COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 13 February 2026 CDPC-AICL(2026)02

EUROPEAN COMMITTEE ON CRIME PROBLEMS (CDPC)

Working Group on Artificial Intelligence and Criminal Law (CDPC-AICL)

Discussion Paper on Deepfakes and Criminal Law

Draft prepared by the Expert Magistrate Alfonso Peralta Gutiérrez in co-operation with the CDPC Secretariat to the Working Group of Experts on Artificial Intelligence and Criminal Law (CDPC-AICL)

Criminal Law Secretariat

DGI-CDPC@coe.int / www.coe.int/cdpc

I. PREAMBLE - BACKGROUND AND MANDATE

The European Committee on Crime Problems (CDPC) has been entrusted by the Committee of Ministers with drafting a legal instrument on criminal liability related to the use of artificial intelligence. This mandate builds on the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225), which establishes common principles for AI governance while leaving space for national legislative diversity.

Within this framework, the Working Group on Artificial Intelligence and Criminal Law (CDPC-AICL) has examined the implications of AI for substantive criminal law, criminal liability, digital evidence and international cooperation. Following earlier feasibility work and a Mapping Study based on national responses, delegations agreed that the breadth of "AI and criminal law" required a more focused approach.

Deepfakes were identified by the majority as a priority topic for further normative reflection, given their rapid technological evolution, their cross-border dimension, and their increasing use in the commission or facilitation of criminal offences.

In January 2026, a targeted questionnaire was circulated to member States to collect views on the desirability, scope and possible form of a future international instrument concerning deepfakes and criminal law. Sixteen member States out of the total of nineteen CDPC-AICL national delegations submitted responses within the established timeframe. The present Discussion Paper builds on those responses and aims to structure further deliberations within the CDPC-AICL Working Group.

II. DEEPFAKE CONTEXT AND EMERGING RISKS

The development of generative artificial intelligence has fundamentally transformed the production of synthetic media. Contemporary deepfakes are no longer rudimentary manipulations easily identifiable by the human eye. High-quality AI systems now generate hyper-realistic audio, video and image content that may be indistinguishable from authentic recordings. Their accessibility through publicly available tools, messaging platforms and subscription-based services has significantly lowered the barrier to entry for malicious actors.

For the purposes of this Discussion Paper, "deepfake" refers to AI-generated or manipulated image, audio, video, or any other content that resembles existing persons, objects, places, entities, or events and would falsely appear to a person to be authentic or truthful (Art 3 (60) EU AI Act).

Deepfakes are not unlawful per se. They may serve artistic, creative, satirical or educational purposes. However, the mapping exercise and comparative developments indicate that deepfake technologies are increasingly used in connection with criminal conduct across several domains.

A. Sexual exploitation and synthetic sexual content

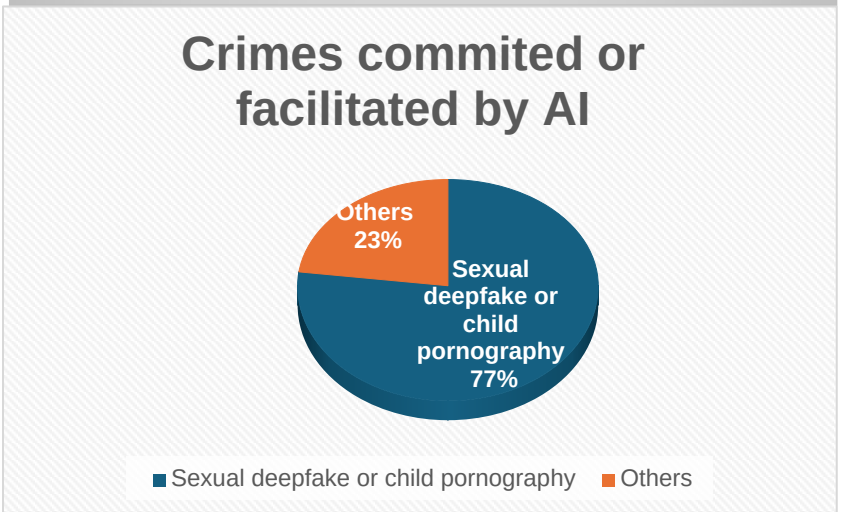
The most widespread criminal use of deepfakes concerns non-consensual intimate content and child sexual abuse material.

In Spain, in the so-called Almendralejo case, minors used a generative AI application accessed via Telegram to digitally "undress" photographs of girls taken from social media. The victims' faces were superimposed onto nude bodies and the manipulated images were circulated via messaging groups. Although criminal convictions were obtained, the case exposed interpretative difficulties regarding the qualification of synthetic nude images under existing pornography definitions.

Similarly, in Denmark, law enforcement authorities identified a producer distributing large quantities of AI-generated child sexual abuse material1 through a subscription-based model. The investigation revealed tens of thousands of synthetic images. This case raised significant legal questions concerning whether fully artificial depictions of minors fall within traditional definitions of child sexual abuse material, which in some jurisdictions presuppose the involvement of a real child.

The scale, speed and anonymity of such production considerably amplify the harm suffered by victims and complicate removal and enforcement efforts.

According to the Mapping Study based on the Compilation of National Responses to the CDPC-AICL Questionnaire on Artificial Intelligence and Criminal Liability (2025), 50% of the member States that responded reported case law or criminal investigations in their jurisdictions involving crimes committed or facilitated by AI. Notably, almost 80% of the reported offences were related to sexual deepfakes or child sexual abuse material. This data suggests that sexual exploitation currently represents the most prominent criminal manifestation of deepfake technology among responding States.



[1] See, *inter alia*, the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS No. 201) and the Convention on Cybercrime (CETS No. 185). In some jurisdictions, criminal definitions of child sexual abuse material presuppose the involvement of a real minor, which may raise interpretative questions where the content is entirely AI-generated.

B. Fraud, impersonation and large-scale economic deception

Deepfake technology has also been used in increasingly sophisticated fraud schemes.

In one high-profile case in 2024, the British engineering company Arup[2] reported losses exceeding USD 25 million after an employee participated in a video conference in which AI-generated representations of senior executives requested urgent financial transfers. The synthetic participants convincingly replicated the appearance and voices of real corporate officers.

Earlier cases in the United Kingdom involved AI-based voice cloning to impersonate a chief executive and induce the transfer of hundreds of thousands of dollars to foreign accounts. Similar frauds have been reported in Italy, where criminals used AI-generated voice messages impersonating children or grandchildren in distress to obtain banking information and one-time authentication codes from victims.

These cases demonstrate how deepfakes significantly enhance traditional fraud mechanisms by increasing credibility, emotional manipulation and operational scalability.

C. Democratic processes, public order and information manipulation

Deepfakes also present risks to democratic institutions and public stability.

During the armed conflict in Ukraine, a fabricated video circulated online purporting to show President Volodymyr Zelensky calling on Ukrainian soldiers to surrender. Although rapidly identified as false, the video illustrated the potential of AI-generated content to undermine morale and create confusion during crisis situations[3].

In France, a fabricated video announcing a coup circulated widely on social media platforms before removal. In the United States, a manipulated image depicting an explosion near the Pentagon briefly triggered stock market fluctuations before being debunked[4].

More broadly, coordinated disinformation campaigns have combined cloned media websites with synthetic images and videos to disseminate false narratives in sensitive electoral contexts.

Such conduct may fall within existing offences relating to electoral integrity, public order, market manipulation or national security. However, the transnational dissemination and rapid viral spread of deepfakes complicate jurisdictional and enforcement responses.

D. Administration of justice and evidentiary integrity

The increasing realism of deepfakes also raises concerns regarding judicial proceedings and evidentiary reliability.

Courts have begun to encounter AI-generated or manipulated audiovisual material submitted as evidence. In one case before a court in California, a video presented in a housing

[2] MAGRAMO, K. (2024, May 17). British engineering giant Arup revealed as $25 million deepfake scam victim. CNN. Available at: https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk

[3] ALLYN, B. (2022, March 16). Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn. NPR. Available at: https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

[4] O'SULLIVAN, D., & PASSANTINO, J. (2023, May 23). Verified Twitter accounts share fake image of ‘explosion’ near Pentagon, causing confusion. CNN. Available at: https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/index.html

dispute was identified by the judge as having been produced using generative AI, prompting questions regarding authenticity and admissibility[5].

The potential for fabricated audio or video recordings to create false alibis, intimidate witnesses or discredit proceedings underscores the need for enhanced forensic capacities and digital literacy within judicial systems.

E. Hybrid threats and national security considerations

In the broader geopolitical context, deepfakes have been incorporated into hybrid strategies combining disinformation, impersonation and digital manipulation.

Operations such as the "Doppelgänger" campaign have involved the cloning of established media outlets' websites and the dissemination of manipulated images and synthetic narratives designed to influence public opinion in politically sensitive periods[6].

In contexts of armed conflict, AI-generated content may be used to fabricate evidence of war crimes, impersonate officials or incite unrest. The attribution of such acts is particularly complex where production, hosting and dissemination occur across multiple jurisdictions.

F. Concluding observations

The above examples demonstrate that deepfakes intersect with a wide range of criminal law domains, including sexual exploitation, fraud, market manipulation, electoral interference, obstruction of justice and national security.

In many instances, existing technology-neutral criminal provisions may apply to conduct facilitated by deepfake technologies. However, the mapping exercise and national responses suggest uncertainties regarding synthetic minors, large-scale automated dissemination, evidentiary authentication, jurisdiction and cross-border cooperation.

These developments raise the question whether existing international instruments sufficiently address the specific risks posed by deepfake technologies, or whether further normative clarification, coordination or guidance at international level may be warranted.

III. COMPARATIVE REGULATION OF DEEPFAKES

Comparative developments indicate that a growing number of jurisdictions have introduced, or are considering introducing, legal measures addressing deepfakes and AI-enabled criminal conduct. These approaches differ significantly in scope, legislative technique and normative ambition.

Broadly speaking, current regulatory responses may be grouped into five models.

A. Specific criminalisation of non-consensual sexual deepfakes

Several jurisdictions have adopted targeted criminal offences addressing the creation or dissemination of non-consensual intimate synthetic content.

In the United States, the TAKE IT DOWN Act[7] criminalises the non-consensual online publication of intimate visual depictions, including computer-generated content, and imposes

[5] PERLO, J. (2025, November 18). AI-generated evidence is showing up in court. Judges say they're not ready. NBC News. Available at: https://www.nbcnews.com/tech/tech-news/ai-generated-evidence-deepfake-use-law-judges-object-rcna235976

[6] AGUILAR ANTONIO, J. M. (2025). Use of Artificial Intelligence by High-Risk Criminal Networks. Expertise France. Available at: https://doi.org/10.5281/zenodo.16750778

[7] Available at: https://www.congress.gov/bill/119th-congress/senate-bill/146/text

removal obligations on certain platforms. At state level, legislation such as California's AB 1836 extends post-mortem personality rights to AI-generated digital replicas[8].

South Korea has strengthened its Deepfake Sexual Crime Prevention Act, criminalising the editing, possession and dissemination of sexually explicit synthetic content and introducing enhanced victim protection mechanisms[9].

Australia is considering amendments to its Criminal Code introducing offences for transmitting non-consensual sexual deepfake material through digital services[10].

In Latin America, Mexico's Olimpia Law and Argentina's Law 27,736 recognise the non-consensual dissemination of intimate content, including AI-manipulated material, as a form of digital or gender-based violence, punishable by criminal sanctions.

These approaches focus primarily on sexual integrity, privacy and gender-based harm.

B. Aggravating circumstances and technology-neutral approaches

Other jurisdictions have chosen to integrate artificial intelligence into existing criminal frameworks through aggravating circumstances or AI-specific offences.

Peru has amended its Penal Code and Cybercrime Law to recognise the use of artificial intelligence in the commission of certain offences as an aggravating factor. [11]

In Europe, Italy's Bill 1146/2024 proposes both a general aggravating circumstance for crimes committed using AI systems and specific AI-related offences, including manipulated content and AI-facilitated fraud. The draft also addresses the liability of operators and other actors involved in the deployment of AI systems, based on traditional principles of intent and negligence. [12]

Under this model, AI is treated either as a method of commission enhancing culpability, or as a distinct technological dimension requiring specific offence definitions.

C. Electoral integrity and public order protections

Some States have introduced targeted provisions addressing the use of deepfake technology in democratic processes.

Latvia amended its Criminal Law in 2024 to introduce offences penalising the use of deepfake technology to influence elections or the appointment of public officials. [13]

In several jurisdictions, discussions are ongoing regarding the risks posed by synthetic media to electoral integrity, public order and institutional trust, particularly in light of coordinated disinformation campaigns and hybrid threats.

These measures reflect a recognition that deepfakes may affect not only individual victims but also collective democratic interests.

---

[8] Available at : https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB1836

[9] Available at: https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=68812&type=part&key=9

[10] Available at: https://www.aph.gov.au/Parliamentary_Business/Bills_LEGislation/Bills_Search_Results/Result?bId=r7205

[11] Available at : https://leyes.congreso.gob.pe/Documentos/2021_2026/ADLP/Texto_Consolidado/32314-TXM.pdf

[12] COMPILATION OF NATIONAL RESPONSES TO THE CDPC-AICL QUESTIONNAIRE ON ARTIFICIAL INTELLIGENCE AND CRIMINAL LIABILITY - 2025.

[13] COMPILATION OF NATIONAL RESPONSES TO THE CDPC-AICL QUESTIONNAIRE ON ARTIFICIAL INTELLIGENCE AND CRIMINAL LIABILITY - 2025.

D. Civil law protections and image rights

In addition to criminal law responses, certain States have strengthened civil protection frameworks concerning honour, image and personal data.

Spain is currently reviewing draft legislation aimed at protecting minors in digital environments and reinforcing civil protection of honour, privacy and image rights in cases involving AI-generated or manipulated content.[14] The draft proposals clarify that consent to the use of an image in one context does not automatically extend to AI-generated uses in another context, and they introduce presumptions of harm in cases of unlawful interference.

Similarly, legislative developments in the United States[15] and other jurisdictions extend personality and image rights to cover AI-generated digital replicas.

These developments demonstrate that deepfakes raise not only criminal law issues but also broader questions of civil liability and personality protection.

E. Regulatory and transparency-based frameworks

In addition to criminal law responses, several jurisdictions and regional organisations have adopted regulatory approaches focusing on transparency obligations, provider responsibilities and preventive governance of AI systems.

The European Union's AI Act[16] defines deepfakes and requires deployers of AI systems generating or manipulating synthetic audio, image or video content to disclose that such content has been artificially generated or altered, subject to specific exceptions, including artistic or law enforcement contexts. Non-compliance may lead to administrative sanctions. However, the AI Act is primarily a market regulation instrument designed to promote the uptake of human-centric and trustworthy artificial intelligence while ensuring a high level of protection of health, safety and fundamental rights in relation to AI systems placed on the market.

On 2 February 2026, the Committee on Legal Affairs of the European Parliament submitted a draft opinion proposing an amendment to the AI Act to include, among the prohibited practices under Article 5, the use of AI systems capable of generating or manipulating sexualised audio, images or videos that facilitate the non-consensual dissemination of intimate or manipulated material. This initiative reflects increasing concern regarding the harmful potential of certain generative AI applications, particularly in the field of sexual exploitation.

Nevertheless, the AI Act does not constitute a criminal law instrument. It does not apply where a criminal offence is perpetrated, facilitated or intensified through the use of AI systems, nor does it apply to natural persons using AI systems in the course of purely personal, non-professional activities. Its enforcement mechanisms rely primarily on administrative and monetary penalties rather than criminal sanctions. While regulatory classification of certain AI systems as prohibited practices may contribute to identifying particularly high-risk or harmful uses of AI, it does not in itself establish a framework of criminal liability.

Beyond the EU framework, China has adopted specific regulations governing deep synthesis services and generative AI systems, imposing obligations on service providers and establishing

[14] See Draft Organic Law on the Protection of Minors in Digital Environments and proposed amendments to the Organic Law on Civil Protection of the Right to Honour, Personal and Family Privacy and One's Own Image (Spain, pending legislative review).

[15] See, for example, California Assembly Bill 1836 (effective 1 January 2025), extending post-mortem personality rights to AI-generated digital replicas.

[16] See Regulation (EU) 2024/… on Artificial Intelligence (AI Act), Art. 3 (definition of deepfake) and Art. 50 (transparency obligations).

administrative enforcement mechanisms aimed at preventing misuse and maintaining public order.[17]

More broadly, regulatory instruments such as the Digital Services Act focus on platform responsibilities, content moderation and due diligence obligations rather than on substantive criminalisation. [18]

These developments demonstrate a growing reliance on transparency, labelling and provider accountability as preventive tools. However, they operate primarily in the domains of market regulation and administrative enforcement, leaving open the question of whether and to what extent criminal law harmonisation may be required in relation to malicious uses of deepfake technologies.

F. Emerging regional and international initiatives

At regional level, the Model Law on AI and Crime[19] developed within the EU-funded EL PACCTO 2.0 programme proposes a framework for criminalising AI-assisted offences and strengthening procedural cooperation in Latin America and the Caribbean. It addresses both substantive offences and investigative measures.

Such initiatives illustrate increasing international recognition of AI-enabled criminal risks, although approaches remain diverse and non-uniform.

G. Concluding observations

Comparative analysis reveals no single dominant regulatory model. Current approaches range from:

- specific criminalisation of sexual deepfakes,
- integration of AI as an aggravating circumstance,
- targeted electoral protections,
- reinforcement of civil image and personality rights,
- transparency-based regulatory frameworks.

While a growing number of jurisdictions have adopted or proposed measures explicitly addressing deepfakes, others rely on technology-neutral criminal provisions and consider existing frameworks sufficient.

This diversity suggests that any potential international instrument would need to carefully assess whether additional guidance at Council of Europe level could provide added value in light of existing criminal and cybercrime conventions.

IV. RESULTS OF THE QUESTIONNAIRE ON A POTENTIAL INSTRUMENT CONCERNING DEEPFAKES

The questionnaire, circulated in January 2026 to the 19 national delegations represented within the CDPC-AICL Working Group, received 16 responses, corresponding to a participation rate of approximately 84.2%.

[17] Available at: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[18] REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

[19] CASSUTO, T., PERALTA, A., & VELASCO, C. (2025). Model Law on Artificial Intelligence and Crime (1.0). Zenodo. https://doi.org/10.5281/zenodo.17281296

Accordingly, the results of the present questionnaire can be regarded as highly representative of the positions within the Working Group, which was specifically mandated by the CDPC to elaborate proposals in this field.

Countries that have answered the questionnaire:

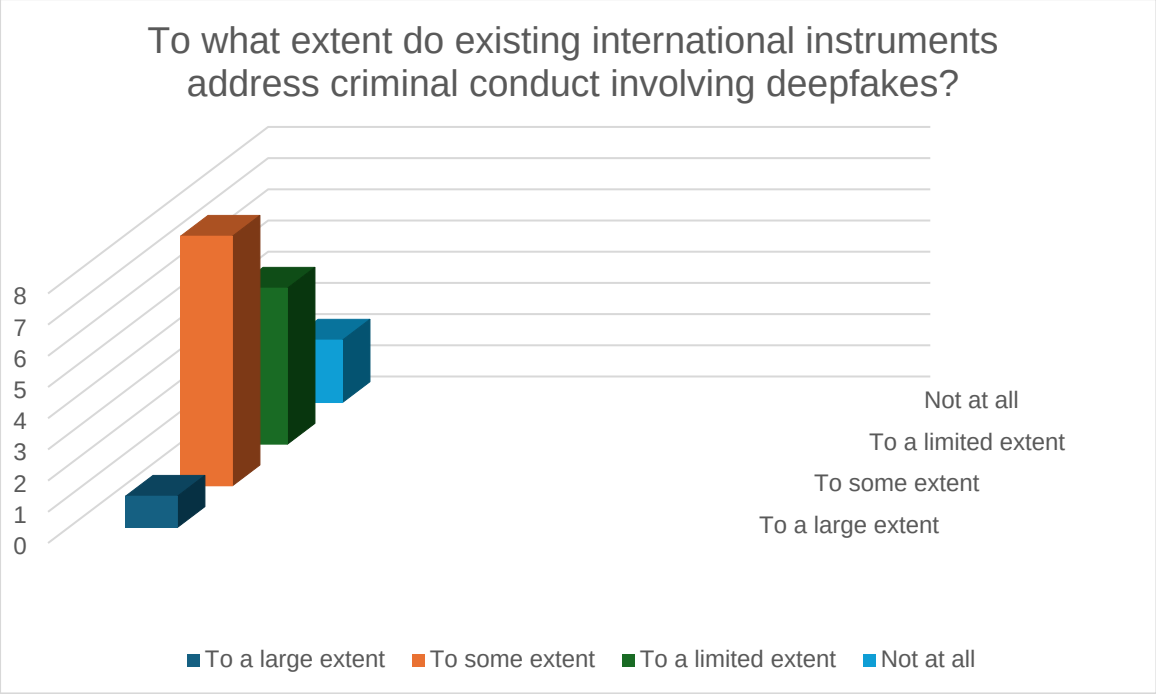
Austria	Malta	Slovenia
Belgium	Netherlands	Sweden
Czechia	North Macedonia	Ukraine
Finland	Norway	United Kingdom
France	Latvia	
Germany	Slovakia	

Part I. General approach

1. To what extent do existing international instruments address criminal conduct involving deepfakes?

The responses show no uniform assessment:
1 State: to a large extent
8 States: to some extent
5 States: to a limited extent
2 States: not at all

A majority therefore considers that existing instruments provide only partial coverage. However, several delegations emphasised the need for further analysis to identify concrete gaps before developing a new instrument.



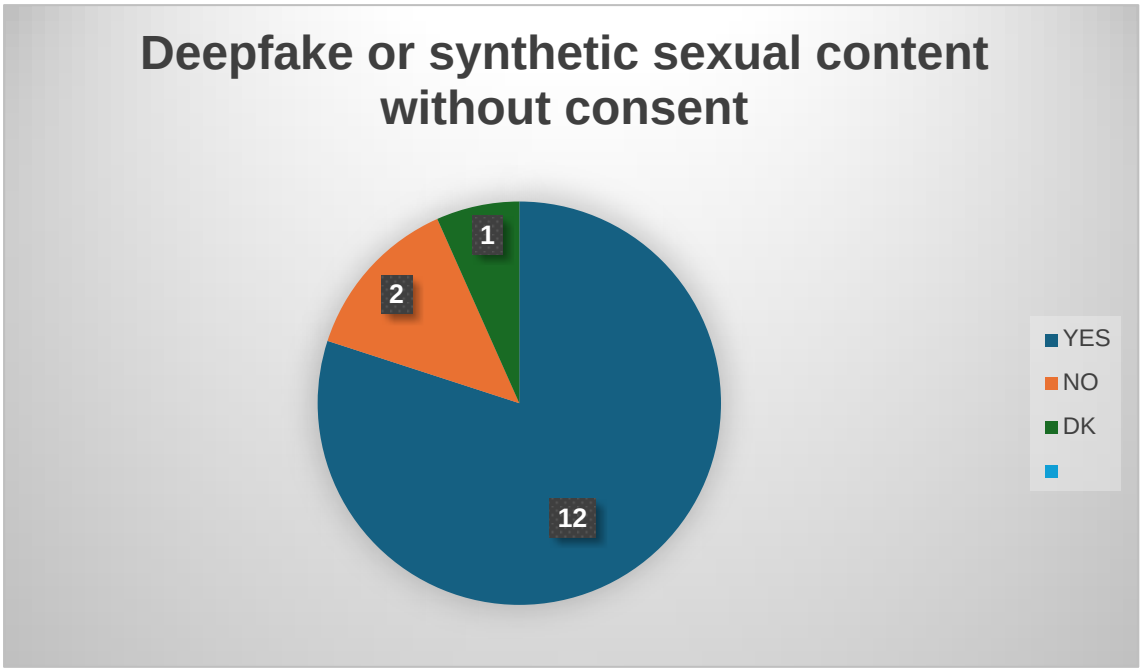
Part II. Substantive criminal law issues related to deepfakes

2. Do you agree that the intentional creation, distribution or making available of deepfake or synthetic sexual content, without the consent of the person depicted and with intent to cause harm or obtain a benefit, should be addressed in a potential international instrument?
☐ Yes

☐ No

A strong majority supported addressing the intentional creation or dissemination of non-consensual sexual deepfake content in a potential instrument. Only two States expressed opposition, while one delegation indicated conditional support depending on the instrument's binding nature.

This area represents the clearest zone of convergence.



3. Do you agree that the creation, possession or dissemination of deepfake material representing child sexual abuse or sexual exploitation, including where the minor is simulated, should be addressed in a potential international instrument?

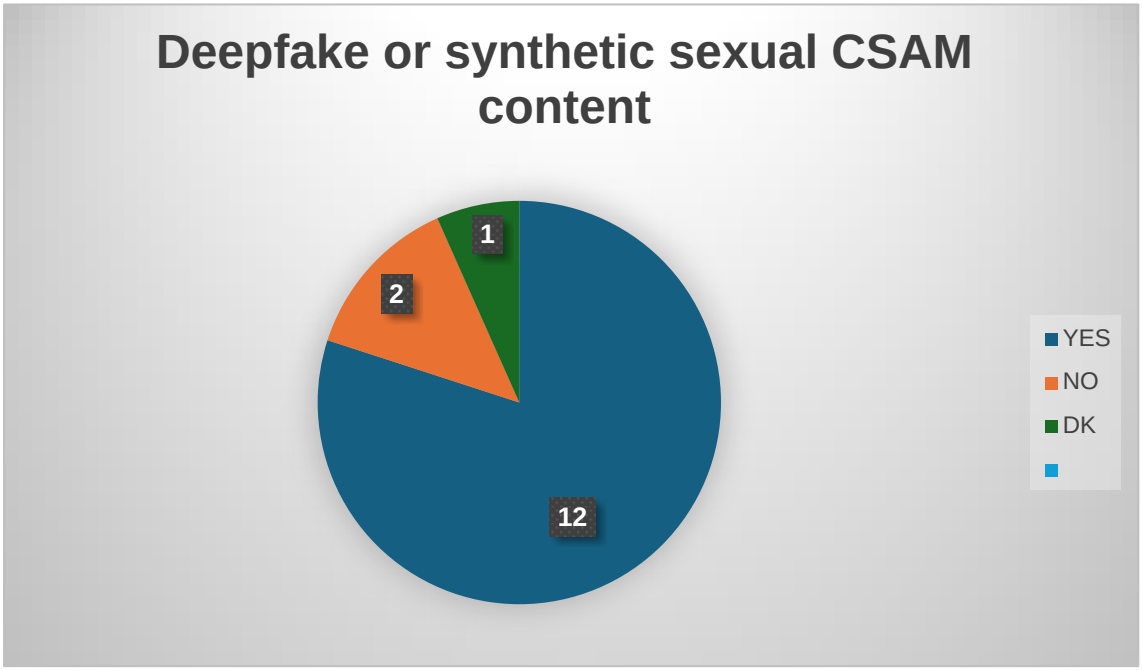
☐ Yes

☐ No

Similarly, a large majority of responding States supported addressing the creation, possession or dissemination of deepfake material representing child sexual abuse, including where the minor is simulated.

Two States opposed inclusion. One delegation emphasised that any new instrument should avoid overlap with existing frameworks such as the Lanzarote Convention and the Convention on Cybercrime.

Overall, there is broad support in this area, subject to concerns regarding duplication of existing child protection instruments.



4. Do you agree that the use of deepfake audio, video or synthetic identity impersonation, generated through artificial intelligence, for the purpose of committing child grooming or facilitating human trafficking, including by misrepresenting one's identity, age or other personal characteristics, should be addressed in a potential international instrument?

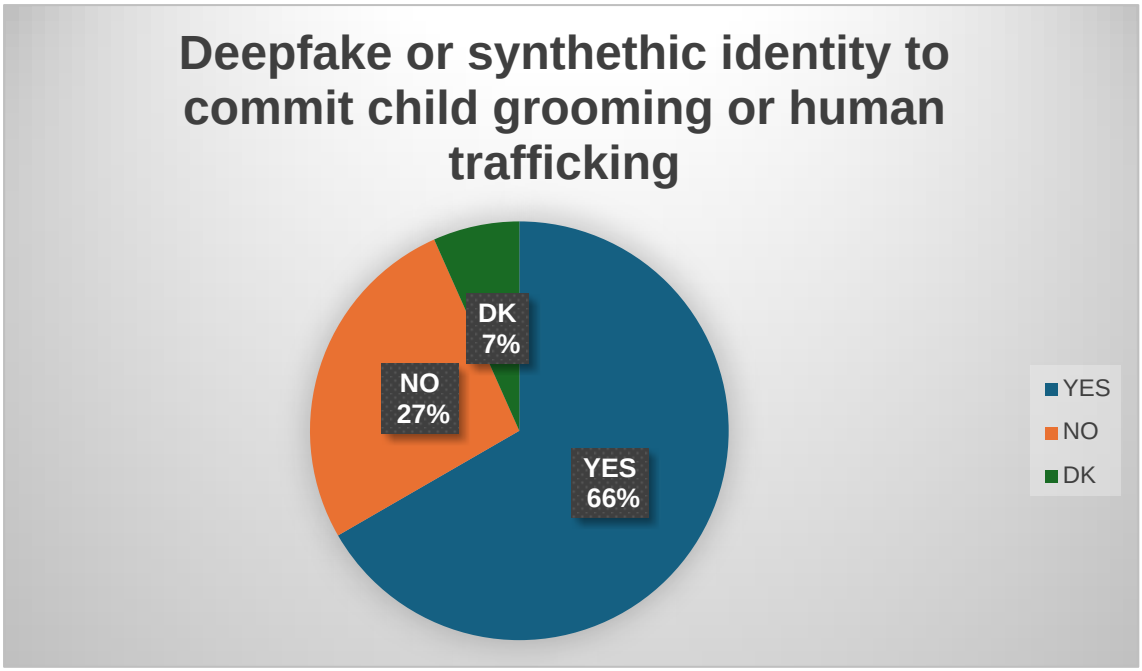
☐ *Yes*

☐ *No*

Ten responding States supported addressing the use of deepfake audio, video or synthetic identity impersonation for the purpose of committing child grooming or facilitating human trafficking.

Four States did not support inclusion. One delegation reiterated concerns regarding potential overlap with existing international conventions addressing child protection and cybercrime.

Support in this area is substantial but less consolidated than in relation to sexual deepfakes and synthetic child sexual abuse material.



5. Do you agree that the malicious creation or dissemination of deepfakes intended to seriously harm a person's honour, reputation or image should be addressed in a potential international instrument?

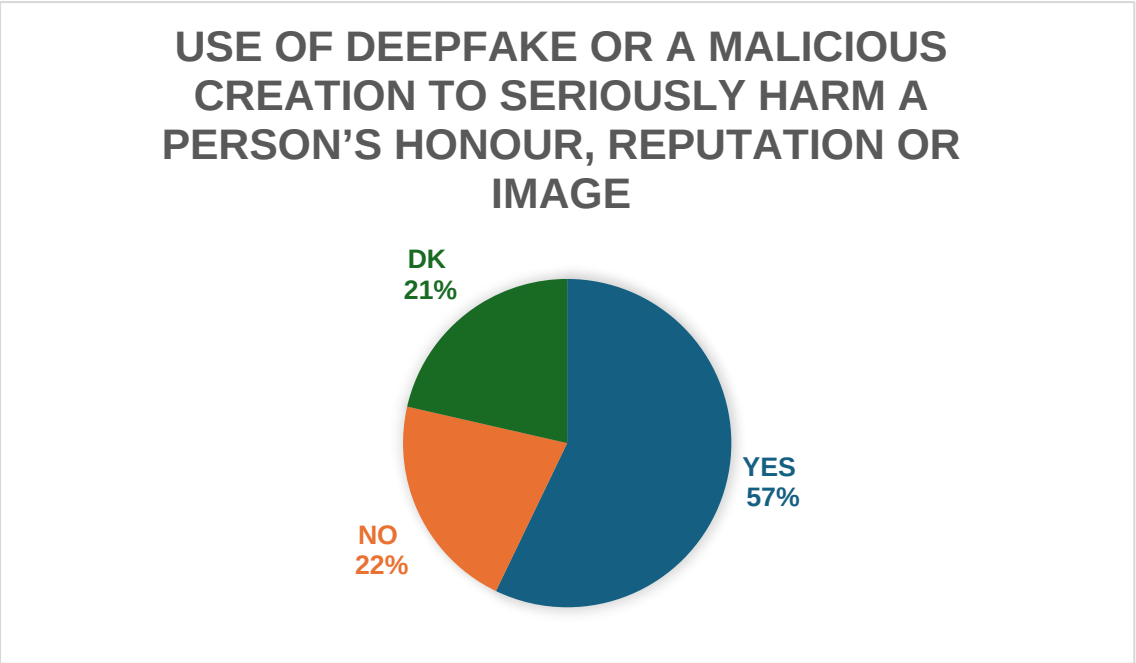
☐ *Yes*

☐ *No*

Eight responding States supported addressing the malicious creation or dissemination of deepfakes intended to seriously harm a person's honour, reputation or image.

Several States did not support inclusion. One delegation considered that the notion was too broad to be incorporated into an international criminal instrument without risking interference with freedom of expression. Another delegation indicated that it was difficult to provide a conclusive response without further clarification of scope.

This topic appears more politically and conceptually sensitive, particularly in light of freedom of expression considerations.



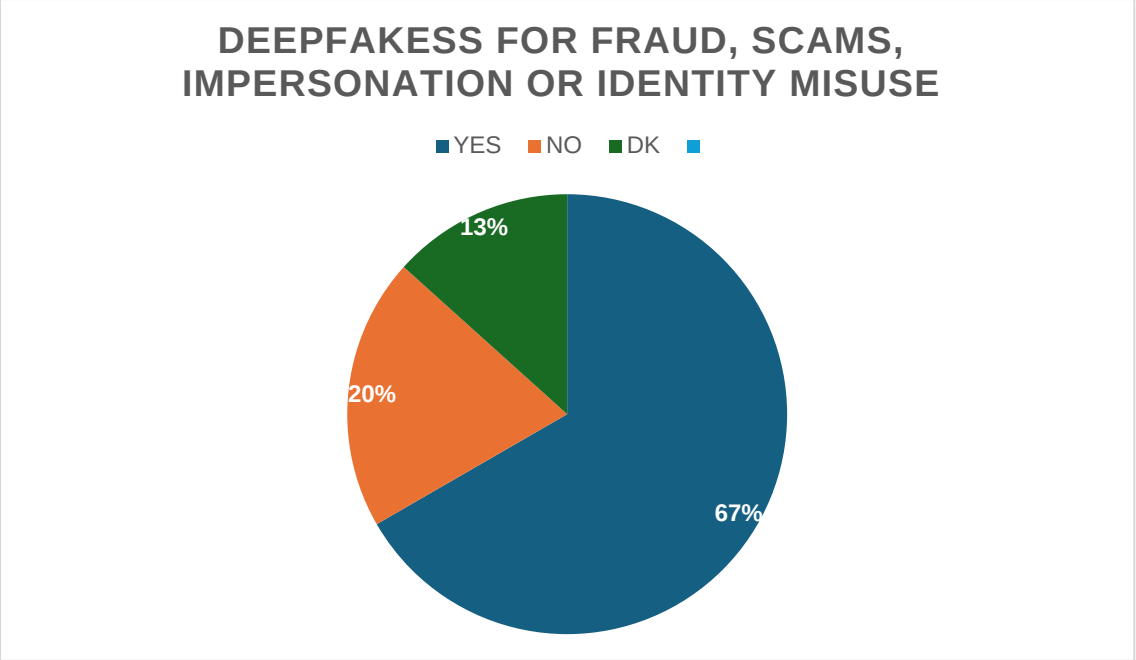
6. Do you agree that the use of deepfakes for purposes such as fraud, scams, impersonation or identity misuse should be addressed in a potential international instrument?
☐ Yes

☐ No

A majority of ten responding States supported addressing the use of deepfakes for purposes such as fraud, scams, impersonation or identity misuse within a potential international instrument.

Three States did not support inclusion of this area. In addition, one delegation stressed that any future instrument should avoid overlapping with existing international conventions, while another delegation indicated that it was difficult to provide a definitive response without clearer definitions of the underlying offences, such as fraud or identity misuse, and without further clarification of the instrument's scope.

The responses suggest that, although there is general recognition that deepfakes may facilitate fraud and impersonation schemes, views diverge as to whether new international standards are necessary, particularly in light of existing technology-neutral criminal provisions addressing fraud and deception.



7. Do you agree that the use of deepfakes to intentionally mislead the public in ways that seriously affect democratic processes, public order or security should be addressed in a potential international instrument?

☐ *Yes*

☐ *No*

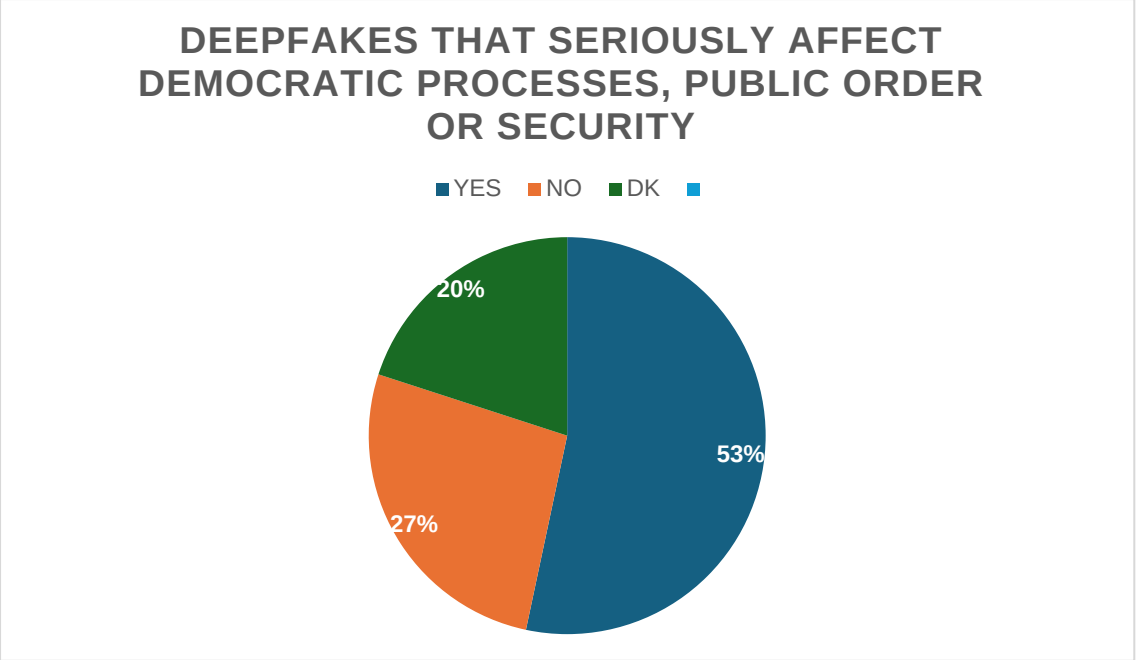
Member States expressed the most divergent views in relation to the possible inclusion of deepfakes used to intentionally mislead the public in ways that seriously affect democratic processes, public order or security.

Eight responding States supported addressing this issue within a potential international instrument. Three States did not support its inclusion.

In addition, one delegation considered that the notion was too broad to be incorporated into an international criminal instrument without risking interference with freedom of expression. Another delegation indicated that a clearer legal distinction would be required, particularly between conduct affecting democratic processes and conduct relating to public order or public security, before taking a definitive position.

Two further delegations adopted a cautious stance, stating that, at this stage, they were not in a position to provide a definitive answer.

Overall, the responses reveal significant conceptual and political sensitivity in this area. Certain malicious uses of deepfakes may raise concerns in relation to electoral integrity or public security. Any potential normative response in this field would require careful consideration of existing national offences and safeguards for freedom of expression.



Part III. Procedural and cooperation issues

8. Do you consider that a potential international instrument on deepfakes should include provisions on one or more of the following issues?

☐ *Criminal liability*

☐ *Jurisdiction*

☐ *Procedural measures*

☐ *Digital evidence, including issues of authenticity and admissibility ç*

☐ *International cooperation, including mutual legal assistance çç*

☐ *Cooperation with relevant online platforms*

☐ *Other closely related aspects (please specify):*

Ten member States supported the inclusion of provisions relating to criminal liability.

Seven member States expressed support for including provisions on jurisdiction. The same number supported addressing digital evidence, including issues of authenticity and admissibility, as well as international cooperation and mutual legal assistance.

Four member States supported the inclusion of procedural measures. An equal number indicated support for provisions concerning cooperation with relevant online platforms.

In addition, certain delegations proposed the inclusion of further elements. Slovakia suggested addressing the liability of service providers and the establishment of data retention periods for computer data. Ukraine proposed including provisions concerning moral damage, including calculation and compensation mechanisms; preventive measures aimed at early detection and prevention of abuse; and mandatory labelling of AI-generated content by persons who generate or use such content.

One delegation, the United Kingdom, indicated that it did not consider additional provisions to be necessary.

Overall, the responses indicate stronger support for including core criminal liability and cooperation elements, while procedural measures and platform cooperation received more limited endorsement.

9. Do you consider that a potential international instrument on deepfakes should address the use of deepfake audio, video or synthetic content for the purpose of manipulating or fabricating evidence in criminal proceedings, including safeguards to ensure the authenticity and reliability of digital evidence?

☐ *Yes*

☐ *No*

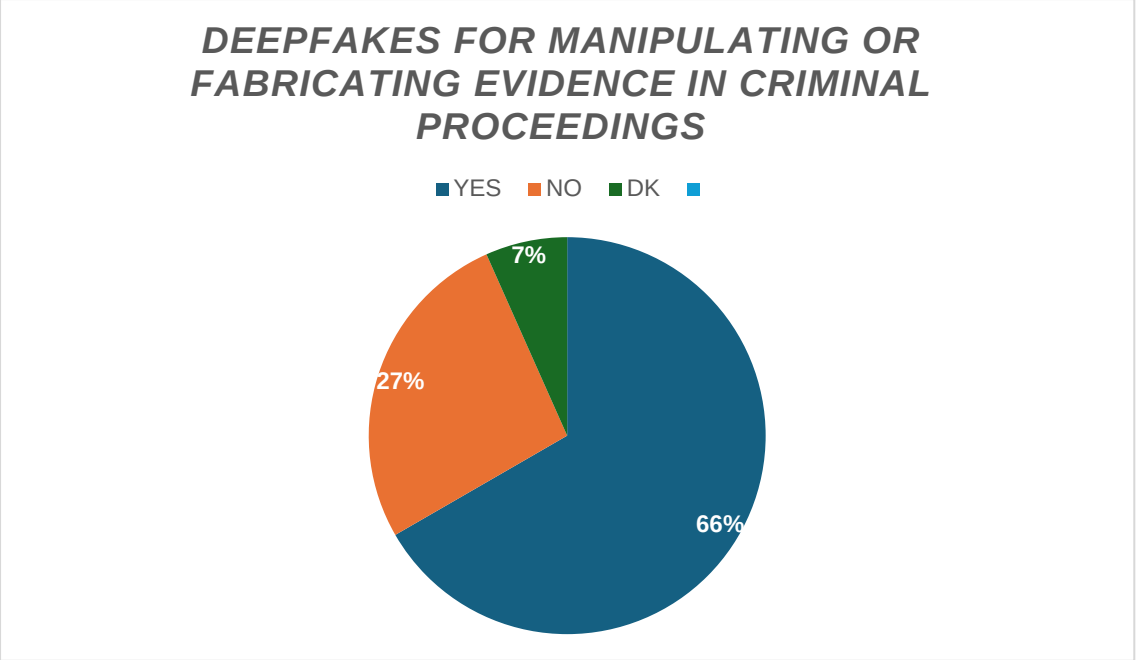
Ten member States responded in the affirmative.

Germany, the Netherlands, Czechia and the United Kingdom did not support inclusion of specific provisions in this area.

France expressed support, subject to the condition that any such instrument be non-binding in nature.

Sweden indicated that, at this stage, it was not in a position to provide a definitive answer.

The responses reflect broad recognition of the potential evidentiary risks posed by deepfake technologies, while also revealing caution among several States regarding the necessity and form of international regulation in this field.



Part IV. Support for further normative work

10. Based on the unanimously agreed focus on deepfakes, do you support the development of an international instrument addressing deepfakes and criminal law, as set out in the current ToRs of the Council of Europe Committee on Crime Problems/CDPC adopted by the Committee of Ministers?

☐ *Yes*

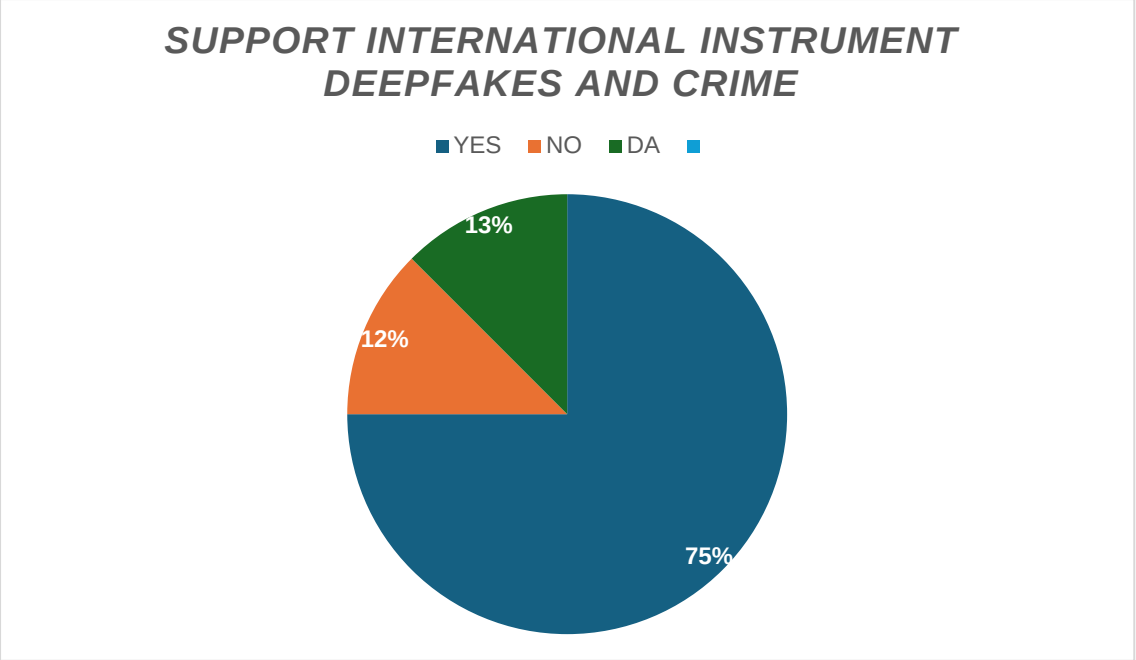
☐ *No*

Twelve member States expressed support for the development of such an instrument.

Germany and the United Kingdom indicated that they do not support pursuing an international instrument in this field.

Two member States did not provide a response to this question.

The responses indicate majority support for continued normative work within the CDPC framework.



Part V. Form of the instrument

11. If you answered "Yes" to the previous question, which form of international instrument would you consider most appropriate?

☐ *Convention*

☐ *Additional Protocol to an existing Convention*

☐ *Recommendation of the Committee of Ministers of the Council of Europe*

Three member States indicated that a Convention would be the most appropriate form.

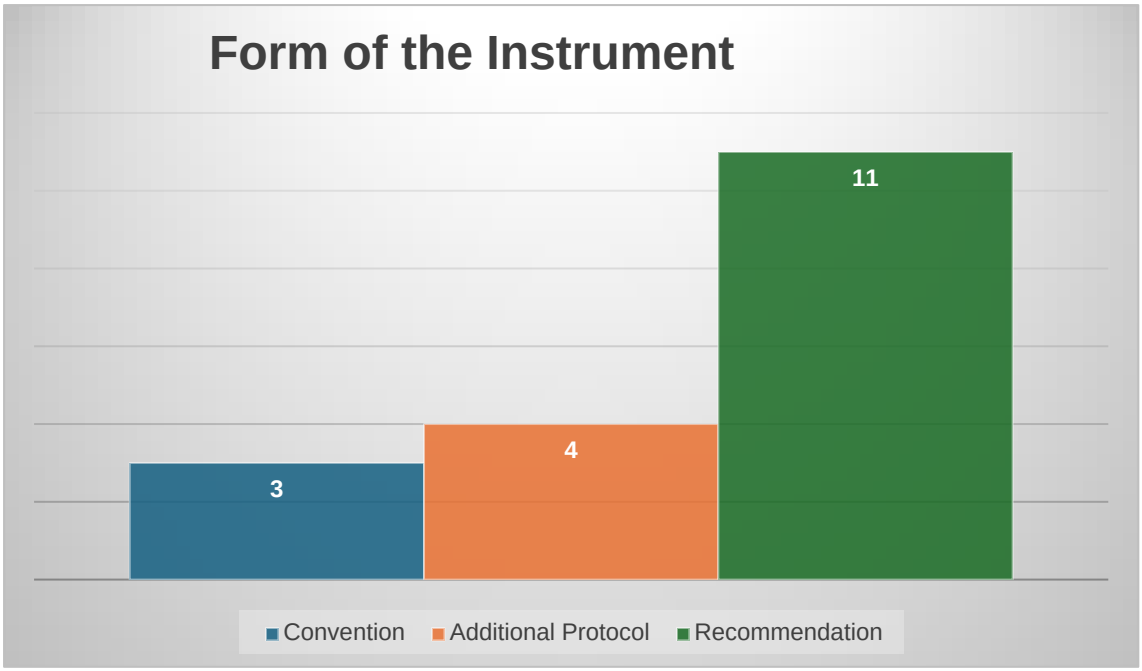
Four member States expressed a preference for an Additional Protocol to an existing convention.

A larger number of responding States, eleven in total, indicated that a Recommendation of the Committee of Ministers would be the most appropriate instrument.

Slovenia expressed openness to all three options.

Sweden indicated that, if pursued, the instrument should take the form of a Recommendation, while also emphasising that the need for an international instrument in this field requires further analysis.

The responses suggest stronger support for a non-binding instrument in the form of a Recommendation than for a binding convention or additional protocol at this stage.



Part VI. Additional observations

12. Without prejudice to the agreed focus on deepfakes, do you wish to signal any other closely related issue that you consider may merit future examination? (Open-ended, optional)

The Netherlands appreciate this exercition to decide on the scope of the instrument and they recognize the value of this structured approach. At the same, the questionnaire is constrained to yes/no-responses which sometimes makes it complex to give a nuanced view on a particular topic. Mainly, under Part 2 of the questions because it covers various forms of deepfakes. Therefore, a 'no' in this questionnaire does not mean they do not see any reason to address it in the international instrument, but the topic still needs some clarification in which way it will be addressed and to what extent it is already covered by other relevant conventions and international instruments (as well as in development).

At this stage, we primarily see scope for a recommendation. For the Netherlands, criminal prohibitions are formulated in a technology-neutral manner: the way content (including deepfakes) is produced does not determine the qualification of the prohibited conduct. This aligns with the UN Cybercrime Convention (and the 2nd Protocol), where the relevant behaviors are already criminalized and deepfakes are not excluded on the basis of the technology used. In addition, the CoE has recently begun the exploration of a legal instrument to counter FIMI.

To demonstrate the added value of the instrument, we think it would be necessary to clarify to what extent the existing instruments cover each topic set out above, and, conversely, where gaps persist that would justify additional information in this additional instrument.

Norway's primary position is that they have not yet identified an urgent need for the development of a new instrument on AI and criminal law. In their answer to the present questionnaire, they provide input on what a new instrument should encompass, assuming that there is agreement to move forward with it. They would also like to note that it would be useful to engage in a more thorough analysis of existing international instruments with the purpose of mapping out a potential gap and ensure that any new provisions do not duplicate existing

instruments, for example in the field of international cooperation in criminal matters. Furthermore, they reiterate the importance of ensuring that the work carried out in this Committee is well coordinated with – and does not duplicate – other CoE initiatives, for instance in the T-CY Working Group on AI and the PC-FIMI. As to the scope of a potential new instrument, the topic of deepfakes appears to be somewhat narrow for a convention or additional protocol, which might be more suitable for a recommendation. If a legally binding instrument were pursued, its scope should preferably be broader. A possible compromise could be a tiered approach, beginning with deepfakes and subsequently addressing other topics, if needed. Although there are advantages to developing a legally binding instrument, it may take considerable time before it could enter into force. Moreover, soft law instruments could facilitate cooperation with service providers and online platforms, as well as contribute to the development of a European consensus on best practices in this area.

Slovenia consider that other specific AI related criminal offences (i.e. AI systems specifically engineered for malicious purposes) and procedural aspects (e.g. safeguarding the right of defence – equality of arms) may merit future examination and be included in the international instrument.

Ukraine suggested to include procedural obligations of the parties regarding disclosure of information on the use of AI, digital competence of judges, prosecutors, investigators, experts and the use of deepfake technology during armed conflicts. They allegued that this technology poses a potential threat and can be used as a tool for information aggression. It can be used as a means of provocation and incitement to violence by creating fake appeals, spreading disinformation to demoralize the population and military personnel, and fabricating knowingly false evidence (audio and video materials) of events or war crimes.

In view of this, the object of legal regulation should be not only the individual activities of separate entities, but also targeted state policy on the development and distribution of content created using artificial intelligence technologies, particularly in cases where the production and dissemination of deepfakes is part of official state policy or an element of hybrid warfare.

Ukraine also propose to regulate the use of artificial intelligence platforms for materials (files, information) that have been uploaded by users to these platforms and contain data whose access is restricted by law.

United Kingdom has allegued that they agree that the misuse of deepfake technology is an important issue, but they do not believe that the framing of this questionnaire reflects a consensus among member States. The various delegations at the December CDPC-AICL meeting had different ideas on whether an instrument should be pursued at all and, if so, what topic any such instrument should cover.

Some delegations were in favour of including scams and fraud in scope of a future deliverable, as such crimes are very closely linked to deepfakes; others were in favour of further exploratory work on broader topics for inclusion (on the understanding that any future instrument would be non-binding); and others preferred a narrower scope or saw no need for an instrument at all.

The UK position remains that we do not see a need for an instrument in this space, and there needs a clear consensus on how to proceed before discussing the specifics of any future deliverable.

They note that there is already a large body of relevant Council of Europe instruments, some of which were developed after the initial agreement on the AICL deliverable. This includes the Framework Convention on AI, which the UK has signed, and the Recommendation on Technology-Facilitated Violence Against Women and Girl, recently approved by the CDPC, while the PC-FIMI's work is likely to cut across related topics. The Cybercrime Convention, which the UK has ratified, also provides an important framework in addressing AI risks.

V. POSSIBLE STRUCTURE AND CONTENT OF A FUTURE RECOMMENDATION ON CRIMINAL DEEPFAKES

This section outlines, for discussion purposes, a possible structure and key elements of a future draft Recommendation concerning deepfakes and criminal law. It is presented in light of the majority preference expressed by the responding national delegations within the CDPC-AICL Working Group, whose participation rate in the questionnaire reached over 84%, and therefore reflects a highly representative sample of positions within the Working Group.

The work is undertaken pursuant to the mandate conferred upon the CDPC by the Committee of Ministers to elaborate an international instrument in the field of artificial intelligence and criminal law. The CDPC is currently operating under this mandate. In light of the clear majority support expressed within the Working Group for continuing normative work, the present outline is intended to facilitate further discussions in execution of that mandate. The inclusion of this outline does not prejudge any decision regarding the adoption, scope or legal nature of a future instrument.

I. Purpose and Scope

A future Recommendation could:

- Clarify its objective as providing guidance to member States on addressing malicious uses of deepfake technologies in the context of criminal law.
- Define its scope as covering the intentional creation, dissemination or use of AI-generated or AI-manipulated synthetic content where such conduct facilitates or constitutes criminal offences.
- Explicitly exclude legitimate artistic, satirical, journalistic or research uses, subject to national law and Article 10 ECHR safeguards.

II. Definitions

The instrument could include a limited set of definitions, such as:

- “Deepfake” or “digital forgery”
- “AI-generated” and “AI-manipulated content”
- “Identifiable person”
- “Digital child sexual abuse material”

Careful drafting would be required to ensure compatibility with existing definitions under the Cybercrime Convention and the Lanzarote Convention.

III. General Principles

A Recommendation could set out guiding principles, including:

- Respect for human rights and fundamental freedoms
- Legality, legal certainty and foreseeability
- Necessity and proportionality in criminal law responses
- Technological neutrality where appropriate
- Avoidance of duplication with existing international instruments

IV. Substantive Criminal Law Considerations

In light of questionnaire responses, the instrument could address the following areas, with varying degrees of emphasis:

1. Non-consensual sexual deepfakes

Given the strong convergence among responding States, a Recommendation could:

- Encourage States to ensure that non-consensual sexually explicit synthetic content depicting identifiable persons is adequately addressed under criminal law.
- Clarify the treatment of AI-generated depictions of minors/children, where consistent with existing child protection frameworks.

2. Deepfakes and child grooming

In view of more nuanced support, the instrument could:

- Invite States to assess whether existing offences sufficiently cover the use of synthetic identity impersonation in child grooming or trafficking contexts.
- Emphasise coherence with the Lanzarote Convention.

3. Fraud, impersonation and identity misuse

The Recommendation could:

- Highlight the risks posed by deepfakes in facilitating fraud and impersonation.
- Encourage review of national legislation to ensure adequate coverage, whether through technology-neutral provisions or specific clarification.

4. Democratic processes and public order

Given the divergence of views, a cautious approach could:

- Acknowledge the potential impact of malicious deepfakes on democratic institutions.
- Invite States to examine whether existing offences provide sufficient protection.
- Reaffirm the need to safeguard freedom of expression.

V. Procedural and Evidentiary Issues

In light of relatively broad support in this area, a Recommendation could include:

- Guidance on assessing authenticity and reliability of digital evidence.
- Encouragement of forensic capacity-building.
- Promotion of digital literacy for judges, prosecutors and investigators.
- Consideration of disclosure obligations concerning AI-generated evidence.

VI. Jurisdiction and International Cooperation

The instrument could:

- Encourage review of jurisdictional frameworks in cross-border deepfake cases.
- Promote effective mutual legal assistance mechanisms.
- Ensure coherence with the Convention on Cybercrime and its Protocols.

VII. Prevention and Awareness

A Recommendation could also address:

- Public awareness initiatives.
- Responsible deployment and labelling of synthetic content.
- Cooperation with relevant stakeholders, including platforms and service providers.

VIII. Monitoring and Future Review

The instrument could foresee:

- Periodic review of technological developments.
- Assessment of whether further binding standards may be required in the future.