Discussion Paper No. 5

EU Network on the Prevention of Gender-Based and Domestic Violence

Artificial intelligence to prevent gender-based violence and domestic violence

1.1 Introduction: the two sides of AI

A previous meeting of the Network, in November 2024, discussed the prevention of non-consensual intimate images (NCII) abuse and cyber hate speech, which are considered structural and systemic problems, requiring a focus on root causes and investment in education on consent and professional training. Artificial intelligence (AI) is used to manipulate images, create messages of hatred, disseminate disinformation, reproduce gender-biases and exacerbate structural inequalities. AI can be exploited to produce **deep fakes**, which consist in the creation of convincing images, audio and video hoaxes. This phenomenon has serious psychological and social impacts on victims. The term describes 'both the technology and the resulting bogus content, and is a portmanteau of deep learning and fake.'[1] Deep fakes might "invent" art works or produce music, but they might be also used for political misinformation, fraud and ruining a person's reputation. Non-consensual deep fake is a form of image-based sexual abuse and AI-generated gender-based violence against women. As a matter of fact, in the majority of cases – studies show[2] – sexual deep fakes are forms of non-consensual pornography. A form of deep fake is deep nude – applications that create realistic nude images from photographs without the consent of individuals.

There is also **another side of artificial intelligence**: if, on the one hand, it can be the cause of violence and reproduce dynamics of subordination and oppression, on the other hand, it can also be a potential resource in tools designed to protect victims and reinforce consent (e.g. AI-driven content moderation, victim-support apps, and informed-consent verification systems), tools to reinforce rules like the Terms of Service on online platforms (e.g. through automatic AI-generated prompts sent to online users), and in detecting illicit content online. New apps and chatbots have been developed to support victims of GBV and DV, by providing a safe and anonymous way to recognise relationship patterns, collect evidence that can be used in potential future proceedings against the abuser, identify biases. As acknowledged also by the Council of Europe, 'when clear, non-biased rules and high-quality data are used, AI may in fact be less prone to bias than human decision-making. Indeed, it can also facilitate the detection of bias through its capacity to collect and analyse large quantities of data.'[3] It is on this positive potential of AI to prevent gender-based violence that this meeting of the Network will focus.

1.2 The Istanbul Convention and the role of AI in preventing and responding to GBV and DV

The Council of Europe Istanbul Convention on preventing and combating violence against women and domestic violence (Istanbul Convention) and its Explanatory Report do not refer expressly either to gender-

[1] https://www.techtarget.com/whatis/definition/deepfake
[2]https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/ https://theconversation.com/deepfake-porn-why-we-need-to-make-it-a-crime-to-create-it-not-just-share-it-227177
[3] Council Conclusions on Advancing Gender Equality in the AI-Driven Digital Age: 6th horizontal review of the implementation of the Beijing Platform for Action by the Member States and the EU institutions, Brussels, 19 June 2025.

based violence in the digital world or to the positive side of AI. However, some guidance, in terms of prevention, support to law enforcement and removal of abusive content, can be found in the first **General Recommendation** by GREVIO, the monitoring body of the Istanbul Convention, dedicated to the digital dimension of violence.[4] GREVIO recommended, for example, relevant to the EU network's discussion, to: promote the inclusion of **digital literacy** and online safety in formal curricula and at all levels of education; encourage the ICT sector and internet intermediaries, including **social media platforms**, to make an active effort to **avoid gender bias** in the design of smart products, mobile phone applications and video games, as well as the **development of artificial intelligence** and - respectively - to create internal **monitoring mechanisms** towards ensuring the inclusion of victim-centric perspectives as well as to advocate stronger awareness of the perspective and experiences of female users, in particular those exposed to or at risk of intersecting forms of discrimination.[5] The General Recommendation also suggested that intermediaries including ISPs, search engines and social media platforms should provide easily accessible **user guidance** to **flag abusive content** and request its removal.[6] With regard to digital violence, one recommendation concerned the need to 'increase capacity-building efforts for criminal justice and law-enforcement professionals to equip them with the necessary expertise and resources on how to use existing legal frameworks to address the digital dimension of violence against women, as well as to develop their **forensic capabilities** on the gathering and securing of electronic evidence without causing secondary victimisation and re-traumatisation of the victim.'[7]

1.3 The EU VAW Directive and the role of AI in preventing and responding to GBV and DV

As it is known, the major recent development in EU law is the Directive on combating violence against women and domestic violence, published in the Official Journal (OJ) on 24 May 2024 (VAW Directive).[8] The **VAW Directive** is based on Article 82(2) and Article 83(1) TFEU. The VAW Directive contains several provisions that are useful to understand **how AI might be used to prevent GBV and DV** and **to protect the victims** (e.g. AI-powered applications for victims), but also to conduct investigations. Prevention through digital literacy is also important to understand how to use apps and chatbots in the most efficient way.

Under Article 14 (1), 'Member States shall ensure that victims can report acts of violence against women or domestic violence to the competent authorities through accessible, easy-to-use, safe and readily available channels', which shall include, 'at least for the cybercrimes' under the scope of the VAW Directive, 'the possibility of **reporting online** or through other accessible and secure ICT, without prejudice to national procedural rules regarding formalising online reporting'. Also, the provision requires Member States to ensure 'the possibility to report online or through other accessible and secure ICT includes the possibility to submit evidence by the means set out in the first subparagraph, without prejudice to national procedural rules regarding formalising the submission of evidence'. With regard to investigations, the VAW Directive obliges Member States to ensure that 'persons, units or services investigating and prosecuting acts of violence against women or domestic violence have adequate expertise in those matters and have effective investigative tools at their disposal to effectively investigate and prosecute such acts, especially for the purpose of **gathering, analysing and securing electronic evidence** in cases of cybercrime as referred to in Articles 5 to 8'.

Concerning the possibility (States 'may') for Member States to issue guidelines for cases concerning violence against women or domestic violence for the competent authorities acting in criminal proceedings, including prosecutorial guidelines, Article 21 sets certain requirements including guidance on how to (b) '**gather and**

[4] GREVIO General Recommendation No. 1 on the digital dimension of violence against women adopted on 20 October 2021, GREVIO(2021)20.

[5] Ivi, para. 52.

[6] Ivi, para. 53.

[7] Ivi, para. 55.

[8] Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence, PE/33/2024/REV/1, OJ L, 2024/1385, 24.5.2024 (VAW Directive).

preserve relevant evidence, including online evidence'. Article 23, **Measures to remove certain online material,** provides the possibility for the competent authorities to issue binding legal orders to remove or to disable access to online publicly accessible material as referred to in Article 5(1), points (a) and (b), and Articles 7 and 8 VAW Directive, but **does not refer to the detection of this material.**

Article 29 VAW obliges States to provide state-wide telephone helplines, 24 hours a day and seven days a week, to provide information and advice to victims, and encourages them to 'also provide **helplines** as referred to in the first subparagraph through other secure and accessible ICT, including online applications'**.**

In terms of prevention, Article 34 (8) states that 'Member States shall ensure that such preventive measures include the development of **digital literacy** skills, including critical engagement with the digital world and critical thinking to enable users to identify and address cases of cyber violence, to seek support and to prevent its perpetration.

Member States shall foster **multidisciplinary and stakeholder cooperation**, including among relevant intermediary service providers and competent authorities, to develop and implement measures to address the cybercrimes referred to in Articles 5 to 8'.

Other relevant legal instruments must be mentioned while analysing the role that the AI can play in preventing GBV and DV.

a) **Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act).**[9] The AI Act lays down, among others, harmonised rules for the placing on the market, the putting into service, and the use of AI systems in the Union; prohibition of certain AI practices; specific requirements for high-risk AI systems and obligations for operators of such systems. According to Recital No. 27 of the Regulation, '**diversity, non-discrimination and fairness** means that AI systems are developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law.' The EU encourages a use of the AI that is founded on these principles. No specific provisions however address the use of AI for preventing GBV and DV, being the act not aimed at introducing rules for AI that is deemed minimal or no risk.[10] Furthermore, in the Regulation, reference to the negative impact of AI on women is scarce and disseminated in just a few provisions. The preamble acknowledges that AI systems might entail risks for fundamental rights, including gender equality, and that 'when improperly designed and used, such systems may be particularly intrusive and may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation'. However, despite the fact that this instrument might have positive implication on combating cyber gender-based violence, there is **no explicit recognition** that AI systems may constitute forms of gender-based violence that disproportionately affect women and girls (e.g. deep fake) and there is no recognition of the connection between the AI act and the VAW Directive.[11]

b) **Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services (DSA).**[12] The DSA centres on users' safety and rights, creating a safer online environment. Among its goals, it establishes rules on specific due diligence

[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), PE/24/2024/REV/1 *OJ L, 2024/1689*, 12.7.2024.

[10] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[11] S. De Vido, *EU Law in light of the Istanbul Convention: Legal Implications after the Accession*, EU Publications Office, 2025.

[12] 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), PE/30/2022/REV/1, *OJ L 277, 27.10.2022, pp. 1–102*.

obligations tailored to certain specific categories of providers of intermediary services. Therefore, the detection of illicit content is part of the risk assessment providers of very large online platforms must perform. Under Article 34, 'providers of very large online platforms and of very large online search engines shall diligently identify, analyse and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services.' Using a systemic interpretation that takes into account the EU legal system, one can argue that platforms should be able to identify risks of GBV that are present in their systems, including the algorithmic ones.

c) **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR).** When it comes to apps, chatbots, collection of evidence through digital systems, the protection of personal data becomes crucial. For example, under Article 5 GDPR, data must be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. The protection of the data victim of GBV and DV intersects many of the issues that will be discussed in the meeting of the EU Network and represents an ethical concern.

In line with the previous meetings, the purposes of this fifth meeting are summarised in the acronym MA.K.E: a) **MA**pping measures and approaches in the use of AI as preventive mechanism, looking into gaps and emerging needs; b) share **K**nowledge and tools for risk assessment and detection of vulnerabilities; c) **E**xchange existing good practices (apps, chatbots, etc.) and effective outcomes.

DAY 1

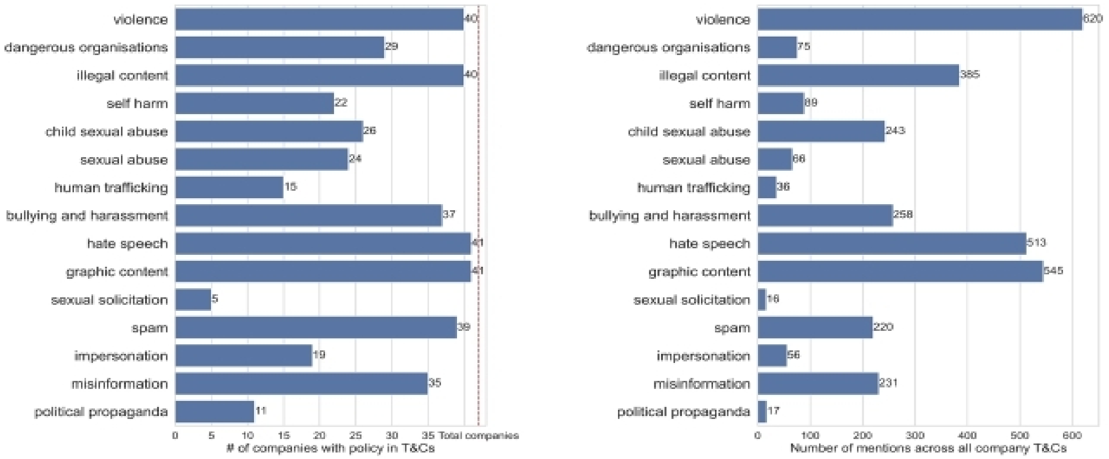
Panel 1 – Detecting illicit content online

There have been several surveys of **computational methods** to detect and address online harm.[13] However, how easy is it to detect illicit content online? Are there systems which can automate and augment content moderation processes at scale? Can humans be replaced by AI in identifying potential risks and abuses? Can GBV always be detected?

In a **study** conducted on 42 online platforms' Terms and Conditions (see table below taken from the study), the authors found that 'violence, graphic content, and hate speech are the most mentioned topics, demonstrating the significant attention paid to them as well as the importance of detecting policy violations regarding these topics. Sexual solicitation and political propaganda are the topics with the fewest mentions.'[14] Child sexual abuse is in the same range of mentions as misinformation, bullying and harassment, and spam while being covered by fewer platforms. In the study, 'violence' is very general, hence it can include GBV but also other forms of violence that are not related to gender.

[13] A. Arora and others, *Detecting Harmful Content on Online Platforms: What Platforms Need vs. Where Research Efforts Go*, ACM Computing Surveys, Volume 56, Issue 3, 72, 1 – 17 retrieved from https://dl.acm.org/doi/10.1145/3603399#fig2

[14] Ibid.



(a) Number of platform T&Cs mentioning each topic

(b) Number of mentions of each topic across all platform T&Cs

One possibility to detect illicit content online is to allow users (as many platforms do) to **manually report** inappropriate and (potentially) illegal content to the service providers for removal. However, this can cause further trauma to the victims who are responsible for reporting, and fear of retaliation which might discourage the use of the reporting itself.

One of the key methods to bypass the voluntary reporting is assessing uploaded content for matches to known problematic content.[15] **Perceptual hashing** – the use of a fingerprint algorithm that produces a snippet, hash or fingerprint of various forms of multimedia – detects illegal content, for example the sharing of NCII. As Somers explains:

> Hashing algorithms are used to extract a unique compact alphanumeric signature, also called a hash or the image's fingerprint, from an image. That newly generated fingerprint can then be compared against fingerprints of other photos that are uploaded to an online service. This makes it possible to actively search out known illegal content and take actions to remove it. For example, Microsoft uses PhotoDNA on its services to scan for illegal images. While hashing is a quick and reliable technique, its main problem is that it can only be used for seeking out known content. It is not an effective technique for identifying new and unknown content.[16]

AI is effective if based on strong underlying datasets.[17] The members of large and obscure online platforms and communities, associated with the 'Manosphere', embrace, spread, and cultivate false, stereotypical, harmful, and misogynistic perceptions and ideas about women, often revolving around women's 'nature', sexuality, bodily autonomy, and self-determination. They use codes and memes that do not necessarily match with existing databases and that might seem harmless at first sight.

It has also been reported that a mechanism of scanning data known as **client-side scanning** has been recently proposed by tech companies and governments as a solution to detect illegal content in end-to-end encryption communications.[18] Client-side scanning broadly refers to 'systems that scan message contents such as images, text and videos, for matches against a database of previously known illegal content, before the message is encrypted and sent to the intended recipient.'[19]

[15] S. Hawkes and others, *Perceptual Hash Inversion Attacks on Image-Based Sexual Abuse Removal Tools*, pre-print, 2024, arXiv:2412.06056v1

[16] C. Somers, *Ensuring Online Safety - The Role of Artificial Intelligence in Combatting Illegal Content Online*, 2023, https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/AI-combatting-illegal-content-online

[17] Ibid.

[18] G. Ralton, *Proposed mechanisms to detect illegal content can be easily evaded, study finds*, 2022, https://www.imperial.ac.uk/news/239291/proposed-mechanisms-detect-illegal-content-easily/

[19] Ibid.

Invited speaker:
Match Group, Badia Berrada, Director of Corporate Affairs in Europe.

Discussion around the tables:

- Share with the participants at your table examples of national practices if any on the use of AI to detect illicit content online. Are these practices/tools promoted by the government or by private companies?
- Enumerate at least three points of strength and three points of weakness of the use of AI to detect illegal content online.
- What role can the EU play in that respect? For example: guidelines, recommendations, elaboration of a database of identified illicit content?

Panel 2 – Risk assessment analysis tools

Risk assessment is envisaged under Article 51 of the Istanbul Convention. It consists of a 'decision-making process through which we determine the best course of action by estimating, identifying, qualifying and quantifying risk.'[20] Its purpose, as reported in a 2019 EIGE report, is to 'reduce harm to female victims of intimate partner violence and their children,' by identifying 'all levels of risk, namely standard, medium and high, as well as victims' specific needs'.[21] At first sight, it seems that only the criminal justice system is requested to use the risk assessment in the decision for, e.g., protection and/or barring orders. However, as GREVIO stressed in its Mid-Term Horizontal Review:

> All relevant professionals, not only law enforcement, are obliged to assess and take steps to manage the safety risks to a particular victim on a case-by-case basis, including the risk of repeated and lethal violence and, if necessary, to provide co-ordinated support.[22]

Some risk assessment tools have been elaborated for all responders, such as the *Ontario Domestic Assault Risk Assessment* (ODARA), which can be used by a wide range of service providers including: shelter workers, victim services workers, attorneys, health care professionals and social workers.[23] Sharing of data and cooperation across law enforcement, health and social services are essential elements for an effective preventive strategy.

An 'individual assessment' to identify victims' protection needs is envisaged in Article 16 VAW Directive, 'in addition to the requirements for the individual assessment set out in Article 22 of Directive 2012/29/EU' and must be performed at the earliest possible stage, 'such as at the time of first coming into contact with the competent authorities or as soon as possible after first coming into contact with them.' We dedicated the **first meeting of the EU Network to risk assessment analysis tools**, and in this session we will focus only on those that are implemented through AI.

One example is the Integral Monitoring System in Cases of Gender Violence (**VioGén System**), a computer application created by the Spanish Secretary of State for Security (SES) of the Ministry of Interior, which has

[20] T.L. Nicholls, S.L. Desmarais, K.S. Douglas, P. Randall Kropp, *Violence Risk Assessment with Perpetrators of Intimate Partner Abuse*, in J. Hamel, T.L. Nicholls, *Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment*, New York, 2017.

[21] EIGE, *Risk Assessment and Management of Intimate Partner Violence in the EU*, 2019, p. 19.

[22] https://rm.coe.int/prems-010522-gbr-grevio-mid-term-horizontal-review-rev-february-2022/1680a58499, para. 451.

[23] https://www.hss.gov.nt.ca/professionals/sites/professionals/files/resources/guide-risk-assessment-safety-planning-ccp.pdf

been functioning continuously in Spain since July 26th, 2007.[24] The system, technically a web application integrated into the Spanish SARA Network, manages over 510,000 cases. The gathered information includes data on the commission of gender-based violence-related crimes, infractions, criminal records, and the penitentiary situation of alleged perpetrators regarding permits or release. It also contains identifying data, personal information (filiation, family, employment status, etc.), and data on assistance and support to victims (e.g., type of help, use of shelters). Access and use are limited to specific professionals, including members of judicial branches, the Public Prosecutor's Office, specialists in law enforcement agencies (LEAs) at all levels, penitentiary administrations, coordination and violence units, integral forensic appraisal units, welfare services, and social services of local entities. Users are given a specific profile with differing levels of privileges to ensure security and compliance with data protection laws[12]. Cases are active when subject to police attention, inactive when they no longer require it but can be reactivated, and are considered 'low risk' (cancelled) only under specific legal circumstances like a firm acquittal or an executed firm conviction.

The VioGén System is useful for coordination and information exchange; risk assessment; monitoring and protection; prevention and alerts. The algorithm estimates and predicts the risk victims face of suffering from recidivism in each reported case. The system employs a probabilistic risk assessment model composed of 37 variables, such as the severity of past violent acts, the presence and nature of threats made by the perpetrator, the perpetrator's history of violence, and the degree of vulnerability of the woman and any children involved, among others. The system's algorithm analyses the data through predefined rules and risk assessment criteria to assign a specific risk level, which can range from "no perceived risk" up to "extreme risk." The level of police protection offered to the victim is defined based on this assigned risk category, ensuring a tailored approach to victim safety. The risk assessments can also be modified and adjusted manually by officers if needed. Depending on the risk level, the protocol specifies the necessary police protection measures to enhance the victim's safety. These measures escalate in intensity as the risk level increases. For instance, in the case of "persistent" aggressors, the high-risk level of recidivism lead to additional protocol measures, including increased surveillance, monitoring of the aggressor and communication to the judicial authorities and victims themselves. The role of police officers remains crucial in this type of risk assessment. Their experience in handling cases of violence against women enables them to interpret the context of the situation and adjust the risk assessment levels when needed, thereby integrating what the algorithmic system may overlook. Consequently, the system serves as a valuable supportive tool.[25]

Invited speaker:

Presentation of **VioGén System -** María Álvarez Fernández, Interior Ministry, Spain

Discussion around the tables:

- Do you have similar experiences in your country?
- Would this system work in your country?
- Which difficulties / challenges do you see in the use of this system?

[24] J.L. González-Álvarez, J.J. López-Ossorio, C. Urruela, M. Rodríguez-Díaz, *Integral Monitoring System in Cases of Gender Violence. VioGén System.* Behavior & Law Journal, 4(1), 2018, 29-40.

[25] https://interoperable-europe.ec.europa.eu/collection/public-sector-tech-watch/viogen-50-discovering-spains-risk-assessment-system-gender-based-violence

DAY 2

Introduction by the European Commission on the AI Act, followed by a Q&A session.

Panel 3 – Ethical considerations of the use of AI

When using AI to prevent GBV and DV, several ethical and privacy concerns must be taken into account. In terms of processing of **personal data**, the GDPR is the legal instrument in force, hence apps and risk assessment tools must comply with these standards. Hence, for example, in VioGén users access the system with a username and a personal and non-transferable key that allows the auditing of their activities, but has limitations, both in terms of the information that they can access and the functionalities they can activate.[26]

Secondly, algorithms can also **reproduce entrenched stereotypes**, biases and myths. VioGén itself has been 'criticised for relying on algorithms that inadvertently reinforced existing gender stereotypes and biases, particularly in the assessment of victims and offenders.'[27] The risk assessment criteria might not grasp the complexities of individual cases, 'leading to misclassification and potentially harmful outcomes for victims/survivors, particularly women.'[28]

Thirdly, there are potential **downfalls and ethical issues** associated with using technology, such as mobile apps and chatbots, to support victim-survivors of domestic abuse and stalking. As reported with regard to the ISEDA project, there are:[29]

a) **Safety and Security Concerns:** Victim-survivors fear being tracked by the perpetrator, having their personal data exposed or unsecure, and being unable to completely delete data from their device.

b) **Accessibility Issues:** Not all victim-survivors may be able to access digital technology due to **socio-economic constraints** or if they belong to **minoritised groups** such as neurodiverse individuals, deaf people, older survivors, or survivors with literacy challenges, as technologies may not be usable in their current forms.

c) **Lack of Human Emulation:** Technology, like chatbots, struggles to emulate 'real' human characteristics such as empathy, the ability to read body language, or make eye contact. Suggestions propose that technology should act as a **pathway to human services** rather than a replacement for human contact.

d) **Maintenance Issues:** The technology itself may not be suitably maintained, which can lead to problems like **outdated information** or **broken hyperlinks**, as found in an evaluation of existing domestic abuse applications.

Fourthly, there are ethical considerations related to **structural and intersectional** inequalities. To have access to apps and chatbots, it is necessary to have digital literacy, access to an updated device, and issues like poverty, social conditions, economic violence, migrant status and others may prevent it, further exacerbating entrenched systems of oppression.[30] Developers must be aware that functions like location services, often used to find local support, can be exploited by perpetrators to **track the victim**.

Finally, there are ethical considerations concerning the status of the victim. As it was argued, a chatbot must be programmed to **quickly identify if a woman needs immediate assistance** (e.g., police or ambulance). In

[26] J.L. González-Álvarez, J.J. López-Ossorio, C. Urruela, M. Rodríguez-Díaz, *Integral Monitoring*, cit.
[27] A. Karagianni, *The EU Artificial Intelligence Act through a Gender Lens*, Friedrich Ebert Stiftung, Bonn, 2025, p. 4.
[28] Ibid.
[29] See in detail, K. Butterby, N. Lombard, *Developing a chatbot to support victim-survivors who are subjected to domestic abuse: considerations and ethical dilemmas.* Journal of Gender-Based Violence, *9*(1), 2025, 153-161. Retrieved Sep 28, 2025, from https://doi.org/10.1332/23986808Y2024D000000038
[30] See also below, panel 4.

situations of immediate danger, a chatbot is unsuitable and should direct users to **emergency services**.[31] Can technology do that? Also, when elaborating a chatbot or an app, language should be gender and trauma sensitive and elaborated in conjunction with professionals from women's services.

Invited speaker:

> **AI Forensics** is a European non-profit that investigates influential and opaque algorithms. Their mission is to apply innovative methods to uncover digital rights violations and provide information to help shape regulatory policies. Speaker: Silvia Semenzin.

The discussion around the tables will be replaced by a general discussion moderated by Tamsin Rose.

Panel 4 – Mobile apps to support victims of GBV and DV

Mobile apps and chatbots have been conceived and implemented to support victims of GBV and DV. Digital tools, such as safety apps, reporting portals, and chatbots, are increasingly being used by victim-survivors of gender-based violence to report unlawful activity and access specialised support and information. Digital tools are an alternative or complementary pathway for more traditional reporting, help-seeking, and prevention, because they are considered by victims as more private, accessible and responsive.[32] Applications can be used by victims to collect and store evidence that can be used during formal proceedings. Chatbots are computer programs that mimic human conversation using text, voice, or a combination of these. Users can access chatbots in a range of ways, such via messaging apps or standalone websites. Advancements in machine learning and natural language processing (NLP) have started a new era of so-called 'intelligent' text-based chatbots over the past decade.[33] There are also chatbots and online reporting tools for reporting sexual violence.
Some examples are the following:

Apps and chatbots

ISEDA: a chatbot to support women who are subjected to domestic violence abuse (DVA) that is part of a wider project, Innovative Solutions to Eliminate Domestic Abuse (ISEDA), which consists of 15 partners from 9 European Countries (only 7 were included in this work package). The wider project aims to use multi-sector expertise via modern technological tools and practices in order to tackle and eliminate DVA. One such tool is the chatbot for victim-survivors who are subjected to DVA. ISEDA chatbot is based on artificial intelligence (AI), but it is not an AI Chatbot. It uses AI for intent and entity detection only and not for answer generation; the answers have been preprogrammed by the ISEDA team. The purpose of the chatbot is to enable victim-survivors to search for information in relation to DVA, such as local legislation, victim-survivors' rights, and information about local support services.[34] https://iseda-project.eu/

IMPROVE: AinoAid™ is at the forefront of the EU-funded IMPROVE project, a collective effort to combat domestic violence. IMPROVE is creating tools to enhance the reporting and detection of domestic violence,

[31] Ibid.
[32] N. Henry, A. Witt, S. Vasil, *A 'design justice' approach to developing digital tools for addressing gender-based violence: exploring the possibilities and limits of feminist chatbots*, in *Information, Communication & Society*, *28*(11), 2024, 1884–1907.
[33] Ibid.
[34] N. Lombard and others, *Introducing a Chatbot to Support Victim-Survivors of Domestic Abuse: Victim-Survivor Perspectives*, in *Violence Against Women*, 2025, 1 ss.

empowering victims to access services and justice. They are teaming up with Police Authorities, Civil Society Organizations, and others to implement innovative solutions. https://ainoaid.fi/

FAIR: is a global network of scientists, economists and activists united in their dedication to finding ways to make AI and related technologies more effective, inclusive and transformational. FAIR-supported projects aim to identify and correct digital biases by fostering collaboration and developing AI solutions that reflect feminist principles. https://aplusalliance.org/feminist-ai-research-network/ Within this network, there is:

- **SafeHER** is an app designed for women transit users in Manila, the Philippines, based on their lived experiences and needs. It provides tools such as SOS alert, live location sharing, scream detection, and a buddy system to enhance their safety on public transport.
- **AymurAI** was developed to address the lack of data on gender-based-violence cases in the Argentinian judicial system, ultimately fostering greater accountability and transparency within the judiciary when it comes to gender-based violence.
- **SOF+IA** is a web-based feminist chatbot created to support victims of technology-facilitated gender-based violence on social-media platforms. It guides users on how to report cases, provides digital self-care tips and evaluates whether a situation can be reported to police.

FollowItApp: a mobile app in Scotland created collaboratively to support stalking victim-survivors, as an example of working with relevant groups in its creation. https://followitapp.org.uk/

NonPossoParlare (*I cannot speak* in Italian): In order to combat the risk of domestic violence, the NONPOSSOPARLARE app has been developed, a project aimed at women who wish to receive information and support from anti-violence centres with no need to talk on the phone. It is aimed in particular at people who have difficulty communicating, forced to stay at home with their psychologically or physically violent partners. The NONOPOSSOPARLARE chatbot responds effectively to an unlimited number of people at the same time, providing users with 24/7 support anonymously and without leaving trace.

SophiaChat: Sophia is a Digital Companion (Chatbot) that offers a safe, anonymous, and accessible way to recognise relationship patterns, explore available options, and securely document important information. Designed to provide guidance, resources, and support, Sophia helps navigate difficult situations with care and confidentiality. https://sophia.chat/

Epowar: set up in 2020 by University of Bath graduates E-J and Maks, is an innovative AI-powered personal safety app for smart watches which harnesses a wide range of original features. It exists to empower women and girls and includes a patent-pending Automatic Attack Detection technology. https://www.epowar.com/

Invited speaker:

ISEDA (Horizon project) https://iseda-project.eu/ Speaker: Kate Butterby, Glasgow Caledonian University

Before the event:
Please provide link and short description of possible apps, chatbots existing in your country.

Discussion around the tables:

- Test one of these apps.
- Are there similar experiences in your country? Are there any data available on their use?
- What are the main difficulties in using this apps?
- What could the EU do? e.g. further develop the existing Horizon projects? Support the creation of an EU-wide app?

Conclusions

In a final session, the main points of the discussion will be summed up, and further suggestions collected from the participants on the role of the EU in the implementation of the VAW Directive and of the Istanbul Convention within the limits of EU competences.