Co-funded by the European Union

EUROPEAN UNION

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Co-funded and implemented by the Council of Europe

June 2026

Information Integrity Conference (DISICON 9)
Pristina 8-9 December 2025

Towards an Ethical, Human Rights, and Gender-Responsive Artificial Intelligence in the Western Balkans: Conference Report

The role and responsibility of the Council of Europe in protecting freedom of expression has been underlined in the "Reykjavik Principles for Democracy", the Reykjavík Declaration – United around our values and in the New Democratic Pact for Europe.

The report was prepared by Patricia Shaw, Council of Europe Consultant, following the DISICON conference, with the support of the action 'Protecting Freedom of Expression and of the media (PRO-FREX)', implemented by the Council of Europe Division for Cooperation on Freedom of Expression under the European Union and Council of Europe's joint programme "Horizontal Facility for the Western Balkans and Türkiye.", in close partnership with the National Democratic Institute (NDI Kosovo). Participants in the workshops with the Regional Information Integrity Form provided written input which contributed to this report; their feedback has been incorporated into the final version.

This publication was produced with the financial support of the European Union and the Council of Europe. Its contents are the sole responsibility of the author(s). Views expressed herein can in no way be taken to reflect the official opinion of the European Union or the Council of Europe.

PRO-FREX enables the beneficiary institutions and civil society organisations to progress towards meeting their reform agendas in the field of freedom of expression and freedom of media, in line with the European standards. It aims to contribute to an improved environment for the exercise, in particular by journalists and media actors, of their rights of freedom of expression, in a more pluralistic and safer media environment, in line with the standards as set by in accordance with Article 10 of the European Convention of Human Rights (ECHR).

www.coe.int/freedomofexpression

Table of Contents

1. Executive Summary

Artificial Intelligence (AI) is rapidly reshaping political, social, and economic systems worldwide. While AI presents transformative opportunities for innovation, efficiency, and improved public services, it also creates significant and systemic risks[1]. Use of AI is particularly challenging in areas which may be susceptible to democratic fragility, limited institutional capacity and/or resources, constrained regulatory oversight, and ongoing political transition. The six Western Balkan Beneficiaries, Albania, Bosnia and Herzegovina, Kosovo*, Montenegro, North Macedonia, and Serbia, face these challenges to greater or lesser degrees.

This document summarises the discussions held during the Information Integrity Conference (DISICON 9), held in Kosovo on 8-9 December 2025. The conference brought together a diverse group of stakeholders from across the Western Balkans, including government officials, civil society organisations, journalists, media regulators, technology experts, women's rights advocates, academic researchers, and international observers.

Participants engaged in structured small-group discussions examining seven critical AI risk areas through the lens of their Beneficiary-specific domestic and regional contexts. The workshop methodology was designed to identify not only the risks themselves, but also the stakeholders who bear those risks, the institutions responsible for addressing them, and suggestions by the author(s) for practical steps needed to build effective governance.

The seven thematic areas examined were:

1. Misinformation, disinformation, and fake news;
2. Misuse of surveillance technologies;
3. Discrimination in automated decision-making;
4. Cybersecurity weaknesses;
5. Opacity in public AI procurement;
6. Gendered harms; and
7. Ensuring inclusive AI governance.

---

[1] According to the EU AI Act Recital 110, "Systemic risks" *should be understood to increase with model capabilities and model reach, can arise along the entire lifecycle of the model, and are influenced by conditions of misuse, model reliability, model fairness and model security, the level of autonomy of the model, its access to tools, novel or combined modalities, release and distribution strategies, the potential to remove guardrails and other factors. In particular, international approaches have so far identified the need to pay attention to risks from potential intentional misuse or unintended issues of control relating to alignment with human intent; chemical, biological, radiological, and nuclear risks, such as the ways in which barriers to entry can be lowered, including for weapons development, design acquisition, or use; offensive cyber capabilities, such as the ways in vulnerability discovery, exploitation, or operational use can be enabled; the effects of interaction and tool use, including for example the capacity to control physical systems and interfere with critical infrastructure; risks from models of making copies of themselves or 'self-replicating' or training other models; the ways in which models can give rise to harmful bias and discrimination with risks to individuals, communities or societies; the facilitation of disinformation or harming privacy with threats to democratic values and human rights; risk that a particular event could lead to a chain reaction with considerable negative effects that could affect up to an entire city, an entire domain activity or an entire community.*

* This designation is without prejudice to positions on status and is in line with UNSCR 1244 and the ICJ opinion on the Kosovo Declaration of Independence.

Following DISICON 9, it was decided to develop and distribute a follow-up survey to workshop participants in January-February 2026, enabling them to continue sharing their expertise and priorities. This survey was designed to finetune the "suggestions for next steps" presented below and ascertain that they are appropriate for the Western Balkans context; identify the specific stakeholders concerned; the way they would follow these steps, including identifying constraints, barriers, and the policy environment needed.

The questionnaire covered eight thematic areas:

1. Beneficiary-specific AI governance structures;
2. Adoption of AI legal and regulatory frameworks;
3. Transparency obligations;
4. Regulation of high-risk AI systems;
5. Protection of democracy and electoral integrity;
6. The role of civil society;
7. The role of the Regional Information Integrity Forum (RIIF);and
8. Protection of women, minorities, and vulnerable groups.

It was agreed that follow-up activities would actively seek gender-specific inputs by engaging gender-specific stakeholder representatives more systematically, recognising that gender perspectives were underrepresented in the initial workshop.

An online workshop was convened on 23 February 2026 to validate and deepen the questionnaire findings. Participants from, Albania, Bosnia and Herzegovina, Kosovo, Montenegro and Serbia, engaged in plenary discussion and Beneficiary-specific breakout groups under Chatham House rules.

Participants emphasised that without foundational rule of law, data governance, transparency, and institutional capacity, there is a risk that not just AI but AI regulation would simply scale existing systemic problems.

1.1 Key Findings from the Workshops and Survey

This document summarises the input provided in the workshops and survey presented above, engaging regional stakeholders. Summary of the findings:

1. Proactively Informing Citizens received the highest urgency rating across all 8 sections: 100% of respondents rated it Top or High Priority, with 73% rating it Top Priority. This signals unanimous commitment to an informed citizenry.

2. Banning Undisclosed Political Deepfakes was rated Top Priority by 87% of respondents, making it the single highest-rated Top Priority item across the entire questionnaire.

3. Democracy and Electoral Integrity showed the strongest regional consensus overall, with most items receiving 85%+ combined Top/High Priority ratings.

4. Monitoring Government AI Deployment and Monitoring AI in Elections both received 86% Top Priority ratings, underscoring civil society's strong emphasis on democratic accountability.

5. Protecting women and minority candidates from AI-enabled hate campaigns was rated Top or High Priority by 93% of respondents, showing this as a near-universal concern.

6. Participants of the workshops underlined that AI Oversight and Governance bodies would need clear definition, terms of reference, and constitutional status.

7. Social Benefits AI regulation was ranked lowest among high-risk categories in Section 4, despite well-documented international precedent (Dutch SyRI, Australia's RoboDebt, UK Universal Credit) showing disproportionate harm to women, minorities, and vulnerable groups.

8. CSO-centred and mixed governance models are strongly preferred. Respondents overwhelmingly favoured mixed or CSO-led approaches for monitoring and oversight, reflecting a desire for multi-stakeholder accountability.

9. The RIIF is seen as a key regional coordination mechanism. Members consistently assigned RIIF lead coordination responsibilities alongside governments and international partners.

10. Sectoral AI Mapping was the highest priority in the Legal/Regulatory section with 64% rating it Top Priority, signalling a strong desire to better understand the AI landscape.

11. Some Western Balkan Beneficiaries provided important insight into the complex administrative structure of some of the Beneficiaries and emphasised the need for Beneficiary-specific-level rather than entity-level coordination. Other Western Balkan Beneficiaries noted the historical and intercultural complexities that the region's heritage brings

13. Many respondents and participants cautioned against risks of simply copy-pasting EU governance structures without commensurate resources, capacity, and institutional infrastructure to support the new legislation and bodies. The Council of Europe Framework Convention on Artificial Intelligence, which was designed to be adaptable and applicable regardless of the institutional capacity or administrative structure, has particular importance in this regard.

1.2 Next Steps Identified from the Workshops and Survey

1. **International Legal Commitment.**

All Western Balkan Beneficiaries are encouraged to consider signing and ratifying the Council of Europe Framework Convention on AI.

2. **Democratic Protection.**

Undisclosed political deepfakes are encouraged to be banned, with election commissions equipped with rapid response units, and collaboration frameworks established with platforms to counter synthetic electoral content. Multilateral observation mechanisms for information integrity across the region are therefore encouraged. This is not only a challenge faced solely by the region but also by the Council of Europe member States and Beneficiaries.

3. **Civil Society and Regional Cooperation.**

CSOs are encouraged to monitor government AI procurement, conduct independent audits, and build public AI literacy. The Regional Information Integrity Forum is encouraged to coordinate

cross-border disinformation monitoring and publish annual reports on AI and democracy in the Western Balkans.

4. **Gendered AI Harms.**

Participants suggested that online gender-based violence, including AI-generated content, should be formally recognised in law. Political parties are encouraged to adopt codes of conduct against gender-based harassment and sexualised intimidation, with dedicated law enforcement units created to address such issues. Mandatory gender equality impact assessments are also encouraged to be introduced for all public-sector AI. Existing practical tools are available for immediate regional deployment.

1.3 The Case for Action

The Western Balkans stand at a pivotal moment. Artificial intelligence is not a future technology; it is already being deployed across the region. The question is: will AI be used to strengthen democratic institutions or undermine them? Will AI be used in public administration, policing, border management, elections, media, and social services to disproportionately harm based on gender ethnicity or being a person in vulnerable circumstances. Participants expressed a shared commitment to collectively influence how AI will shape the region's future, aiming for a future grounded in democratic governance, human rights protections, and gender equality, rather than one defined by unaccountable algorithms, opaque surveillance, or discriminatory automation

The urgency of action in 2026 is driven by several converging factors:

- **The Framework Convention is open for ratification.** Beneficiaries that sign and ratify now will be helping to shape the implementation standards and interpretive guidance that will govern AI worldwide for decades to come.
- **The EU AI Act is being implemented.** As the EU's comprehensive AI regulation enters full operation, Western Balkan Beneficiaries pursuing EU accession should seek to demonstrate alignment with its requirements, as all benefit from internationally consistent approaches. Delay in alignment risks creating an ever-widening governance gap that will become harder and more expensive to close.
- **AI risks are escalating rapidly.** The proliferation of generative AI, deepfake technologies, and agentic AI systems means that the harms identified during these discussions are intensifying with each month of inaction. Election cycles across the region will increasingly be targeted by AI-enabled manipulation. Surveillance technologies are becoming more sophisticated and more widely deployed. Discriminatory automated decision-making is expanding into new domains.
- **Geopolitical pressures are intensifying.** The Western Balkans remain a theatre of geopolitical competition, with foreign actors - including those less committed to democratic governance - offering AI technologies and partnerships that may not align with human rights standards. Anchoring the region's AI governance in the Framework Convention and EU AI Act provides a counterweight to these pressures.

- **The chilling effect on women's political participation is documented and measurable.** Regional monitoring has demonstrated that heightened visibility during election cycles increases women's exposure to online abuse, with misogynistic attacks persisting through campaigns, parliamentary negotiations, and local races as a consistent phenomenon. Women should not be put in a position where they have to choose between visibility and safety, or between leadership and dignity. Protecting women in politics is not only a matter of gender equality; it is a prerequisite for resilient, inclusive, and ethical democratic systems.
- **Civil society and public expectations are rising.** The DISICON 9 workshops demonstrated that stakeholders across the region - from government officials to civil society leaders to journalists - are aware of AI risks and demanding action. Failure to respond will erode public trust in democratic institutions at a time when that trust is already fragile.

The combination of the Framework Convention and the EU AI Act offer the Western Balkans a uniquely advantageous governance pathway for:

- International legitimacy;
- EU accession alignment;
- Human rights foundation;
- Gender responsiveness; and
- Practical and adaptable frameworks.

2. The International Regulatory Landscape: The Council of Europe Framework Convention and the EU AI Act

Views expressed in this chapter can in no way be taken to reflect the official opinion of the European Union or the Council of Europe.

2.1 The Council of Europe Framework Convention on Artificial Intelligence

The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225) is the first legally binding international treaty on AI. Adopted by the Council of Europe Committee of Ministers on 17 May 2024 and opened for signature on 5 September 2024 in Vilnius, Lithuania.

The Framework Convention was negotiated by the 46 Council of Europe member Beneficiaries, the European Union, and 11 non-member Beneficiaries including the United States, Canada, Israel, Japan, and others. As of June 2026, signatories include Albania, Andorra, Armenia, Bosnia and Herzegovina, Canada, Georgia, Iceland, Israel, Japan, Liechtenstein, the Republic of Moldova, Montenegro, North Macedonia, Norway, San Marino, Switzerland, Ukraine, the United

Kingdom, the United States of America, Uruguay, and the European Union (representing its 27 Member-States).

The Convention establishes fundamental principles for AI governance throughout the whole AI system lifecycle, such as:

- Protection of human rights, including privacy and non-discrimination
- Preservation of democracy and democratic institutions
- Upholding the rule of law
- Transparency and accountability of AI systems and algorithms
- Mandatory human oversight for high-risk AI systems
- Non-discrimination, gender equality, and protection of vulnerable groups
- Access to effective remedies for individuals harmed by AI systems
- Inclusive and participatory regulatory processes
- Possibility for authorities to introduce bans or moratoria on certain AI applications (so-called red lines), calling predominantly for safe innovation.

The Convention is technology-neutral and covers AI systems used by public authorities, private actors acting on behalf of public authorities, and private actors' use of AI. It establishes a Conference of the Parties (CoP) as its follow-up mechanism to monitor implementation, issue recommendations, and ensure long-term compliance.

A core feature is its AI risk-management framework, supported by the HUDERIA Methodology and Model (Risk and Impact Assessment of AI Systems from the Perspective of Human Rights, Democracy and the Rule of Law), a non-legally binding tool developed by the Council of Europe to help Beneficiaries assess AI-related risks and impacts.

The Convention places particular emphasis on the obligations of states towards individuals whose human rights, democratic participation, or rule of law protections may be affected by AI systems. Its methodology (HUDERIA) centres the assessment on the impact to the rights-holder - meaning the person subject to a consequential decision or systemic effect - rather than primarily on technical categories of risk.

For the Western Balkans, stakeholder discussions highlighted that key challenges relate to the deployment of AI by public authorities, for example in surveillance, biometric policing, welfare, and electoral processes.

2.2 The European Union AI Act

The EU AI Act, which entered into force in stages from August 2024, introduces a risk-based classification system for AI systems:

- **Unacceptable risk:** AI systems that pose an unacceptable risk to fundamental rights are prohibited outright, including social scoring, manipulative subliminal techniques, and real-time biometric identification in public spaces for law enforcement (with narrow exceptions).
- **High risk:** AI systems used in critical areas like justice, elections, policing, welfare, migration, employment, education, are all subject to obligations concerning design and

development requirements, risk management, human oversight, transparency, data governance (including specifically for bias), quality management, conformity assessments, technical documentation, and registration requirements. Where there are public sector deployments, public authorities are also required to conduct a fundamental rights impact assessment (FRIA).

- **Limited risk:** AI systems interacting with the public (chatbots, deepfake generators) are required to meet minimum transparency obligations, including disclosure that content is AI-generated[2].
- **Minimal risk:** Most AI systems fall into this category and are largely unregulated, although adherence to voluntary codes of conduct is encouraged.

Like the Framework Convention, the EU AI Act is developed with a human-centric perspective, its Article 1 states the purpose of the EU AI Act is to:

"[I]mprove the functioning of the internal market and prompt the update of human-centric and trustworthy AI while ensuring a high level of protection of health, safety and fundamental rights enshrined in the Charter[3], including democracy, the rule of law, and environmental protection, against the harmful effects of AI systems in the union and supporting innovation".

This appears also to be at the heart of the positive change the Western Balkans want to see for the region.

Key obligations of the EU AI Act with direct relevance to the Western Balkans include fundamental rights impact assessments for high-risk AI deployed by public sector, data governance, heightened transparency for public-facing AI systems, bans on harmful AI practices including biometric mass surveillance, independent oversight authorities for compliance and enforcement, and strict rules on deepfakes and synthetic media. Furthermore, the EU AI Act seeks to address the risks of undue external interference with the right to vote, adverse effects on democracy and the rule of law, in particular where AI systems can influence the outcome of an election or referendum or voting behaviour of natural persons, by classifying such systems as high-risk AI[4].

While the EU AI Act establishes important obligations relating to risk mitigation, transparency, and accountability, its scope is principally oriented toward product safety and regulatory conformity within the single market. By comparison, the Council of Europe Framework Convention addresses the broader rights-based impacts of AI, focusing on state responsibilities to individuals. Both frameworks are complementary, offering Governance tools suited to different aspects and stages of AI policy development.

Western Balkan stakeholders recognised that alignment with both may strengthen human rights protections, democratic integrity, and regional cooperation.

2.3 Why Both Frameworks Matter for the Western Balkans

[2] See Article 50, EU AI Act.
[3] Charter of fundamental rights of the European Union 2012/C 326/02, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT
[4] Recital 62 EU AI Act

All six Western Balkan Beneficiaries are at various stages of EU accession. Alignment with the EU AI Act is not merely aspirational, it is a legal and political prerequisite for EU membership.

The Council of Europe Framework Convention also plays an essential role by providing a broader human rights foundation that addresses obligations toward individuals whose rights, democratic participation, and rule of law protections may be affected by AI. Importantly, the Convention should not be viewed as a temporary measure or a placeholder before EU accession; its commitments are permanent and remain relevant regardless of a country's EU status. Both frameworks are complementary but cannot substitute for one another, and together they offer a stronger basis for rights-based AI governance in the region.

3. Horizontal Preconditions: The Enablement Environment

This section identifies the cross-cutting preconditions that demonstrably need to be in situ for effective AI governance across the Western Balkan Beneficiaries. These are not merely contextual factors but foundational requirements without which AI-specific regulation cannot function. All of the aspects identified clearly overlap and undergird the Framework Convention and the EU AI Act.

3.1 Rule of Law as a Prerequisite

Any AI governance approach needs to take into account the overall standards in the relevant Beneficiary concerning adherence to the rule of law and the extent to which the judiciary is impartial and independent, the capacity of regulators to enforce the law, the functioning of data protection authorities and/or other authorities and fora (including ombudsperson institutions) to provide procedural mechanisms for access to redress for individuals or groups harmed or potentially harmed by AI.

3.2 Data Governance as a Foundation

AI governance cannot be built without robust data governance. To get AI right requires a "data-first" approach i.e. getting data regulation in order and enforceable as a prerequisite for getting ready for AI regulation. This includes alignment with data protection principles established in the Council of Europe Convention 108 + and the EU's GDPR:

- Lawfulness, Fairness, and Transparency: Data needs to be processed legally, in a way people expect, and with open communication about how it is used.
- Purpose Limitation: Data should only be collected for specified, explicit, and legitimate purposes and not used for anything else.
- Data Minimisation: Organisations need to only collect and process the minimum amount of data necessary for the purpose.
- Accuracy: Personal data needs to be kept accurate and up to date, with inaccurate data erased or rectified.
- Storage Limitation: Data should only be stored for as long as necessary to fulfil the purpose for which it was collected.

- Integrity and Confidentiality (Security): Data needs to be processed securely to protect against unauthorised access, loss, or destruction.
- Accountability: Data controllers are responsible for complying with the above principles and needs to be able to demonstrate their compliance.

Across the region, data protection laws aligned with Convention 108+ (which has been ratified by Albania, Bosnia and Herzegovina, Croatia, North Macedonia, Serbia, and Slovenia) and GDPR are being adopted. However, even where laws are aligned, implementation and enforcement need to be visible and commensurate to the risks. It is acknowledged that where data protection authorities exist across the region, they often have limited budgets, limited human resources, and/or lack technical capacity for holding data controllers and processors (especially concerning those utilising AI and/or methods of autonomous decision making) accountable. Institutions without sufficient authority may undermine the law, creating a fragmented level of privacy and data protection across the region. Lessons from both GDPR and Convention 108 should inform preparations for AI governance in the Western Balkans.

3.3 Institutional and Regulatory Capacity

According to the feedback of the Beneficiaries, all six Western Balkan Beneficiaries appear to have limited institutional capacity to oversee AI deployment. Regulatory bodies often lack the technical expertise, financial resources, and political independence necessary to monitor, audit, and enforce AI governance standards whether within the public sector or private sector. Conference participants underlined that effective accountability and scrutiny mechanisms (including judicial scrutiny) are needed, especially where increasingly more complex AI systems are being implemented, making traceability and investigations even harder. The complexity of the technologies being rolled out, combined with a lack of resources, could risk being used to undermine government institutions (and not enable them to give full effect to their democratically legitimated mandates) and disempowers populations who should be afforded necessary human rights protections (and be enabled to hold government institutions to account).

3.4 Transparency

Meaningful public sector transparency is key to both trust and democratic legitimacy. Where public or private institutions refuse to release information, this significantly undermines trust and fails to build capacity within populations to adequately understand AI developments and deployments. Transparency was identified by Beneficiaries as a necessary precondition for accountability: without transparency, accountability is impossible, but where the mechanisms to enforce transparency are themselves absent or dysfunctional, these are highly encouraged to be rectified as a priority. For transparency to be effective, it needs to be designed with its intended audience in mind: open-source code may serve academia, but citizens and their representatives through civil society organisations require more accessible and explainable forms of disclosure.

The EU integration process is the most promising lever for regional engagement in good transparency practice. To this end, Beneficiary-specific commissioners on access to information need significantly greater support and capacity to ensure transparency is given full and proper effect both concerning data use and how AI is used in public sector decisions, particularly those with human rights-critical outcomes.

3.5 Political Stability and Democratic Resilience

Participants suggested that AI governance structures need to be insulated from political pressure and be designed to withstand changes in government wherever possible. Beneficiaries within the region with upcoming elections are encouraged to urgently prioritise information integrity especially within their governance structures and particularly where AI is used for public sector communications.

It was underlined during the Conference that the region is particularly susceptible to foreign influence operations, including through AI-enabled misinformation and disinformation campaigns. Geopolitical misinformation and disinformation campaign manifest in the digital space, with AI tools increasingly being used to amplify divisive narratives, undermine democratic institutions, and influence electoral processes. Generative AI manipulated information does not only manifest in social media platforms where there is a shared information space where the same or similar content can be viewed by others, but also where users can interact with informative content specifically tailored for them in an isolated and automated way, a so-called "audience of one"[5]. This threatens to further fragment dissemination of information and informative content. This shift in consumption of information undermines a shared and pluralistic information space, which is essential for democracy and political stability.

Robust information integrity and AI governance both in social media platforms and in the use of generative AI is therefore also a matter of national and regional security.

3.6 Civil Society Ecosystem

Civil society organisations play a critical role in monitoring for AI governance and providing informational autonomy to individuals and groups. However, the civil society ecosystem faces severe challenges including limited funding, brain drain (with human resources leaving the region), project-dependent capacity, insufficient technical expertise, and limited voice to represent concerns in the civic space in some jurisdictions.

There is a specific operational gap: CSOs often lack the knowledge either to know what to ask for in terms of transparency, or to understand what they receive when transparency is provided. Most CSOs in the region are led by people with social sciences or humanities backgrounds, not technical ones, while technology-led organisations lack human rights knowledge and experience. Skills and talent matching across these groups is essential.

Strengthening civil society capacity for AI oversight is a horizontal priority.

3.7 Economic Constraints

Limited public budgets constrain the ability of governments to invest in AI governance infrastructure. International donor support and regional pooling of resources is therefore critical, particularly in the early stages to enable capacity to be built. Pooling resources in the region to

[5] Guidance Note on the Implications of Generative Artificial Intelligence for Freedom of Expression Adopted by the Council of Europe Steering Committee on Media and Information Society (CDMSI) at its 28th Plenary meeting, 3-5 December 2025, https://rm.coe.int/cdmsi-2025-15rev-guidance-note-on-the-implications-of-generative-artif/488029df80

ensure AI regulation (and the institutions and infrastructure necessary for it to succeed) can be implemented and be effective is encouraged to be considered a priority.

Rather than being seen as innovation stifling, AI regulation has been shown to crucially support business and citizen certainty. Where there is legal certainty with commensurate protections, this in turn has been seen to promote safer and ethical adoption of other technologies. There is therefore no reason why this could not aide promotion of AI adoption where it to is used for responsible and safe innovation which is human-rights respecting. There is greater potential scope for economic growth through access to the EU single market. Barriers to AI adoption and EU market access are encouraged to be addressed by Beneficiary-specific governments in conjunction with Civil Society.

3.8 Digital Infrastructure, Connectivity, and Digital Skills

The Western Balkans face significant disparities in digital infrastructure, both between Beneficiaries and between urban and rural areas within Beneficiaries. Whilst it is acknowledged that mobile connectivity has greater saturation enabling internet use by individuals, reliable broadband connectivity, cloud computing capacity, and data centre infrastructure are unevenly distributed, and with that the access and skills to utilise it. Uptake is also higher in Western Balkan Beneficiaries where populations are younger and more likely to be digital natives.

According to the OECD, "digital skills are low, particularly among older age groups, with progress limited by insufficient opportunities for upskilling or training"[6]

Without adequate digital infrastructure, digital inclusion, and digital skills, AI governance frameworks risk being theoretical rather than practical. Investment in digital infrastructure and education for digital inclusion is a prerequisite for effective AI oversight, and e-government. Digital connectivity and access to government services and information for citizens, academia, civil society, an independent and diverse media, and businesses is essential for government and AI decision transparency. Once these groups have access to the digital, it is vital that firstly access is inclusive, and secondly, they have the commensurate skills to access and receive not only government services, but most importantly reliable information to be a well-informed citizenry. Civil society and the media can play an important role in awareness raising of issues and challenges concerning the information space for the public, whilst also providing opportunities for fact-checking.

The right to seek, receive and impart information is an important human right, and access to a verified information and an integrous information space is a vital preventative measure against misinformation, disinformation and fake news to ensure use of digital infrastructure and connectivity is safe, ethical and promoting human rights.

3.9 Education: AI literacy and socio-technical understanding

[6] Economic Convergence Scoreboard for the Western Balkans, 10 June 2025, https://www.oecd.org/en/publications/economic-convergence-scoreboard-for-the-western-balkans-2025_bc0babf3-en/full-report/digital-transformation-cluster_7cafb35e.html#:~:text=Recent%20data%20suggest%20that%20further,only%2030%25%20of%20EU%20levels.

Foundational education is a missing piece in the puzzle. This is not just about AI education for young people but capacity building for CSOs and wider society themselves. Much AI literacy to date currently focuses heavily on "how to use AI" rather than critical thinking about AI's impact on human dignity, autonomy, freedom of expression, freedom of opinion, and equality as well as the physical person in terms of cognition. More needs to be embedded in education systems from the ground up, to improve jobs and future workers but also in understanding of AI amongst other emerging technologies, and their wider societal implications for people and populations in the short, medium and long term, to create a well-informed citizenry.

3.10 EU Accession Context

All six Beneficiaries are at various stages of EU accession, creating both an opportunity and an obligation. Legislative alignment ought not to be at the expense of institutional efficacy otherwise this would result in a mere box-ticking exercise that could undermine integrity not just of one Beneficiary-specific context but potentially the whole region.

Participants suggested that the opportunity is to put in place Beneficiary-specific AI governance structures that do not simply duplicate those established at EU level but provide effective governance in the regional context. However regional experience shows that mere alignment with EU legislation can simply translate rather than also harmonise to the local context, risking legislative instrument which are not operationally feasible, given limited leverage over major technology companies, insufficient institutional capacity, and/or limited stakeholder capacity.

For some Beneficiaries, the complex administrative structure consisting of two levels of entities with competences distributed between local entity and Beneficiary-specific levels creates a specific challenge because agreement cannot be reached on the appropriate level of governance. Securing political will to initiate and complete reforms is therefore a key prerequisite.

4. Risks for the Region impacting Information Integrity, Human Rights, and Gender Equality

There are seven interconnected risk areas that particularly impact information integrity, human rights and gender equality in the region that demand urgent attention in order to ensure their protection across the Western Balkans. This document focuses on two in particular:

4.1 Misinformation, Disinformation, and Fake News

AI-generated propaganda, deepfakes targeting political leaders, synthetic content specifically attacking women (particularly women operating in the public sphere and/or political roles) and minorities, is a critical threat to democracy and gender equality in the region, and where it incites both hatred and fear, it risks political stability[7].

[7] Information Disorders in Kosovo 2023, NDI, May 2024, https://www.ndi.org/sites/default/files/INFORMATION%20DISORDERS%20IN%20KOSOVO%20-%202023-compressed_0.pdf

AI-enabled manipulation of elections has been seen in the Western Balkans. Monitoring of Kosovo's 2025 Beneficiary-specific domestic and local elections found that AI-generated and deepfake content was actively deployed to discredit women politicians, with fabricated videos circulating on social media platforms reaching hundreds of thousands of views. Gendered disinformation, hate speech, AI-generated content, and sensationalism now operate as an integrated ecosystem rather than as a separate phenomenon, making undermining of the shared information space more frequent, shareable, and harder to challenge.[8]

To address this requires urgent action to address how media and technology companies (particularly those involved in the information space) are regulated and respect international and domestic legislation, including human rights.

Independence and diversity of media ownership across the region is key to success in this area, both across traditional audiovisual media services and the largely unregulated online media platforms, along with putting in place measures to hold technology companies accountable for their outputs.

The fact that technology companies are currently not required to be transparent over data-gathering practices, algorithmic design, and the training data used for large language models, needs to be addressed to ensure appropriate and effective oversight and accountability mechanisms are put in place. Alignment with both the AI Framework Convention and the EU AI Act and would facilitate greater transparency and accountability in this regard.

4.2. Gendered AI Harms

While anyone can be impacted by AI-related harms, women in politics and women in the public sphere, women activists and female journalists face specific and heightened risks. These include AI-generated sexual deepfakes, disproportionate online harassment (sometimes also spreading to offline harassment), and manipulation and misrepresentation of images or altered videos.

These harms are captured by the concept of "gendered information disorders": a range of harmful information practices including disinformation, misinformation, harassment, and abusive narratives that disproportionately target women in public life because of their gender, with the aim or effect of discrediting, silencing, or excluding them.[9]

Regional evidence from Kosovo provides the most comprehensive dataset to date. Analysis of 660,758 social media comments during the 2025 election period found that 14% (91,525 comments) contained hate speech directed at women in politics - demonstrating that misogynistic discourse is structurally embedded in Kosovo's political online environment as a persistent feature of public debate, not an exception triggered by isolated events. In 2025, 96.2%

[8] Democracy for Development (D4D), unSOCIAL MEDIA II: Sensational online hate against women in politics (2025). Research produced with NDI through SDC, and Pikasa Analytics. The report monitored 162 media outlets and analysed 660,758 social media comments during Kosovo's 2025 national and local elections.

[9] NDI / DISICON 9, Journalist Guidelines for Covering Women in Public Life: Ethical, Accurate, and Responsible Reporting in the Digital Age (December 2025). Developed from the DISICONNECT workshop session with women journalists at DISICON 9. The definition of "gendered information disorders" is drawn from the Glossary of these Guidelines.

of hate-speech comments were authored by men (up from 89.6% in 2023), showing a consolidation of male-dominated gatekeeping over public discourse.

Documented cases include AI-generated deepfake videos of Kosovo's President and former President circulated on social media with offensive captions, and a fabricated AI video depicting a woman MP in a non-existent personal relationship to discredit her publicly. Women from minority communities face compounded discrimination, with gendered attacks intersecting with ethnic or nationalist rhetoric - a dual level of targeting rarely seen against men. Online hostility has shifted from occasional event-triggered spikes to sustained, systemic hostility across Beneficiary-specific domestic and local election cycles, with attacks following recognisable patterns of moral de-legitimisation, sexualised shaming, patronising infantilisation, and intellectual undermining.

Biased datasets that exclude or misrepresent women and minorities contribute to discriminatory outcomes, may be an area to address, but this is not the sole issue to be addressed concerning AI design, development and deployment to minimise discrimination.

Enforced accountability and responsibility for AI generated outputs that incite and result in gender-based harm and intimidation is vital. The formal recognition of online gender-based violence as violence in law, establishment of dedicated taskforces to address AI-related crimes against women, specialist departments for online gender-based crime, judicial training to ensure courts take such cases seriously, and coordinated civil society advocacy and policy research would be a tremendous start to address some of the ensuing issues.

Practical tools to address these harms have already been developed through NDI, including a Practical Guide for Women in Politics covering threat identification, first response protocols, incident documentation, platform reporting, and well-being support, and Journalist Guidelines for Covering Women in Public Life providing editorial standards on gendered framing, AI verification, and election coverage.[10]

5. Suggestions for Next Steps

This section presents the Author's suggestions for next steps which could be undertaken by various Beneficiaries stakeholders. These suggestions are based on the input provided by participants in the workshops and in the survey.

5.1 Sign and Ratify relevant Council of Europe instruments

Participants agreed that measures should be undertaken to sign and ratify relevant Council of Europe's instruments such as AI Framework Convention and Convention 108+. All participants could support the national processes of establishment of Beneficiary-specific AI governance structures

[10] NDI / Our Secure Future / DISICON 9, A Practical Guide for Women in Politics: Addressing Gendered Information Disorders, Online Violence, and Cyber Threats (2025). Developed from the DISICONNECT workshop session at DISICON 9, facilitated by experts from Our Secure Future, Council of Europe, and Lithuania. See also Kosovo's constitutional incorporation of the Istanbul Convention and GREVIO General Recommendation No. 1 on the digital dimension of violence against women.

5.2 Protect Democracy and Elections from AI Threats

Participants agreed that measures should be taken to ensure AI systems are not used to undermine the integrity, independence, and effectiveness of democratic institutions and processes, and to protect individuals' fair access to and participation in public debate and their ability to freely form opinions.

- Apply a structured assessment framework (such as COBRA) to identify, monitor, and mitigate risks relating to democratic processes, including but not limited to political deepfakes, digital manipulation, and AI-enabled hate campaigns.
- Develop effective mechanisms within election commissions and relevant bodies to address AI-related threats such as synthetic political content, ensuring responses are consistent with Article 5 of the Framework Convention
- Develop training programmes for political parties on responsible AI use.
- Establish collaboration frameworks with platforms to detect and remove synthetic electoral content.
- Provide targeted protections for women and minority candidates against AI-enabled hate campaigns, reflecting Convention's non-discrimination and democratic participation requirements

Multi-lateral communications and observation mechanisms: Practical and holistic measures are needed to combat disinformation in and around democratic processes. The region could provide valuable input to other Beneficiaries or Member-States (and vice versa), by establishing multilateral observation mechanisms for sharing pertinent information and finding, providing for effective cooperation and coordination to help maintain public trust.

5.3 Strengthen Civil Society and Regional Cooperation

For Civil Society Organisations

- Monitor government AI deployment and procurement.
- Conduct independent audits of AI systems in use.
- Identify and report discriminatory AI outcomes with a focus on gender and minority impacts.
- Participate in policymaking consultations at all stages.
- Build public awareness of AI rights and risks through media engagement and public education.
- Train journalists on AI literacy and investigative techniques for AI accountability.

Recent Council of Europe key standards and documents include:

- Guidance Note on the implications of generative AI for Freedom of Expression,
- Guidance note on countering the spread of online mis- and disinformation through fact-checking and platform design solutions
- Guidelines on the responsible implementation of artificial intelligence systems in journalism
- Resisting Disinformation: 10 Building Blocks to Strengthen Information Integrity
- Specific to North Macedonia: Artificial Intelligence in the Media of North Macedonia: Toward Common Standards for Regulation and Self-Regulation

The Journalist Guidelines for Covering Women in Public Life, developed by NDI and the Swiss Agency for Development and Cooperation may assist here by providing ready-made editorial standards addressing gendered framing, headline and visual practices, AI and deepfake verification, and election coverage protocols, including a newsroom policy template that can be adopted across the region.

For the Regional Information Integrity Forum (RIIF)

- Coordinate regional monitoring of AI-driven disinformation.
- Establish a shared repository of case studies, alerts, and best practices.
- Organise annual regional roundtables on AI governance.
- Engage directly with the EU AI Office, Council of Europe, and NDI.
- Publish an annual monitoring report on AI and democracy in the Western Balkans, with RIIF as a key platform.

5.4 Address Gendered AI Harms

- Formally recognise online gender-based violence (including AI-generated content) as violence in law. Adopt the concept of gendered information disorders as a defined category encompassing disinformation, misinformation, harassment, and abusive narratives that disproportionately target women.
- Require political parties to adopt internal codes of conduct prohibiting sexist language during campaigns or public communication, actively condemning attacks on women members across party lines, appoint digital safety focal points, and provide technical and psychological support to women candidates exposed to harassment. Digital violence is really encouraged to be treated as a political integrity issue, not simply a personal matter.
- Establish dedicated law enforcement units for online gender-based crime.
- Mandate judicial training on AI-facilitated gender-based violence.
- Require mandatory gender equality impact assessments for all public-sector AI tools.
- Introduce penalties and removal obligations for AI-generated gendered violence content.
- Establish accessible reporting channels for women targeted by AI abuse.
- Create AI harm redress mechanisms accessible to women and minorities.

Please note: Two practical tools have already been produced through NDI and are available for immediate deployment across the region:

(1) A Practical Guide for Women in Politics: Addressing Gendered Information Disorders, Online Violence, and Cyber Threats (NDI/Our Secure Future, 2025); and

(2) Journalist Guidelines for Covering Women in Public Life: Ethical, Accurate, and Responsible Reporting in the Digital Age (NDI/DISICON, 2025).

Additionally, the unSOCIAL MEDIA II monitoring report (Democracy for Development, 2025) provides a replicable methodology for monitoring gendered online violence during elections that could be extended across the Western Balkans to develop a standardised methodology for auditing AI systems for gender bias and gendered harms, adapted to the Western Balkan context.

Recent and relevant Council of Europe Standards on Freedom of Expression:

- Recommendation CM/Rec(2026)4 of the Committee of Ministers to member States on online safety and empowerment of users and content creators and its Explanatory Memorandum
- Guidance Note on the Implications of Generative AI for Freedom of Expression
- Resisting Disinformation: 10 Building Blocks to Strengthen Information Integrity
- National Media and Information Literacy (MIL) Strategies – Practical Steps and Indicators
- Feasibility Study on Benefits and Challenges to Freedom of Expression in Immersive Realities
- Guidance note on countering the spread of online mis- and disinformation through fact-checking and platform design solutions
- Guidelines on the responsible implementation of artificial intelligence systems in journalism
- Recommendation CM/Rec(2022)11 on principles for media and communication governance
- Recommendation CM/Rec(2022)12 on electoral communication and media coverage of election campaigns
- Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression
- Recommendation CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age
- Guidance note on the prioritisation of public interest content online

The Council of Europe Framework Convention on Artificial Intelligence

- Convention text
- Explanatory Report

HUDERIA – Risk and impact assessment of AI systems

Steering Committee for New and Emerging Digital Technologies (CDNET)

The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) ("Convention 108") and Protocols