogues foster continuous communication and a shared understanding of AI-driven harms among policymakers, regulatory bodies, and cybersecurity experts from all Western Balkan countries.[102] This proactive engagement addresses the "Fractured Digital Expanse of the Western Balkans" by promoting a collective understanding of threats. It integrates national responses into a broader regional defense, ensuring the pervasive influence of AI meets a coordinated, rather than disparate, approach. The RCC's role is crucial in facilitating this exchange, transforming individual national concerns into a cohesive regional agenda.

- **Phase 1 (Short-Term): Initiate Cross-Border Training Programs for Regulators:** Building harmonized expertise to address the complex challenges of AI-driven harms requires initiating cross-border training programs for regulators across the Western Balkans. These programs focus on AI governance, digital forensics, and media and information literacy, ensuring a consistent level of technical and legal understanding across the region. The training addresses issues such as the "Linguistic Seam-Rip," equipping regulators to understand and respond to nuanced hate speech in Bosnian, Serbian, and Croatian languages.[103] This collaborative capacity building strengthens the "Digital Immune System" by enhancing institutional readiness, mitigating the "Moderation Desert" created by insufficient platform investment.

- **Phase 2 (Medium-Term): Develop a Common Regional Position on AI Governance:** The Western Balkans develops a common regional position on AI governance. This unified stance presents a cohesive front to tech companies and international bodies, increasing the region's leverage in demanding equitable resource allocation for content moderation and data protection.[104] This common position, informed by EU standards and the "Acquis Digitalis" as a primary paradigm, facilitates the harmonization of national legislations and regulatory practices. It directly counters the "Digital Double-Bind" by fostering unity in digital policy, ensuring AI does not inadvertently reinforce political fragmentation within the region. This strategic alignment is essential for countering cross-border disinformation campaigns and promoting collective digital security.

- **Phase 2 (Medium-Term): Establish a Regional AI Ethics Advisory Board:** A Regional AI Ethics Advisory Board provides independent expert guidance on responsible AI deployment across the Western Balkans. This board comprises academics, civil society representatives, technical experts, and youth advocates, ensuring a multi-stakeholder perspective. Its mandate includes advising on the ethical implications of AI systems, particularly concerning historical trauma, ethnic sensitivities, and fundamental rights. The board also assesses potential "Algorithmic Blind Spots" that could perpetuate discrimination or amplify divisive narratives. This proactive measure ensures AI development and deployment adhere to human-centric principles, building public trust and mitigating the "Political Pilav Paradox" by addressing systemic biases before they manifest digitally.

- **Phase 3 (Long-Term): Create a Western Balkans Digital Single Market Framework:** A long-term objective involves creating a Western Balkans Digital Single Market Framework. This framework, leveraging the EU "paradigm" for digital integration, fosters a secure and integrated digital economy across the region. It harmonizes regulations on digital services, data flows, and cybersecurity, aligning them with the EU's Digital Single Market. This initiative promotes economic growth and innovation and

102 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.3, p. 29.
103 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint , 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
104 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective", European Western Balkans, April 4, 2025.

strengthens collective digital security by establishing common standards for platform accountability and consumer protection. It ensures the ubiquitous cross-border nature of AI-driven disinformation meets a unified regulatory environment, fortifying the "Digital Immune System" against malign influence and facilitating the Western Balkans' path toward EU integration.

## 9.4 Strengthening Oversight and Accountability in AI System Deployment (36-40)

- **Mandate AI Impact Assessments for Public Sector Deployments:** Mandating AI Impact Assessments (AIAs) for all public sector AI deployments across the Western Balkans is essential for risk mitigation. This includes AI systems in public services, electoral processes, and information dissemination. The AIA identifies potential algorithmic biases, discrimination risks, and the amplification of divisive narratives, particularly those 'redolent' with historical trauma.[105] It assesses the impact on fundamental rights, including freedom of expression and non-discrimination, and proposes mitigation measures prior to deployment. This step incorporates human rights considerations into AI system design and operation, preventing inadvertent perpetuation of "The Political Pilav Paradox."

- **Develop AI Audit and Certification Standards:** Establishing robust AI audit and certification standards across the Western Balkans creates clear benchmarks for trustworthy AI. These standards align with principles from the EU AI Act, providing a framework for independent third-party auditing of AI systems, particularly those classified as high-risk.[106] Certification ensures AI systems meet specified criteria for data quality, accuracy, transparency, and accountability. This involves rigorous testing for algorithmic bias, especially concerning ethnic and religious sensitivities, and evaluating content moderation effectiveness for Bosnian, Serbian, and Croatian languages. Such standards counter the "obfuscation" of algorithmic impacts, offering verifiable assurance of responsible AI deployment.

- **Implement Algorithmic Transparency Requirements for Platforms:** Requiring tech companies to disclose their systems' internal workings is a necessary step toward accountability. This recommendation establishes stringent algorithmic transparency requirements for platforms operating in the Western Balkans. These requirements mandate public disclosure of how content recommendation engines prioritize information, how moderation algorithms detect and act on harmful content (especially for B/S/C languages), and the human oversight mechanisms in place.[107] This transparency enables researchers, civil society, and regulators to identify "Algorithmic Blind Spots" and diagnose the "Linguistic Seam-Rip," fostering informed public discourse and targeted interventions against AI-driven disinformation.

- **Establish Independent AI Ethics Review Boards:** Establishing independent AI Ethics Review Boards at national and regional levels provides impartial oversight and guidance. These boards, composed of multidisciplinary experts (including ethicists, legal scholars, technical specialists, and youth representatives), offer independent review and guidance on AI policy and specific deployments.[108] Their mandate includes assessing the societal impact of AI, particularly on youth and vulnerable groups, and advising on measures to prevent the 'perfidious' exploitation of post-conflict trauma and ethnic divisions. This oversight strengthens public trust in AI technologies, especially in a region where trust in institutions is often 'chary'.[109]

- **Create Mechanisms for Redress Against Algorithmic Harms:** Citizens require clear avenues to challenge and rectify 'perfidious' AI decisions. This recommendation establishes accessible mecha-

105 https://www.kdz.eu/system/files/downloads/2025-04/AI%20at%20local%20level.pdf, "Artificial Intelligence in Local Government: Driving Innovation, Bridging Gaps, and Shaping the Digital Transition in the Western Balkans and Moldova," KDZ – Centre for Public Administration Research, NALAS, April 2025, p. 7.
106 Ibid., p. 10
107 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint , 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
108 Communications Regulatory Agency (CRA) Bosnia and Herzegovina, "GSR-25 Contribution: Best Practice Guidelines," presented at the GSR-25 Consultation, accessed https://www.itu.int/itu-d/meetings/gsr-25/wp-content/uploads/sites/33/2025/06/GSR-25_Contribution_Best-Practice-Guidelines_RAK-Bosnia-and-Herzegovina.pdf
109 Ibid.

nisms for redress against algorithmic harms, including streamlined complaint processes, independent mediation services, and legal avenues for individuals to seek compensation for damages caused by biased or discriminatory AI systems. This is particularly important for youth, who may be disproportionately affected by AI-driven hate speech or disinformation but lack recourse. This ensures that the "Regulatory Recalcitrance" of existing legal frameworks does not leave citizens unprotected. It builds public trust and reinforces the principle that AI technologies must serve society responsibly, especially in contexts 'redolent' with historical distrust and systemic injustices.

## 9.5 Long-Term Measures for Societal Resilience (Youth Focus) (41-45)

- **Foster a Culture of Critical Digital Engagement from Early Education:** Long-term societal resilience in the Western Balkans necessitates integrating comprehensive digital and media literacy as a core competency from early education. This curriculum extends beyond basic online safety, emphasizing "algorithmic awareness" to instruct youth on how AI algorithms personalize content and how external actors exploit these algorithms for manipulation.[110] Key elements include lateral reading, inoculation against common manipulation tactics, provenance verification, and practical exercises in source triangulation.[111] This approach cultivates a discerning citizenry, reducing susceptibility to "cognitive capture" by AI-driven disinformation. It addresses the "Digital Iron Curtain" by empowering youth to critically deconstruct algorithmic influence, transforming them into active digital sentinels.

- **Support Research into AI's Impact on Youth Mental Health and Polarization:** The dynamic landscape of AI-driven disinformation and the "Balkan Algorithmic Fracture" require targeted research into AI's impact on youth mental health and polarization. This research provides an evidence base for interventions. Investigations should examine how algorithmic amplification of divisive narratives, particularly those evoking historical trauma and ethnic tensions, affects psychological well-being, fosters echo chambers, and contributes to radicalization.[112] This includes longitudinal studies on synthetic media exposure and the psychological resilience of youth in multi-ethnic societies. Such evidence informs the development of "Resilience Programs for Post-Conflict Trauma Amplified by AI," ensuring interventions are empirically grounded and culturally sensitive.

- **Promote Inter-Ethnic and Inter-Religious Dialogue in Digital Spaces:** Countering the amplification of divisions by AI algorithms requires promoting inter-ethnic and inter-religious dialogue in digital spaces. This measure cultivates "The Digital Youth Forum," a conceptual meeting ground for youth across the Western Balkans. The forum, facilitated by technology, engages youth in dialogue, critical analysis, and the co-creation of counter-narratives. It fosters inter-ethnic understanding and resilience against AI-driven harms. The forum supports the "Digital Diplomat," an AI-powered platform designed to identify and bridge informational divides. It recommends diverse perspectives and facilitates moderated, cross-ethnic conversations. This AI, explicitly trained on "pro-social" metrics, prioritizes content that promotes empathy, critical thinking, and mutual understanding, actively mending the "Balkan Algorithmic Fracture" and countering "Digital Memory Wars."

- **Develop Resilience Programs for Post-Conflict Trauma Amplified by AI:** The persistent impact of post-conflict trauma, exacerbated by AI-driven historical revisionism, necessitates specialized resilience programs. These programs address the psychological impacts of AI-amplified divisive narratives, particularly those targeting ethnic and religious sensitivities. They integrate psychosocial support with critical digital engagement, offering safe spaces and mentorship programs for youth to process trauma and develop coping mechanisms. Such initiatives, a core component of "The Digital Immune System," empower youth to critically analyze historical narratives encountered online, fostering healing rather than perpetuating "Digital Memory Wars." The "Trauma-Informed Digital Storytelling for Reconciliation" initiative is a key element, guiding youth to create narratives that promote empathy, inter-ethnic under-

110 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en, Case Study 2: Media Literacy Education, and Looking Ahead: Generative AI.
111 Ibid.
112 Conclusion Paper: Online radicalisation and P/CVE approaches in the Western Balkans, published by the Radicalisation Awareness Network (RAN), European Commission, based on a paper prepared by Nejra Veljan after consultation with Fenna Canters and Rositsa Dzhekova (RAN Staff). https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf (Page 4)

standing, and critical engagement with historical accounts.[113]

- **Cultivate Digital Storytelling for Peacebuilding Among Youth:** Empowering youth to create positive counter-narratives is a crucial long-term measure for fostering lasting social cohesion. This involves cultivating digital storytelling for peacebuilding, an initiative that equips youth with the skills to produce short videos, podcasts, and interactive web experiences that promote reconciliation and inter-ethnic understanding. These programs are trauma-informed, acknowledging the persistent impact of past conflicts. They guide youth to construct narratives that foster empathy and healing, rather than perpetuating historical revisionism or nationalist propaganda.[114] This effort builds a powerful reservoir of positive, unifying content that directly counters the "Digital Memory Wars" fueled by the "Balkan Algorithmic Fracture." By transforming youth into "digital peacebuilders," this measure fortifies "The Digital Immune System" with proactive, empathetic content, ensuring that the dynamic landscape of AI-driven disinformation is met with enduring narratives of unity.

## 9.6 Sustainable Funding and International Support Mechanisms

Ambitious strategic roadmap for the Western Balkans, which encompasses legal reforms, institutional strengthening, and societal resilience initiatives, requires sustained financial backing and robust international collaboration. Without a framework for sustainable funding, these efforts risk remaining mere aspirations rather than demonstrable, on-the-ground impacts. This sub-section identifies five mechanisms for securing the necessary resources to translate the strategic vision into reality, ensuring the long-term viability of digital resilience efforts.

These funding and support mechanisms are essential for translating the strategic roadmap into demonstrable, on-the-ground impact, ensuring the long-term viability of digital resilience efforts and providing a comprehensive vision for implementation that leads directly into the Conclusion.

Timeline bar chart visually represents the phased approach to implementation, underscoring the continuous commitment required from domestic and international partners. This ongoing investment is crucial for building a resilient digital ecosystem in the Western Balkans.

## 9.7 Synthesizing Findings and Prioritizing Action

Impact of AI on the amplification of hate speech, extremism, and disinformation in Bosnia and Herzegovina and the wider Western Balkans requires a concise synthesis of findings and a ranked set of actions. The report documents how AI-driven processes — manifesting as what we term the "Balkan Algorithmic Fracture" and the "Ephemeral Nature of Truth" created by synthetic media — exploit existing ethnic divisions and historical grievances.[115] The analysis identifies youth as a high-risk group: they face disproportionate exposure online and will determine the region's capacity for long-term resilience.

Responding effectively requires a multi-pronged strategy. First, harmonize state- and entity-level legislation with European standards, notably the Digital Services Act (DSA), to reduce legal fragmentation described in Section 3.1.[116] Second, establish a coordinated BiH Digital Governance Body to align policy across institutions and to set digital-specific standards for countering online hate. Third, strengthen legal definitions and enforcement mechanisms for online hate speech so that remedies address platform-mediated harms.

This analysis also documents gaps in platform responses and moderation capacity. International technology firms show limited responsiveness to the region's linguistic and contextual particularities, producing a

113 Council of Europe, Programme of the World Forum for Democracy 2024: Democracy and Diversity, 6-8 November 2024, Strasbourg, Lab 1, p. 14.
114 Ibid.
115 The 'perfidious' nature of AI-driven manipulation lies in its capacity to mimic authenticity while sowing deep-seated distrust.
116 "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective", European Western Balkans, April 4, 2025.

moderation gap for Bosnian/Serbian/Croatian (B/S/C) content.[117] Policy and operational responses must close that gap through targeted vendor engagement, procurement of localized moderation capacity, and investment in language-specific detection tools.

Priorities, sequenced by timeframe, are as follows.

Immediate (0–12 months)

- Harmonize legislation and adopt digital-specific amendments to hate speech and media laws.
- Draft a BiH AI Act that adapts key principles from the EU AI Act to the national and entity contexts.

Strengthen the Communications Regulatory Agency (CRA) BiH with technical teams for platform oversight and digital investigations.[118]

Medium term (12–36 months)

- Launch a Western Balkans Digital Cooperation Forum to coordinate cross-border policy, share threat intelligence, and align standards.[119]
- Develop a Regional Early Warning System for AI-driven disinformation to support rapid detection and coordinated responses.
- Scale youth-focused programs: youth-led fact-checking networks, media and information literacy modules, and an algorithmic awareness curriculum integrated in schools and civil-society training.[120]

Long term (36+ months)

- Secure sustainable funding through EU Pre-Accession Instruments (IPA III) and targeted international donor programs to support resilience and capacity building.[121]
- Foster public-private partnerships for technical solutions, including language models and moderation tools tailored to local languages and dialects.
- Institutionalize workforce development for digital governance, platform oversight, and civil-society monitoring.

Strategic use of EU integration levers offers a practical route to system-level reform: adopting the DSA and AI Act provisions will align market incentives, increase platform accountability, and provide technical benchmarks for enforcement. Complementing legal change with operational improvements - localized moderation capacity, CRA upgrades, regional threat-sharing, and youth resilience programs - converts regulation into measurable outcomes.

117 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint arXiv:2506.09992, 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1

118 Namanja Sladaković and Milica Novaković, "Bosnia and Herzegovina," in IBA Alternative and New Law Business Structures Committee, July 2024, accessed https://www.ibanet.org/medias/anlbs-ai-working-group-report-july-2024-4-bosnia-herzegovina.pdf?context=bWFzdGVyfFB1YmxpY-2F0aW9uUmVwb3J0c3w1NTQ3OHxhcHBsaWNhdGlvbi9wZGZ8YURJeUwyaGlaQzg1TVRNNE1qQTJOREE0TnpNMEwyRn-ViR0p6TFdGcExYZHZjbXRwYm1jdFozSnZkWEF0Y21Wd2lzSjBMV3AxYkhrdE1qQXlOQzAwTFdKdmMyNXBZUzFvWlhKNlpX-ZHZkbWx1WVM1d1pHWXw4M2UzMmYzYTRlOGEwZWJiYTk1ZTkyYThkN2MxNGI1NmU4ZDFhODUxOGY5ODQ0OTM1NWIz-YTc1MmFlODkwZGZlfQ==

119 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.3, p. 29.

120 Radicalisation Awareness Network (RAN). (2022). Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper. European Commission. Page 5. https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf

121 "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024," Regional Cooperation Council, December 2024, Section 2.3.5, p. 34.

## 9.8 Youth, Cooperation, and the 'Ineluctable' Path Forward

Youth, despite exposure to AI-driven manipulation, are not mere recipients of digital harms. They serve as agents for societal resilience, capable of forming a "Digital Immune System" through deliberate empowerment. This approach recognizes that young people, deeply integrated into the "ubiquitous" digital landscape, can actively identify, counter, and establish defenses against AI-driven disinformation and extremism.[122] Their active participation is an "ineluctable" component of a durable democratic future.

"Digital Immune System" functions through interconnected youth networks, peer education, and direct engagement with policy-making. Initiatives such as "Youth-Led Fact-Checking Networks" enable active monitoring of online content, identification of AI-driven disinformation, and production of verified counter-narratives in local languages.[123] These networks directly challenge the "Digital Iron Curtain" by fostering critical assessment of algorithmically curated information. An "Algorithmic Awareness and Critical Thinking Curriculum" educates youth on AI algorithm function, content personalization, and manipulation tactics, thereby strengthening resilience against "cognitive capture."[124]

"Linguistic Seam-Rip" presents a notable vulnerability, as AI moderation systems frequently fail to detect nuanced hate speech in Bosnian, Serbian, and Croatian languages.[125] This deficiency necessitates a "Localized AI Moderation and Language Data Initiative," driven by "The Acquis Digitalis." This initiative mandates tech companies invest in comprehensive, culturally nuanced datasets and human linguistic expertise, directly addressing the moderation gap that leaves youth vulnerable to subtle, culturally embedded hate speech.

Robust regional cooperation and increased international support are "ineluctable" to counter the "profligate" digital tide. Collective action, sensitive to linguistic subtleties and historical trauma, constitutes the sole viable bulwark. The "Regional Digital Diplomacy and Counter-Narrative Hub" functions as a collaborative platform, uniting youth activists, media professionals, and digital experts across the Western Balkans. This hub develops and disseminates positive, unifying counter-narratives that directly challenge AI-amplified hate speech and extremism, especially content "redolent" with historical revisionism.[126]

"Trauma-Informed Digital Storytelling for Reconciliation" initiatives empower youth to create digital stories that promote empathy and inter-ethnic understanding, directly countering the "Digital Memory Wars" fueled by algorithmic bias.[127] These programs, mindful of past conflicts' persistent impact, guide youth to construct narratives that foster healing. The overarching framework connects to a vision for a fortified Western Balkans, where democratic processes are resilient, and citizens, particularly youth, are equipped to navigate the "ubiquitous" digital landscape securely. The unwavering commitment to EU integration provides the "ineluctable" leverage to drive these reforms, ensuring the region transitions from a vulnerable frontier to a model of digital resilience.

122 Conclusion Paper: Online radicalisation and P/CVE approaches in the Western Balkans, published by the Radicalisation Awareness Network (RAN), European Commission, based on a paper prepared by Nejra Veljan after consultation with Fenna Canters and Rositsa Dzhekova (RAN Staff).
https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf (Page 3)
123 Ibid., Page 5.
124 https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en, Case Study 2: Media Literacy Education, and Looking Ahead: Generative AI.
125 Amel Muminovic and Amela Kadric Muminovic, "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages," arXiv preprint arXiv:2506.09992, 2025, Section I-A, Paragraph 4. Available at: https://arxiv.org/html/2506.09992v1
126 Council of Europe, Programme of the World Forum for Democracy 2024: Democracy and Diversity, 6-8 November 2024, Strasbourg, Page 16.
127 Ibid., Lab 1, p. 14.

# Appendix A: Glossary of Terms

## A.1 Definition of Key Terms

- **Acquis Digitalis:** This concept asserts that the EU accession process provides inherent external leverage for comprehensive digital governance reform across the Western Balkans. It advocates for the strategic adoption and harmonization with the Digital Services Act (DSA) and the forthcoming AI Act, viewing them as critical tools for national security and democratic consolidation.

- **Algorithmic Amplification:** The mechanism by which AI systems, through content recommendation engines and personalization, increase the visibility and reach of specific content. This often inadvertently promotes divisive or harmful narratives.

- **Algorithmic Bias:** Systemic and recurring errors in an AI system's output resulting from flawed assumptions in the machine learning process. This occurs when training data misrepresents the population or contains embedded societal prejudices, leading to skewed or unfair outcomes.

- **AI Hallucinations:** Inaccurate information generated by AI chatbots or large language models, presented as factual despite being fabricated. These impact public trust and democratic processes.

- **Balkan Algorithmic Fracture:** This concept describes how AI's content recommendation systems and personalization algorithms exploit and intensify existing ethnic, religious, and political divisions within BiH and the Western Balkans. Algorithms, optimizing for engagement, create tendentious echo chambers, reinforcing pre-existing biases and historical grievances.

- **Code-Switching:** The practice of fluent speakers alternating between two or more languages or dialects within a single conversation or text. This alters the context and intent of seemingly innocuous phrases, posing challenges for AI moderation.

- **Content Recommendation Engines:** AI systems that suggest content to users based on their past behavior, preferences, and interactions. These engines prioritize engagement, which can create "filter bubbles" by reinforcing existing beliefs.

- **Deepfakes:** Media content, including images, audio, or video, generated or manipulated using AI to create realistic but fabricated representations. This includes altering faces in videos, generating realistic voice clones, or creating entirely artificial events.

- **Digital Services Act (DSA):** A comprehensive EU regulation establishing a legal framework for online platforms and digital services. It creates a safer digital space, protects users' fundamental rights, and enforces platform accountability.

- **Digital Divide of Governance:** Challenges arising from Bosnia and Herzegovina's fragmented legal and institutional structures, resulting in disparities and inefficiencies in developing and implementing digital governance policies, especially concerning AI.

- **Digital Immune System:** This concept positions youth not merely as recipients of digital literacy but as active, self-sustaining components of societal defense against AI-driven harms. Empowering them with advanced media literacy and critical thinking skills transforms them into frontline defend-

ers against AI-driven manipulation, capable of generating counter-narratives and fostering collective resilience.

- **Digital Iron Curtain:** This concept highlights the acute vulnerability of youth, the primary demographic exposed to AI-driven harms. The pervasive presence of social media, combined with low media literacy, creates a "digital iron curtain" that limits exposure to diverse perspectives and increases susceptibility to cognitive capture by extremist narratives.

- **Disinformation:** False, inaccurate, or misleading information intentionally designed, presented, and promoted to cause public harm or generate profit.

- **Entity-Level:** Refers to the administrative divisions within Bosnia and Herzegovina (e.g., Republika Srpska, Federation of BiH, Brčko District) and their respective governmental and legal competencies.

- **EU AI Act:** A pioneering EU regulation ensuring the safe use of AI systems and establishing a clear legal framework. It employs a risk-based approach, introducing proportionate, effective, and binding rules for AI systems.

- **Extremism:** Advocacy or practice of extreme political or religious views, often characterized by intolerance towards opposing views and the use of violence or illegal means to achieve goals.

- **Hate Speech:** Publicly inciting hatred, discrimination, or violence against an individual or group based on characteristics such as origin, religion, or ethnicity.

- **Hybrid Threats:** Sophisticated campaigns combining conventional and unconventional tactics, such as cyberattacks, disinformation campaigns, and economic coercion, to destabilize governments and exploit societal divisions.

- **Linguistic Seam-Rip:** A critical enforcement gap in AI content moderation capabilities for Bosnian/Serbian/Croatian (B/S/C) languages. This results from linguistic complexities, cultural nuances, and their status as "low-resource languages," allowing AI systems to fail in detecting sophisticated hate speech and disinformation.

- **Low-Resource Languages:** Languages for which there is a scarcity of digital data (e.g., text corpora, labeled datasets) available for training Artificial Intelligence models, thus hindering the effectiveness of automated language processing tools.

- **Natural Language Processing (NLP) for Moderation:** The branch of AI that enables computers to understand, interpret, and generate human language. In content moderation, NLP models analyze text for harmful keywords, sentiment, and context but often struggle with the nuances of low-resource languages, slang, or culturally specific derogatory terms.

- **Platform Accountability:** The legal and ethical responsibility of online platforms (e.g., social media companies) to address harmful content, ensure transparency in their operations, and mitigate systemic risks on their services.

- **Post-Conflict Society:** A society undergoing recovery and reconstruction after a period of armed conflict. Such societies often contend with unresolved historical grievances, ethnic divisions, and fragile democratic institutions, making them particularly vulnerable to information manipulation.

## References

**Atlantic Council.** "How the US and Europe Can Counter Russian Information Manipulation About Nonproliferation." *Atlantic Council*, October 4, 2024.

https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/how-the-us-and-europe-can-counter-russian-information-manipulation-about-nonproliferation/.

**Atlantic Council, in collaboration with the US Department of State's Office of Cooperative Threat Reduction.** "Issue Brief: How the US and Europe Can Counter Russian Information Manipulation About Nonproliferation." *Atlantic Council*, October 4, 2024.

https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/how-the-us-and-europe-can-counter-russian-information-manipulation-about-nonproliferation/.

**Bradshaw, Samantha, and Philip N. Howard.** "The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation." *Working Paper 2019.2*. Oxford, UK: Project on Computational Propaganda, Oxford Internet Institute, University of Oxford, 2019.

https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2019/09/CyberTroop-Report19.pdf.

**Carnegie Endowment for International Peace.** "Countering Disinformation Effectively: An Evidence-Based Policy Guide." January 31, 2024.

https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en.

**Communications Regulatory Agency (CRA) Bosnia and Herzegovina.** "GSR-25 Contribution: Best Practice Guidelines." Presented at the GSR-25 Consultation. Accessed [Date of Access].

https://www.itu.int/itu-d/meetings/gsr-25/wp-content/uploads/sites/33/2025/06/GSR-25_Contribution_Best-Practice-Guidelines_RAK-Bosnia-and-Herzegovina.pdf.

**Council of Europe.** "Programme of the World Forum for Democracy 2024: Democracy and Diversity." Strasbourg, 6-8 November 2024.

https://rm.coe.int/programme-of-the-world-forum-for-democracy-2024-democracy-and-diversit/1680b11b3a.

**European Commission.** "Horizon Europe Work Programme 2023-2025 - 5. Culture, Creativity and Inclusive Society." Decision C(2024) 2371 of 17 April 2024.

https://research-and-innovation.ec.europa.eu/document/download/6e3460b4-dd63-4555-8acb-3f999f638624_en.

**European External Action Service (EEAS).** "3rd EEAS Report on Foreign Information Manipulation and Interference Threats: Exposing the architecture of FIMI operations." March 2025.

https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3nd-ThreatReport-March-2025-05-Digital-HD.pdf.

**European Parliament.** "Resolution of 1 June 2023 on foreign interference in all democratic processes in the European Union, including disinformation (2022/2075(INI))." P9_TA(2023)0219.

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0219_EN.pdf.

**European Western Balkans.** "Why the Digital Services Act is needed in the Western Balkans: An institutional and market perspective." April 4, 2025.

https://europeanwesternbalkans.com/2025/04/04/why-the-digital-services-act-is-needed-in-the-western-balkans-an-institutional-and-market-perspective/.

**International Bar Association (IBA).** "Bosnia and Herzegovina." In *IBA Alternative and New Law Business Structures Committee*, July 2024.

https://www.ibanet.org/medias/anlbs-ai-working-group-report-july-2024-4-bosnia-herzegovina.pdf?context=bWFzdGVyfFB1YmxpY2F0aW9uUmVwb3J0c3w1NTQ3OHxhcHBsaWNhdGlvbi9wZGZ8YURJeUwyaGlaQzg1TVRNNE1qQTJOREE0TnpNMEwyRnViR0p6TFdGcExYZHZjbXRwYm1jdFozSnZkWEF0Y21Wd2IzSjBMV3AxYkhrdE1qQXlOQzAwTFdKdmMyNXBZUzFvWlhKNlpXZHZkbWx1WVM1d1pHWXw4M2UzMmYzYTRlOGEwZWJiYTk1ZTkyYThkN2MxNGI1NmU4ZDFhODUxOGY5ODQ0OTM1NWIzYTc1MmFlODkwZGZlfQ==.

**International IDEA.** "The ethical questionof electoral AI #3." April 02, 2025.

https://www.idea.int/news/ethical-conundrum-electoral-ai-3.

**International Telecommunication Union (ITU) Office for Europe.** "Digital Development Country Profile Bosnia and Herzegovina" (final - March 2023).

https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Documents/Publications/2023/Digital%20Development%20Country%20Profile%20Bosnia%20and%20Herzegovina%20%5Bfinal-%20March%202023%5D.pdf.

**International Telecommunication Union (ITU) Europe Region.** "ITU Europe Region Initiatives 2023-2025."

https://www.itu.int/en/ITU-D/Regional-Presence/Europe/Pages/Regional%20Initiatives/2022/ITU-Europe-Region-Initiatives-2023-2025.aspx.

**Kostić, Bojana, and Caroline Sinders.** "Responsible Artificial Intelligence: An overview of human rights' challenges of Artificial Intelligence and media literacy perspectives in the context of Bosnia and Herzegovina." *Council of Europe*, June 2022. Accessed via

https://rm.coe.int/mil-study-3-artificial-intelligence-final-2759-3738-4198-2/1680a7cdd9.

**KDZ – Centre for Public Administration Research, and NALAS (Network of Associations of Local Authorities of South-East Europe).** "Artificial Intelligence in Local Government: Driving Innovation, Bridging Gaps, and Shaping the Digital Transition in the Western Balkans and Moldova." April 2025.

https://www.kdz.eu/system/files/downloads/2025-04/AI%20at%20local%20level.pdf.

**Muminovic, Amel, and Amela Kadric Muminovic.** "Large Language Models for Toxic Language Detection in Low-Resource Balkan Languages." *arXiv preprint arXiv:2506.09992*, 2025. Available at:

https://arxiv.org/html/2506.09992v1.

**Open Society Institute – Sofia.** "Media Literacy Index 2021." March 14, 2021.

https://osis.bg/?p=3750&lang=en.

**Public Media Alliance.** "How Western Balkans public media empowered youth to tell fake from fact." April 2, 2025.

https://www.publicmediaalliance.org/how-western-balkans-public-media-empowered-youth/.

**Radicalisation Awareness Network (RAN), European Commission.** "Online radicalisation and P/CVE approaches in the Western Balkans: Conclusion Paper." June 16, 2022.

https://home-affairs.ec.europa.eu/system/files/2023-05/ran_paper_online_radicalisation_p-cve_approaches_in_wb_16062022_en.pdf.

**Regional Cooperation Council (RCC).** "Common Regional Market Action Plan 2025-2028 (CRM 2.0)."

https://www.rcc.int/files/user/docs/0e5e72bb8334509a1feb954cdc7a3e54.pdf.

**Regional Cooperation Council (RCC) Secretariat.** "DRAFT REPORT ON THE ACTIVITIES OF THE REGIONAL COOPERATION COUNCIL SECRETARIAT For the period 01 October 2024 – 28 February 2025."

https://www.rcc.int/files/user/docs/a9b1d91ff97948db2d1e8adf2d6a2662.pdf.

**Regional Cooperation Council (RCC).** "From Dialogue to Action: Working Group Outcomes and Recommendations from the 9th Regional Security Coordination Conference 2024." December 2024.

https://www.rcc.int/download/docs/WWG-Outcomes-and-Recommendations-from-the-9th-Regional-Security-Coordination-Conference-2024-2.pdf/3121b2bcbc36c0dd085d11d1826b961c.pdf.

**Safer Internet Center Bosnia and Herzegovina.** "Sigurnodijete.ba - Home Page." Accessed June 12, 2024.

https://www.sigurnodijete.ba/en/.

**UNESCO.** "UNESCO Supports Launch of Coalition for Freedom of Expression and Content Moderation in Bosnia and Herzegovina." June 20, 2023.

https://www.unesco.org/en/articles/unesco-supports-launch-coalition-freedom-expression-and-content-moderation-bosnia-and-herzegovina.

**World Economic Forum.** "The Global Risks Report 2024." January 2024.

https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.

**World Economic Forum.** "The Global Risks Report 2025." January 2025.

https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf.

**Other Useful Links and Research Guides**

**Council of Europe Resources on AI and Human Rights:** Explore the Council of Europe's work on artificial intelligence and its impact on human rights, democracy, and the rule of law for a broader understanding of the European regulatory context.

https://www.coe.int/en/web/artificial-intelligence.

**EU Digital Services Act (DSA) Information:** Access official EU documentation and explanatory materials on the Digital Services Act to understand its scope, obligations, and enforcement mechanisms.

https://digital-strategy.ec.europa.eu/en/policies/digital-services-act.

**EU AI Act Information:** Consult official EU resources detailing the Artificial Intelligence Act, its risk-based approach, and its implications for AI deployment across member states and potential candidate countries.

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689.

**Open Society Foundations - Media Literacy Initiatives:** Explore the work of the Open Society Foundations on media literacy, disinformation, and digital rights, which often includes regional analyses and reports relevant to the Western Balkans.

https://www.opensocietyfoundations.org/policy-and-advocacy/media-and-internet.

**Regional Cooperation Council (RCC) Digital Agenda:** Review the RCC's ongoing initiatives and strategy documents concerning the Digital Agenda for the Western Balkans to understand regional cooperation efforts and alignment with EU digital policies.

https://www.rcc.int/priority_areas/65/digital-transformation.

**ITU Child Online Protection (COP) Guidelines:** Access the International Telecommunication Union's global guidelines and resources for child online protection, which provide a framework for national adaptation and implementation.

https://www.itu-cop-guidelines.com/.

**Global Risks Report (World Economic Forum):** Consult the World Economic Forum's annual Global Risks Report for broad insights into emerging societal, technological, and geopolitical risks, including those related to AI and disinformation.

https://www.weforum.org/reports/the-global-risks-report-2025.

**arXiv.org - NLP and AI Research:** Explore arXiv for the latest pre-print research in Natural Language Processing (NLP) and AI, particularly studies focusing on low-resource languages and toxic language detection, such as the work by Muminovic and Kadric Muminovic.

https://arxiv.org/list/cs.CL/recent.

**Atlantic Council - Digital Forensic Research Lab (DFRLab):** Review DFRLab's analyses for insights into disinformation campaigns, manipulated media, and the operational tactics of state and non-state actors in the digital space.

https://www.atlanticcouncil.org/programs/digital-forensic-research-lab/.

This research was conducted as part of the ‘AIMindItWithCare’ project, with financial support from the Government of the Kingdom of the Netherlands through the Matra program. The views and conclusions expressed are solely those of the JaBiHEU association and the author of the research and do not necessarily reflect the views of the donor.