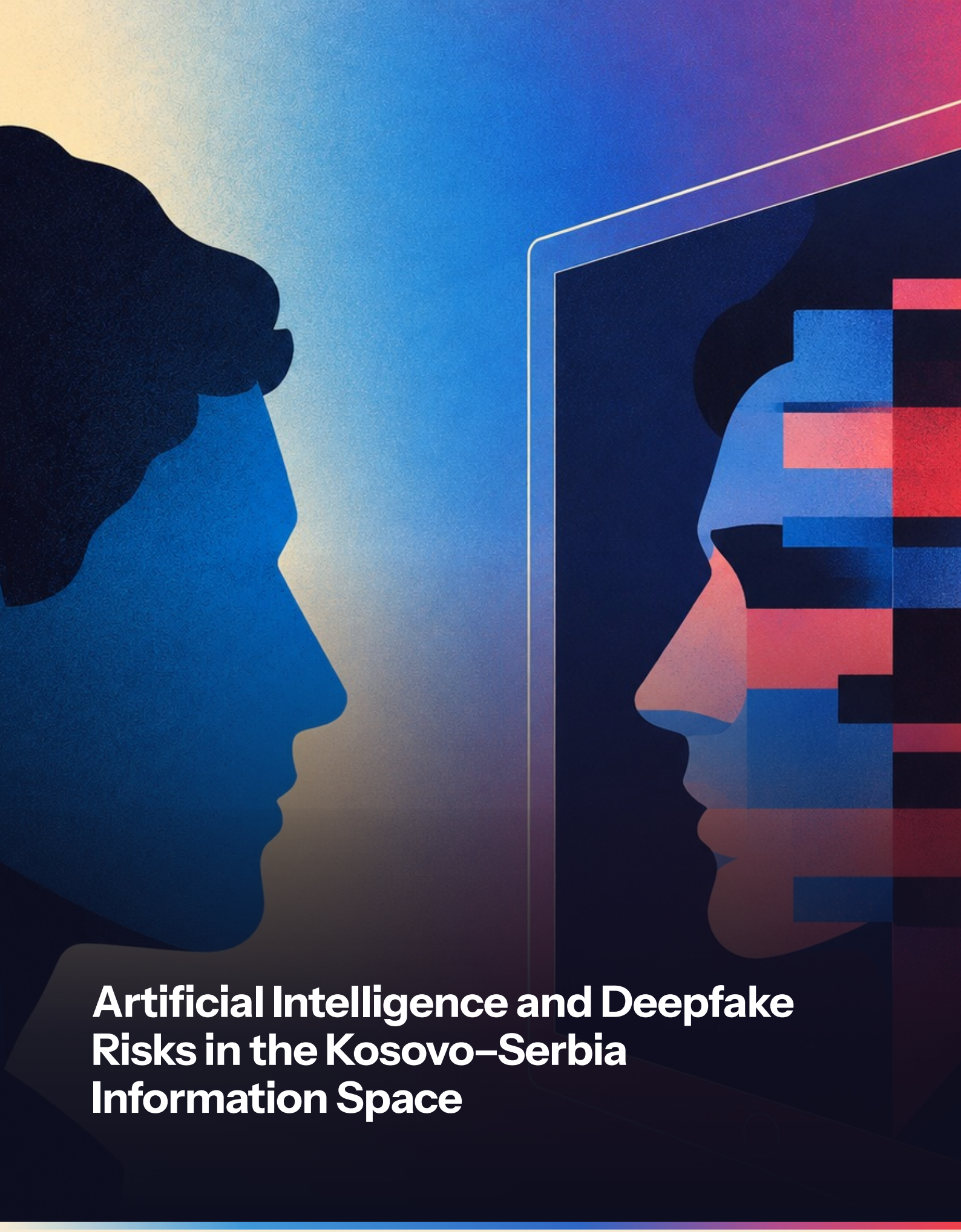


Artificial Intelligence and Deepfake Risks in the Kosovo–Serbia Information Space

NEW
SOCIAL
INITIATIVE

Artificial Intelligence and Deepfake Risks in the Kosovo–Serbia Information Space

PUBLISHED BY:
New Social Initiative - NSI
www.newsocialinitiative.org

AUTHOR:
Yllka Bajoku

MENTOR:
Ramadan Ilazi, Kosovar Centre for Security Studies (KCSS)

October 2026, North Mitrovica

This publication was produced by the New Social Initiative with support from the National Endowment for Democracy. The views expressed are those of NSI alone and do not necessarily reflect those of NED.

Contents

Executive Summary 04

01/ Introduction 05

02/ Problem Context 06

— Banjska case – manipulated audio
— AI deepfake of Donald Trump – Kosovo and Serbia
— Kurti and Hovenier – manipulated audio
— AI-generated video portraying Serbian youth supporting Kosovo's recognition

03/ Impact of Deepfakes and AI 09

04/ Conclusion and Recommendations 11

References 13

Executive Summary

This paper examines the growing role of artificial intelligence (AI)-generated deepfakes in shaping the information space between Kosovo and Serbia. It focuses on how manipulated audio, video and images can be used to produce or amplify disinformation, influence public perceptions, increase political and interethnic tensions, and undermine trust in the media and public institutions. These risks are particularly significant in the Kosovo–Serbia context, where political narratives, ethnic divisions and unresolved historical disputes make the information environment especially vulnerable to manipulation.

The paper analyses selected cases of AI-generated or manipulated content in order to identify the main objectives and patterns behind their use, including the potential to create interethnic tensions, reinforce existing political narratives, discredit institutions or public figures, and influence public opinion during periods of heightened political sensitivity. It also assesses the preparedness of governments, media organisations, civil society and educational institutions to identify and respond to such threats.

The paper argues that reducing the impact of deepfake-driven disinformation requires a combination of stronger digital and media literacy, improved fact-checking and verification mechanisms, appropriate legal and regulatory responses, and greater cooperation among institutions, the media, civil society and international partners. Ultimately, it provides evidence-based recommendations aimed at strengthening democratic and societal resilience against AI-enabled information manipulation in Kosovo and Serbia.

01/
Introduction

The rapid development of artificial intelligence (AI) is transforming the way information is created, shared and consumed. While these technologies offer significant opportunities, they also create new risks for the integrity of the information environment and for democratic processes (Council of Europe, 2026).

One of the most significant challenges is the growing use of AI-generated and manipulated content, including deepfakes, which can imitate a person's voice, appearance or actions and make fabricated material appear authentic (UNESCO, 2025). As synthetic media become more sophisticated and accessible, distinguishing between authentic and manipulated information can become increasingly difficult for citizens (UNESCO, 2025).

The risks associated with AI-generated disinformation are particularly serious in politically sensitive and polarised environments, where manipulated content can be used to influence public perceptions, reinforce existing narratives and contribute to confusion or mistrust (Council of Europe, 2026). Deepfakes can also pose risks to democratic institutions and public stability, particularly when fabricated content involves political figures or circulates during periods of political tension or crisis (Council of Europe, 2026).

In the context of Kosovo and Serbia, these risks require particular attention due to the sensitive political environment, existing ethnic divisions and the continued influence of competing political narratives.

AI-generated or manipulated content can therefore have consequences beyond ordinary misinformation, potentially contributing to polarisation and undermining trust in institutions and reliable sources of information (NDI, 2024; Ilazi et al., 2025).

02/
Problem Context

The growing accessibility and sophistication of artificial intelligence have made it easier to create and distribute manipulated audio, video and images that can appear authentic to audiences. Synthetic media, including deepfakes, can convincingly imitate a person's voice or likeness, making it increasingly difficult to distinguish between authentic and fabricated content (Naffi, 2025).

This development presents particular risks when manipulated content is used in politically sensitive environments. Deepfakes involving political leaders or public officials can be used to impersonate them, manipulate public perceptions and threaten democratic processes, making authenticity and transparency increasingly important (Parliamentary Assembly of the Council of Europe [PACE], 2026). The problem becomes even more significant when such content circulates during periods of political tension, when audiences may already be exposed to competing narratives and high levels of mistrust (National Democratic Institute [NDI], 2024).

The Kosovo–Serbia information space is particularly vulnerable to these risks. Political tensions, ethnic divisions and competing narratives can create conditions in which false or manipulated content is used to influence perceptions and intensify existing divisions. In such an environment, disinformation can undermine trust between communities and institutions and complicate efforts to normalise relations between Kosovo and Serbia (Ilazi et al., 2025).

Tensions between Kosovo and Serbia during 2023 – including the elections boycotted by Kosovo Serbs in April, the violent clashes between protesters and Kosovo Police and KFOR peacekeepers in May, the detention of three Kosovo Police officers by the Serbian police in June, and the armed attack against the Kosovo Police in Banjska in September – were key events during which the information space in Kosovo was flooded with information disorders (NDI, 2024).

Banjska case – manipulated audio

Following the 2023 Banjska attack, an audio recording was circulated claiming that Kosovo Police officers had refused to help a wounded Serbian attacker, allegedly saying "even if he dies, nothing." Fact-checkers found that the recording did not contain the claimed statement, making it an example of how manipulated or misleading audio can be used to spread disinformation during a crisis (Hoti, 2023).

AI deepfake of Donald Trump – Kosovo and Serbia

In 2025, an AI-generated video circulated on social media showing Donald Trump apparently speaking about tensions between Kosovo and Serbia. In the video, Trump appeared to make threatening statements towards Serbia, creating the impression that the United States would take military action if Serbia attacked Kosovo. The video was not a genuine statement by Trump: his appearance and voice had been manipulated using AI to make it seem as though he had made statements that he had never actually made. The case is particularly significant because it involved a highly sensitive political and security issue in the Western Balkans. A fabricated statement from a major political figure such as the US president can easily influence public opinion, increase political tensions and create confusion about the actual position of the United States (Sbunker, 2025).

Kurti and Hovenier – manipulated audio

The escalation of tensions in northern Kosovo in 2023 was also accompanied by the emergence of AI-enabled manipulation in the information space. A notable example occurred in June 2023, following the detention of three Kosovo Police officers by Serbian authorities and amid heightened tensions in northern Kosovo. A deepfake audio recording was circulated online, falsely presenting a conversation between Kosovo Prime Minister Albin Kurti and US Ambassador to Kosovo Jeffrey Hovenier. The fabricated recording portrayed the two officials as discussing the arrest of Kosovo Serbs in the north and included statements suggesting that the Kosovo authorities intended to arrest Serbs regardless of whether they had committed any crime. The recording was subsequently identified as a deepfake and publicly denied by the relevant authorities. According to the National Democratic Institute (NDI), the original post containing the recording received more than 150,000 views, demonstrating the potential reach of synthetic media during a period of heightened political and interethnic tensions (NDI, 2024).

AI-generated video portraying Serbian youth supporting Kosovo's recognition

In October 2025, an AI-generated video circulated on social media portraying a group of Serbian youth, with a young woman in the foreground calling on Serbia to recognise "the reality and Kosovo." The video also displayed the message "Recognize Kosova!" and presented the statement as representing the views of Serbian youth. The video was found to be artificially generated: Hibrid.info identified the Sora logo in the footage, while Hive Moderation assessed a 92% probability that the video was AI-generated. Hibrid.info therefore classified the content as deceptive. The case is particularly relevant to Kosovo–Serbia relations because it demonstrates how generative AI can fabricate a synthetic public voice and attribute a politically sensitive position to an apparently identifiable group (Rizanaj, 2025).

Rather than manipulating the statement of a real Serbian public figure, the video created an apparently authentic scene involving ordinary Serbian youth and used it to communicate a highly sensitive message concerning Kosovo's status (Rizanaj, 2025).

These cases illustrate how deepfake technology can be used to fabricate political communication and attribute to influential public figures statements that they never made. By placing fabricated statements within an already sensitive Kosovo–Serbia context, such content can undermine trust in political institutions, distort perceptions of government actions and further intensify existing ethnic and political divisions. The cases therefore demonstrate that the risks associated with AI-generated content are not merely hypothetical in the Kosovo–Serbia information space, but have already emerged during periods of heightened political tension (NDI, 2024).

These developments demonstrate that the growing use of AI-generated and manipulated content adds a new layer of risk to an already fragile information environment, making the ability to detect and counter synthetic media essential for preventing further escalation of tensions and protecting public trust in the Kosovo–Serbia context.

03/

Impact of Deepfakes and AI

The growing accessibility of artificial intelligence has made it easier to create realistic synthetic audio, images and videos, making it increasingly difficult to distinguish authentic information from manipulated content. This creates risks for information integrity, particularly in political and electoral contexts (UNESCO, 2025).

It is important to distinguish deepfakes from disinformation. Disinformation refers to false or misleading information deliberately disseminated to deceive or influence audiences, while deepfakes are AI-generated or AI-manipulated audiovisual content. Deepfakes can therefore be used as a tool to produce and spread disinformation (UNESCO, 2025; ITU, 2024).

The growing use of deepfakes and AI-generated content poses significant risks to the information environment surrounding Kosovo–Serbia relations. The National Democratic Institute (NDI) identified deepfakes and AI-generated content as an emerging form of information disorder in Kosovo during 2023 (NDI, 2024). Deepfakes can fabricate realistic images, videos and audio recordings that make individuals appear to say or do things that never actually occurred (UNESCO, 2024). Such content can make it increasingly difficult for audiences to distinguish authentic information from manipulated or synthetic material (Naffi, 2025).

The risks are particularly significant in politically polarised and ethnically sensitive environments, where manipulated content can reinforce existing narratives and deepen social divisions (UNESCO, 2024). In the Kosovo–Serbia context, the 2023 information environment demonstrated how manipulated content could be used around politically and ethnically sensitive events, including the crisis in northern Kosovo and tensions between the two sides (NDI, 2024).

The use of synthetic media to imitate political figures can further undermine trust in political communication, because fabricated statements may appear to originate from people who hold real political authority (UNESCO, 2024). This creates a particular risk during periods of heightened tension, when audiences may react to manipulated content before its authenticity can be verified.

Addressing these risks requires stronger institutional capacity to identify, verify and respond to manipulated and AI-generated content. UNESCO emphasises the importance of media and information literacy in helping citizens critically assess information and recognise the challenges created by AI-generated content (UNESCO, 2025). Public institutions should therefore invest in media and AI literacy programmes that equip citizens with the skills needed to question sources, verify information and identify potentially manipulated content (UNESCO, 2025). Cooperation between state institutions, journalists, fact-checking organisations, civil society and digital platforms should also be strengthened to facilitate faster identification and public clarification of sus-

pected deepfakes. UNESCO highlights the importance of cooperation among media actors, technology platforms and other stakeholders in promoting information integrity in an increasingly AI-driven environment (UNESCO, 2024).

At the same time, any state response should protect freedom of expression and avoid measures that could be used to restrict legitimate political speech. This balanced approach is essential for ensuring that efforts to counter AI-enabled manipulation strengthen, rather than weaken, public trust and democratic resilience.

04/

Conclusion and Recommendations

AI-generated disinformation and deepfakes are becoming an increasingly important challenge for the Kosovo–Serbia information space, with manipulated content contributing to polarisation, mistrust and the spread of misleading narratives (NDI, 2024; Ilazi et al., 2025). The response should therefore focus on prevention rather than reaction, by strengthening fact-checking capacities, improving AI and media literacy, and establishing clear institutional mechanisms for the rapid identification, verification and communication of manipulated content (SEE Check, 2026; Ajanović, 2026).

In Kosovo, the response should be coordinated through existing institutional mechanisms, with responsibilities clearly assigned among relevant state and non-state actors. The Office of the Prime Minister should provide overall inter-institutional coordination and support the development of protocols for the identification, verification and public communication of high-impact AI-generated disinformation. The Independent Media Commission should monitor manipulated content disseminated through audiovisual media. The Central Election Commission should incorporate AI-generated disinformation into election-related monitoring and response mechanisms, particularly during electoral periods. The Kosovo Police should address cases in which deepfakes or other forms of synthetic content are linked to criminal activity, fraud or cyber-enabled threats, in cooperation with the competent prosecutorial authorities. Fact-checking organisations, independent media, civil society and research institutions should complement these mechanisms through independent monitoring, verification and early warning. Relevant educational institutions and public bodies should also strengthen AI and media literacy among journalists, public officials and young people.

In Serbia, a comparable approach should be implemented through existing national institutions and mechanisms. Relevant government bodies should coordinate the institutional response and public communication, while the Regulatory Authority for Electronic Media (REM) should monitor manipulated content disseminated through regulated broadcasters. The Republic Electoral Commission (RIK) should incorporate AI-generated disinformation into election-related monitoring and reporting mechanisms, particularly during electoral periods. The Ministry of Information and Telecommunications should support measures related to media and digital literacy and cooperation with relevant stakeholders, while the Ministry of Internal Affairs and competent cybercrime structures should address cases involving the criminal misuse of synthetic content. Independent fact-checking organisations, media, civil society and research institutions should provide complementary monitoring, verification and early-warning capacity.

Given the cross-border nature of the information environment, Kosovo and Serbia should establish designated institutional points of contact within their existing institutional structures to facilitate rapid information-sharing on verified or suspected AI-generated disinformation, particularly during elections and periods of political tension. These focal points should operate through clear reporting, verification and escalation protocols and should be able to engage, where appropriate, with fact-checking organisations, media, civil society, researchers and relevant online platforms.

Cooperation with social media platforms should be strengthened to improve the reporting, verification and labelling of manipulated content and, where legally justified, the timely restriction or removal of harmful content, while safeguarding freedom of expression and access to information.

Both sides should also provide regular training for public institutions, journalists and fact-checkers on detecting and verifying synthetic content, while strengthening AI and media literacy among young people. The EU AI Act (2024) and relevant European Commission initiatives provide an important reference framework for developing such safeguards.

Ultimately, building resilience against AI-generated disinformation requires a whole-of-society approach in which national institutions retain clearly defined responsibilities, while independent media, fact-checkers and civil society provide independent verification and early warning, and bilateral mechanisms facilitate rapid cross-border coordination (EEAS, 2026; Ilazi et al., 2025).

In this context, strengthening resilience against AI-generated disinformation is not only a technological challenge but a shared responsibility, essential to protecting public trust, democratic stability and sustainable relations between Kosovo and Serbia.

References

Ajanović, A. (2026). *Next generation what we know: Mis/disinformation in the Western Balkans.* British Council. https://doi.org/10.57884/EESX-5952

Beselica, L. (2025, October 29). Video e krijuar me inteligjencë artificiale e disa të rinjve serbë për njohjen e Kosovës. *Kallxo.* https://kallxo.com/krypometer/video-e-krijuar-me-inteligjence-artificiale-e-disa-te-rinjve-serbe-per-njohjen-e-kosoves/

Council of Europe. (2026). *Guidance note on the implications of generative artificial intelligence for freedom of expression.* Council of Europe. https://edoc.coe.int/en/artificial-intelligence/12525-guidance-note-on-the-implications-of-generative-artificial-intelligence-for-freedom-of-expression.html

European External Action Service. (2026, March 12). *4th EEAS annual report on foreign information manipulation and interference threats.* https://www.eeas.europa.eu/eeas/4th-eeas-report-foreign-information-manipulation-and-interference-threats_en

European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* EUR-Lex. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Frau-Meigs, D. (2024). *User empowerment through media and information literacy responses to the evolution of generative artificial intelligence (GAI).* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000388547

Hoti, I. (2023, September 29). Rrena në mediat serbe: Audio-inçizimi nuk thotë "edhe sikur të vdesë, asgjë". *Kallxo.* https://kallxo.com/krypometer/rrena-ne-mediat-serbe-audio-incizimi-nuk-thote-edhe-sikur-te-vdese-asgje/

Ilazi, R., Loshaj, J., Pavlović, T., & Bytyçi, A. (2025). *How does disinformation target and affect interethnic relations in Kosovo?* Kosovar Centre for Security Studies. https://qkss.org/en/publikimet/ne-c-menyre-dezinformimi-shenjestron-dhe-ndikon-ne-marredheniet-nderetnike-ne-kosove

International Telecommunication Union. (2024). *Detecting deepfakes and generative AI: Report on standards for AI watermarking and multimedia authenticity workshop.* https://www.itu.int/pub/T-AI4G-AI4GOOD-2024-7

Naffi, N. (2025, October 1). *Deepfakes and the crisis of knowing.* UNESCO. https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing

National Democratic Institute. (2024). *Information disorders in Kosovo: 2023 report.* https://ndi.org/publications/information-disorders-kosovo-2023-report

Parliamentary Assembly of the Council of Europe. (2026). *Violence and hate speech against politicians: A threat to democracy* (Resolution 2646). Council of Europe. https://pace.coe.int/en/files/35987/html

Rizanaj, F. (2025, October 21). Video me pretendimin për të rinjtë serbë, e gjeneruar me inteligjencë artificiale. *Hibrid.info.* https://hibrid.info/video-me-pretendimin-per-te-rinjte-serbe-e-gjeneruar-me-inteligjence-artificiale/

Sbunker. (2025, July 3). *Manipulimi me AI: Pretendohet rrejshëm se Trumpi paralajmëroi sulm ndaj Serbisë për shkak të Kosovës.* https://sbunker.org/sr/disinfo/manipulimi-me-ai-pretendohet-rrejshem-se-trumpi-paralajmeroi-sulm-ndaj-serbise-per-shkak-te-kosoves/

SEE Check. (2026, June 29). *State of disinformation in the SEE region: SEE Check's third report.* https://seecheck.org/index.php/2026/06/29/state-of-disinformation-in-the-see-region-see-checks-third-report/

UNESCO. (2024). *AI and the future of journalism: An issue brief for stakeholders.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391214

UNESCO. (2025). *Towards user empowerment: A multistakeholder action plan for integrating media and information literacy on digital platforms.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000394855



NEW
SOCIAL
INITIATIVE