AI and Open Data for GBV Prevention Among Youth: Ethical Governance and Policy Frameworks in the EU

1st Musharu, Timothy
Dept of Computer Science
University of Oldenburg)
Oldenburg, Germany
Timothy.musharu@uol.de

2nd Gomez, Jorge Marx
Dept of Computer Science
University of Oldenburg
Oldenburg, Germany
Jorge.marx.gomez@uol.de

3rd Rosales Megan
Developmental Psychology And Didactics
University Of Alicante
Alicante, Spain
ignasi.navarro@ua.es

4th Navarro Soria, Ignacio
Developmental Psychology And Didactics
University Of Alicante
Alicante, Spain
megan.rosales@ua.es

***Abstract*—This paper examines the utilization of artificial intelligence (AI) and open data for gender-based violence (GBV) prevention among youth in the European Union. A systematic review of 74 sources—comprising 45 peer-reviewed academic articles, 12 EU policy documents, and 17 pieces of grey literature from 2019 to 2024—was conducted to evaluate the technical performance, ethical implications, and policy coherence of AI-driven GBV interventions. The analysis indicates that while AI tools, such as predictive analytics and social media monitoring, offer promising capabilities for early detection and prevention, their effectiveness is constrained by their current pilot-stage implementation and the lack of longitudinal evidence. In addition, critical ethical challenges persist, including issues of data privacy, algorithmic bias, and insufficient youth participation in the design process, compounded by fragmented open data practices. Moreover, although the EU has established a robust policy framework—including instruments like the forthcoming AI Act and the 2024 GBV Directive—the practical application of these policies remains inconsistent across Member States, particularly in terms of interoperability and standardized reporting. These findings underscore the need for enhanced open data integration, participatory design approaches, and harmonised policy implementation to ensure that AI interventions are both effective and ethically sound. Addressing these gaps will enhance the EU's ability to deploy evidence-based AI solutions that protect vulnerable populations while upholding digital rights.**

***Index Terms*—Gender-based violence (GBV), artificial intelligence (AI), EU policy, algorithmic bias, Open data**

I. Introduction

Gender-based Violence (GBV) among youth has increased on digital and online spaces, raising complex challenges for prevention and intervention. An estimated 1 in 2 internet users have encountered Technology assisted GBV [2]. GBV in this context encompasses physical and sexual violence, as well as stalking, harassment, and abuse facilitated through technology [1], [2].GBV is defined as violence committed against a person because of his or her sex or gender. It is forcing another person to do something against his or her will through violence, coercion, threats, deception, cultural expectations, or economic means. Although the majority of survivors of GBV are girls and women, LGBTIQ+, boys and men can also be targeted through GBV [2]. Technology facilitated GBV includes cyberstalking, sexual harassment, dating violence, and online grooming. This phenomenon demands innovative, scalable solutions to protect vulnerable populations such as women and youth. Globally, Fifty-eight percent of young women report experiencing some form of online harassment or abuse[3], with social media platforms and encrypted messaging apps amplifying risks for adolescents across the EU [4], [5]. These trends coincide with the EU's prioritization of digital transformation under initiatives such as the Gender Equality Strategy 2020-2025[6] and the Digital Services Act (DSA) [7], which emphasizes the use of technology to combat systemic inequalities. Artificial intelligence (AI) has emerged as a dual-edged tool in this context. On one side, AI offers innovative capabilities, but on the other side, AI poses several ethical risks. While AI systems form predictive risk models to chatbots for survivors, promise proactive GBV prevention, their deployment raises ethical dilemmas tied to data privacy, algorithmic bias, and transparency. For example, Spain's VioGén system, which uses machine learning (ML) to assess domestic violence risks, has faced criticism for bias in risk scoring[8].In addition generative AI tools like deep fakes aid online harassment [9]. The EU's regulatory frameworks, such as the AI Act and General Data Protection Regulation (GDPR), aim to mitigate these risks but lack specificity in addressing youth-centric vulnerabilities or

EU-Funded

the ethical nuances of open data sharing in GBV prevention. This systematic literature review (2019–2025) evaluates 74 sources (45 academic articles, 12 EU policy documents and 17 grey literature sources) to answer the following research gaps:

- **Technical Efficacy:** How are AI tools such as predictive analytics and social media monitoring applied to youth-focused GBV prevention, and what are their limitations?
- **Ethical Risks:** What governance challenges arise from AI's use in sensitive contexts, particularly regarding bias, privacy, and youth agency?
- **Policy Alignment:** How do EU frameworks balance innovation and ethical safeguards in AI-driven GBV interventions, and where do inconsistencies persist?

The analysis draws on AI tools from EU member states, such as Spain's VioGén, CESAGRAM's online grooming detection tools and global comparisons such as India's intimate partner violence prediction models [10] to highlight region-specific lessons. By situating AI within the EU's social sustainability agenda, this review advocates for Business Information Systems driven governance frameworks that harmonize technical innovation with youth protection, human rights, and open data ethics.

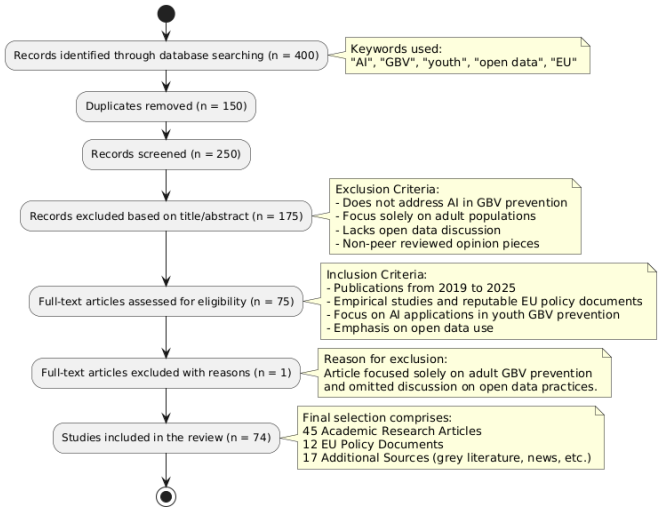

Fig. 1: Systematic Desk Review methodology

II. Methodology

This study employed a systematic desk review of 74 sources retrieved from Scopus, Web of Science, Google Scholar, and various EU policy portals, covering publications from 2019 to 2024. The final selection includes 45 peer-reviewed academic articles, 12 EU policy documents, and 17 grey literature sources see Table 1. Rigorous inclusion criteria prioritized empirical studies and policy texts addressing AI applications in youth-focused gender-based violence (GBV) prevention, while excluding research that either did not involve technological dimensions, focused solely on adult populations, or consisted of non-peer-reviewed opinion pieces unless they provided critical policy insights. Given the centrality of open data to this review, additional emphasis was placed on sources that discuss data-sharing practices—such as the use of crime statistics, helpline usage figures, and judicial outcomes—and their integration into AI-driven systems. An iterative coding process was used to synthesize the technical, ethical, and policy dimensions pertinent to AI for GBV, mapping use cases, identifying governance challenges, and assessing EU regulatory alignment. Fig . 1 shows the PRISMA flow diagram which visually illustrates the review process from identification through screening and eligibility to final inclusion.

TABLE I: Systematic Desk Review Details

Aspect	Details
Total Sources Analyzed	74 items in total: 45 academic research articles, 12 EU policy documents, and 17 additional sources (grey literature, news, etc.)
Indices/Databases Searched	Scopus, Web of Science, Google Scholar, and EU policy portals (e.g., EUR-Lex, European Commission websites)
Inclusion Criteria	• Publications from 2019 to 2025 • Focus on AI applications in the prevention of gender-based violence (GBV) among youth • Empirical studies and reputable EU policy documents • Written in English • Keywords "AI", "GBV", "Youth","Open Data", "EU"
Exclusion Criteria	• Studies not addressing an AI component in GBV prevention • Research focusing solely on GBV interventions without a technological dimension • Studies targeting only adult populations • Non-peer reviewed opinion pieces (unless providing critical policy insights)
Open Data Considerations	• Priority given to studies or policy documents that explicitly discuss the availability, use, or governance of open data for GBV prevention • Examination of how open datasets (e.g., crime statistics, helpline usage, judicial outcomes) inform AI-driven GBV interventions • Exclusion of sources lacking any mention of data-sharing or transparency aspects, unless otherwise relevant

III. Ethical Challenges of AI in GBV Prevention

While AI offers innovative tools to tackle GBV, it also raises significant human and ethical rights challenges- especially in a sensitive domain involving vulnerable youth. Key concerns identified in literature include algorithmic bias, data privacy, risks of surveillance, and questions of inclusivity and accountability. Fig. 2 visually organizes the ethical challenges of deploying AI in GBV prevention among youth, structured into four core themes identified in literature [11]: Bias and Discrimination, Data Privacy and Consent [11], Over-Surveillance and Misuse, and Lack of Inclusivity and Accessibility [12].

A. Bias and Discrimination

AI systems deployed in GBV prevention risk perpetuating societal biases due to reliance on historical datasets that underrepresent marginalized groups, such as LGBTQ+ youth and male victims or encode gender stereotypes [17], [18], [19]. For example, one audit for Spain's VioGén system, which uses machine learning to assess domestic violence risks, faced criticism for bias in risk scoring during audits, reflecting broader concerns about algorithmic discrimination in predictive policing tools. Biased algorithms could also prioritize some victims over others or stigmatize certain communities. The EU's Ethics Guidelines for Trustworthy AI explicitly call for "diversity, non-discrimination, and fairness" in AI systems [20], requiring audits to identify and mitigate biases during development. In addition, the tech industry's lack of diversity – women constituting less than twenty-five percent of AI professionals [21] exacerbates these risks as homogeneous teams may overlook the intersectional vulnerabilities that exist in the GBV context [21]. This imbalance means AI tools may not fully account for experiences of women and gender based minorities or may encode subtle sexist assumptions. The EU's AI act classifies GBV-related AI tools as high risk , mandating bias assesments and transparency reports. Yet challenges remain in operationalizing these requirements, partcularly for systems that are trained on incomplete or culturally narrow datasets. This further pushes the need for a collaborative multi-dimensional and diverse approach to buidling AI algorthms for equity and equality. Ethically it is crucial to audit AI for Bias and involve diverse stakeholders in their development. This is enshrined in the EU's AI act, which calls for diversity, non discrimination and gender inclusion in AI systems- Principles which should form a guide for any GBV-related AI deployments.

B. Data Privacy and consent

Technological interventions for societal challenges data and further require processing of sensitive data [22]. From incident reports , criminal records or social media interactions form part of the data needed which raises critical questions about GDPR compliance and informed consent,anonymity and data protection. For instance AI chatbots collecting youth mental health data must ensure encryption and anonymization to avoid retraumatization or deterrance from seeking help. The EU's law enforcement directive [23] prohibits fully automated decisions in policing without safeguards, yet gaps remain in youth-specific protections [24]. Surveillance-type data collection is a particular worry: if AI tools monitor online spaces for grooming or harassment, they must tread carefully to avoid unjustified intrusion into private communications. There is a fine line between proactive monitoring for public safety and the risk of creating a surveillance environment for youth, which could surpress their online expression. The ethics become especially fraught if AI is applied to predict individuals' propensity for violence or victimization, essentially profiling people based on data [14], [25]. Misuse of such profiles could label someone "high-risk" in police databases without their knowledge, potentially affecting their rights. For example, The CESAGRAM Project, which monitors online grooming, highlights the tension between proactive safety measures and intrusive surveillance [14]. While its AI detects harassment patterns, critics warn that unchecked data collection could discourage freedom of expression, violating Article 7 (Privacy) and Article 8 (Protection of personal data) of the EU Charter (data protection) [25].

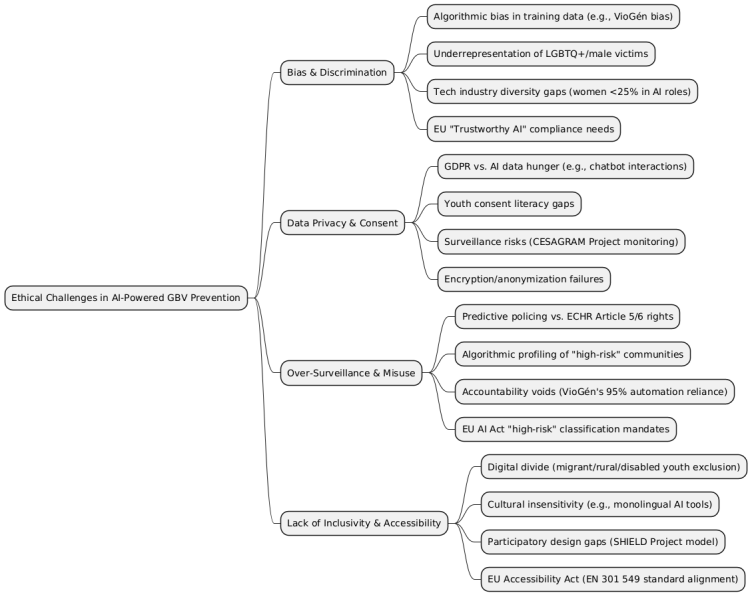

Fig. 2: Ethical challenges for AI-powered GBV prevention

Any AI system that handles youth data or survivor information must be designed with strict privacy-by-design measures: data minimization, encryption, and clarity on who can access the outputs [26]. Advocacy groups also stress the importance of informed consent – users (be they young people using a chatbot or victims whose case data feed a model) should know that AI is being used and have some agency over it. In summary, maintaining confidentiality and trust is paramount; otherwise, the very communities interventions are poised to help may feel "spied on" or exploited by AI systems.

C. Risks of Over-Surveillance and Misuse

Closely related to privacy and consent are broader civil liberaty concerns. AI if not properly governed could aid in an overreach of surveillance under the pretext of preventing GBV. As an example predictive policing tools might propt authorities to intervene preemptively (before any crime is committed) based on AI algorithms suggestions which clashes with legal principles of innocence and due process under the European Convention on Human Rights [27].Although the AI Act clasifies social scoring as "unnacceptable risk" which seeks to discourage this , underlying algorithmic calculations can give a false aura of objectivity that justifies intrusive measures. There is also a danger that designating certain neighborhoods or groups as "high risk" leads to disproportionate policing and surveillance in those communities, reinforcing stigma and distrust. Now, In the context of youth, consider an AI that flags a teenager's messages as indicative of being abusive or harrassing – if shared with law enforcement or schools without context, it might lead to punitive action rather than education or rehabilitation, potentially derailing a young person's future based on an algorithmic judgment. The accountability for AI errors or misuse is another ethical dilemma: If an AI chatbot gives a survivor dangerously poor advice, or a risk assessment tool fails to predict a lethal incident (or falsely predicts one), who is responsible? Ensuring human oversight is critical, yet as one audit noted, sometimes "humans are not in the loop" [28] – Spanish police were found to stick with the VioGén system's risk score ninety-five percent of the time, with minimal independent judgment. In the presence of technological aid, humans tend to over-rely on automation as opposed to their flawed judgment. This underscores the goal for the EU AI Act, which seeks for enforceable accountability mechanisms, transparency and rigorous testing with mandated human-in-the-loop control for high-risk classified systems. Using an ethical lens to view this ensures that AI use must be aligned with human rights, which ensures prioritization of safeguards over speed or availability, no matter how well intentioned tools that encroach on personal freedoms or operate opaquely have the propensity to do more harm than good in the long haul.

D. Inclusivity and accessibility

AI tools must be designed and deployed with the marginalized voices in mind. In today's world, AI tools often fail to address the digital divide or cultural nuances [29]. For example, chatbots assuming everyone has laptops or smartphone access exclude rural, disabled or low-income youth, while monolingual models overlook linguistic diversity in migrant communities [30]. The EU-funded SHIELD Project [31] advocates participatory design, engaging youth, youth workers, experts and LGBTQ+ advocates to refine AI responsiveness to intersectional needs. The digital gender divide remains an issue even within Europe [22], [32], some girls or vulnerable youth have less access to technology or feel less confident using it, so a purely AI-based support system might leave behind those who are offline or uncomfortable with chatbots. Inclusivity also means accounting for different languages and dialects (an AI monitoring tool must be trained on the languages youth use, including slang), and different contexts of GBV (from dating violence in teens to abuse in migrant communities). The ethical principle of beneficence requires that AI for GBV must not just work for the "average" user but for all potential users, especially those at higher risk. Efforts like participatory design – involving youth, youth advocacy groups, survivor networks, and intersectional experts in creating AI tools – can improve inclusivity. The SHIELD[31] project explicitly mentions increasing awareness of ethical AI practices and engaging civil society and policymakers to ensure AI solutions are accepted and effective. Lack of inclusivity is also crucial for long-term adoption and safety of the tool: if certain groups (say, non-binary youth or minority ethnic groups) don't see their experiences reflected in an AI's responses or risk criteria, they may mistrust or ignore the tools. Finally, inclusive [33] AI needs to be transparent and explainable to users. A young person should be able to understand, at least in simple terms, why an AI chatbot is asking certain questions or why a risk score or recommendation was given to their case. Building that understanding and trust is key to ethical adoption of AI in this domain. In summary, deploying AI against GBV among youth must navigate a tightrope between leveraging data for good and upholding the rights and dignity of individuals. Guidelines such as the [20] Ethics guidelines for trustworthy AI should not only be known but also adhered to. Bias mitigation, strict data governance, human oversight, and inclusive design are not just ideal practices – they are ethical imperatives. As recommended by the EU , building trustworthy AI should incorporate the [34] Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment to be compliant. The next sections will see how these considerations are being addressed (or not) in policy frameworks, particularly within the EU.

IV. EU Policy Landscape: AI, GBV, and Youth

The European Union has, in recent years, sharpened its policy focus both on advancing AI in a human-centric way and on combating GBV (including violence against youth and online abuse). This Section critiques the EU policy landscape on AI, GBV and Youth. In this respect, several key EU strategies and legal instruments intersect to shape how AI can be used in GBV prevention:

1) **EU Gender Equality Strategy (2020–2025):** This strategy frames the EU's commitment to end gender-based

violence as a top priority for social Sustainability. It calls for using all available tools – legislation, funding, awareness – to prevent and combat GBV [35]. While not explicitly mentioning AI, the strategy underlines the need to address emerging challenges like cyber violence and to engage technology actors in solutions. The European Commission's High-Level Expert Group on AI also issued Ethics Guidelines [20] that prioritises fairness and non-discrimination [22] in AI design development and deployment. The synergy of these policies is evident in [36]. The Commission highlights AI as a driver of economic progress but stresses that it must not perpetuate discrimination. This provides a broad mandate for [34]"trustworthy AI" that supports gender equality – implicitly encouraging AI applications that help protect women and youth, so long as they align with EU values of privacy, non-discrimination, and transparency.

2) **Legislation on Violence Against Women and Online Abuse:** The [37] Directive on Combating Violence Against Women criminalizes online GBV such as non-consensual intimate image sharing (Cyber-flashing), cyberstalking and mandates technology driven compliance for digital platforms [35]. For example:
   - Content Moderation: Platforms may deploy AI filters to detect and remove illegal abusive content such as deepfake pornography within 24 hours, per Article 12
   - Victim Support: Member States are required to establish prevention programs auch as AI-powered reporting channels (such as chatbots like IMPROVE [38]) to ensure 24/7 access to legal and psychological aid.

   The directive obliges Member States to implement robust prevention measures, victim support mechanisms, and improved access to justice [35]. Each EU country is in the process of transposing this directive into national law, creating a more uniform framework where AI solutions for GBV can be deployed, knowing the legal definitions and duties are aligned across the Union.
3) **EU Accession to the Istanbul Convention:** The Istanbul Convention [39] is a gold-standard treaty on preventing and combating violence against women. The EU as a bloc formally acceded to the Convention in 2023 reinforcing its commitments. Notably, Article 17 of the Istanbul Convention calls on states to encourage the ICT sector to participate in violence prevention policies [40]. This policy direction has spurred collaboration with tech companies and innovators to curb online harassment and abuse. We see its influence in EU initiatives that engage social media firms in addressing hate and harassment such the [41] Code of Conduct on countering illegal hate speech, which covers misogynistic hate. Accession of this bill means the EU must coordinate implementation of the Convention's provisions – potentially including partnerships for safer internet environments for youth and developing new tools (where AI can play a role) to detect and prevent violence [40] Many EU countries have already integrated Istanbul Convention obligations into national action plans, some of which mention leveraging technology for prevention (for example, Spain's national strategy references improving data systems like VioGén to fulfill Convention duties[13])
4) **National Policies and Pilot Programs:** Individual EU Member States have launched their own AI-related GBV initiatives, offering case studies of policy in practice. Spain stands out: under the State Pact against Gender Violence (2017–2022), Spain invested in enhancing the VioGén risk assessment platform with AI capabilities [13]. Although it is without controversy, this reflects a policy decision to harness AI to strengthen protection orders and policing of domestic violence. An audit by a Spanish NGO (Eticas) in 2022, supported by some policymakers, pressed for greater transparency and oversight of VioGén's algorithm[42]. The Spanish Interior Ministry defended the system's data-driven approach, citing continuous improvements and academic partnerships [42]. This debate has informed policy discussions on how to integrate AI ethically – for instance, by involving women's advocates in system design and ensuring victims are informed about how risk scores are used[26]. France, while focusing more on legal measures, has also discussed using technology for protection such as considering electronic tracking of domestic violence offenders, which could involve AI in monitoring geolocation data [43] (though privacy concerns are being debated in policy forums such as Commission nationale de l'informatique et des libertés - CNIL [French data authority]. Nordic countries like Finland and Denmark have strong digital education policies; their governments have supported campaigns and school programs to educate youth on digital citizenship and relationship ethics, sometimes using AI-based interactive content (aligning with a broader policy of promoting safe digital environments for youth) esnshrined in [44]. In Eastern Europe, some governments (e.g. Lithuania, Poland) have been slower to adopt AI in social sectors, but EU-funded pilots (through Horizon Europe or Justice Programme) are bringing tools like victim support chatbots to these countries as well, often in collaboration with NGOs. Overall, EU nations vary in their approach, but a common trend is that AI is emerging: policy interest in data-driven approaches to GBV is growing, provided these tools respect the stringent European data protection and human rights frameworks[27].
5) **EU Artificial Intelligence Act (Draft)**: The EU is in the final stages of negotiating the AI Act, a horizontal regulation of AI that will likely come into force in 2025–2026. This law will categorize AI systems by risk. Fig. 3 shows the proposed structure of the AI act. Notably, AI systems used by law enforcement for risk assessment or profiling are expected to be classified as "high-risk," subject to requirements like human oversight, transparency, and audits[23]. If a predictive policing tool for domestic violence is deployed, it would need to

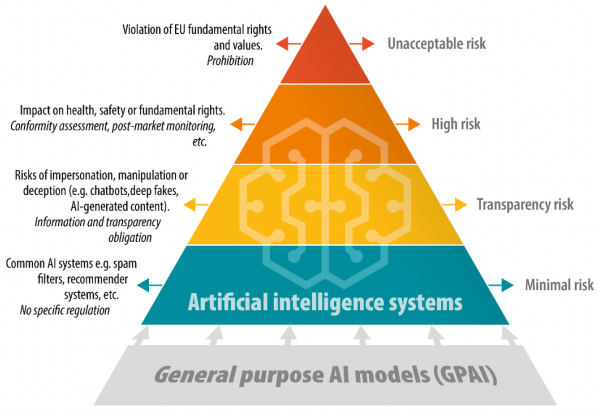

Fig. 3: Summary of the AI ACT

comply with these standards. Some uses may even be prohibited if deemed too harmful – for example, there have been parliamentary debates on banning social scoring and certain forms of biometric surveillance [42]. While GBV prevention tools are not targeted for bans, any AI that could materially affect people's rights (which includes most GBV applications) will be tightly regulated. The AI Act also emphasizes data quality – ensuring training data is relevant, representative, and free from errors – which directly addresses the bias concerns in GBV algorithms. Additionally, the Act will mandate risk assessments and documentation for high-risk AI, meaning developers of, say, a GBV risk prediction model in an EU country must provide authorities with details on how it works, its limitations, and mitigation measures for risks like discrimination. This regulatory push reflects the EU's commitment to "human-centric AI" and will shape the development of future GBV tech, anticipating that compliance with the Act will become part of project planning (as seen in EU-funded projects that already incorporate ethics reviews). Policymakers in Europe thus set a clear expectation that AI can be used in public services (like violence prevention) only under strict guardrails to protect fundamental rights and safety [45]. Such an approach contrasts with more laissez-faire environments elsewhere and is intended to foster public trust in AI interventions.

6) **Other Related Digital Policies:** Other EU policy instruments indirectly support AI against GBV. The Digital Services Act [46] (DSA) 2022 imposes duties on online platforms to address illegal content and systemic risks. Large social media companies must assess and mitigate risks such as the spread of misogynistic harassment or the impact on minors' mental health [47]. Many will deploy AI filters and recommender tweaks to comply. The EU Strategy on the Rights of the Child [44]calls for making the internet safer for children and echoes the need to tackle online sexual exploitation – here the use of AI to detect grooming or abusive material (like known child sexual abuse images via hash-matching) is encouraged, though balanced with privacy (the debate on scanning messages with AI for abuse material is ongoing at EU level). Furthermore, the EU's funding programs (Horizon Europe, Digital Europe) are investing in research and innovation on "tech for good." Projects like SHIELD, IMPROVE, and ISEDA [48] (Innovative Solutions to Eliminate Domestic Abuse) and CESAGRAM are backed by EU funds and often bring together universities, companies, and NGOs across member states. These projects not only produce tools but also yield policy guidance, such as best practice manuals and ethical frameworks that feed into EU policy discussions. For example, findings from Horizon 2020's IMPRODOVA [49] project on police response to domestic violence have been used to train officers and inform EU justice policy about the value and limits of data-driven approaches [38].

In summary, the EU's policy landscape is becoming increasingly supportive of AI-assisted GBV prevention, but always with an eye on ethics and rights. The Union's comprehensive approach – binding laws on violence, broad AI regulations, and funding of pilot initiatives – creates a framework within which AI solutions can be scaled responsibly. There is also recognition at the EU level that technology must be part of the solution to GBV (especially as abuse goes digital), yet that it must operate under "European values" of privacy, equality, and accountability. The EU's policy landscape uniquely merges technological ambition with rights-based caution, contrasting with the U.S.'s market-driven model and China's surveillance-heavy tactics. By anchoring AI governance in existing frameworks (GDPR, Istanbul Convention), the EU positions itself as a global leader in ethical AI for social good, though implementation asymmetries persist. This policy posture, distinct from some other regions, has both advantages and challenges, which we explore next by looking at the international outlook on AI for GBV.

V. International Comparisons and Key Lessons

To put the EU's approach in context, it's useful to compare it with other regions' strategies for AI in GBV prevention. Different legal cultures and resource levels have led to varying approaches, highlighting gaps and offering lessons:

A. Australia ,United States and Canada

In North America, there is significant technological innovation around addressing GBV, but less centralized policy guidance than in the EU. In the U.S., law enforcement agencies and courts have experimented with risk assessment algorithms for domestic violence, often at the city or county level. For example, some police departments use actuarial risk tools (similar to the U.K.'s DASH, but augmented with local data) to decide how to follow up on domestic calls. A collaboration between researchers and Greater Manchester Police, cited earlier, involved a University of Chicago economist – indicative of transatlantic knowledge sharing on predictive policing [50]. However, the U.S. lacks a federal equivalent of the EU's AI Act or a comprehensive digital safety law[51]. This means

AI use is governed by a patchwork of state laws and ethics codes. The upside is quicker experimentation (fewer regulatory hurdles); the downside is potential inconsistency and lower baseline protections for privacy. For instance, Protective AI tools might be deployed with minimal transparency in one jurisdiction, while another city might ban automated decision-making in policing altogether due to bias concerns. Civil liberties organizations in the U.S. have challenged predictive policing and face recognition on constitutional grounds, which indirectly influences GBV-related uses[51], [52]. On the social media front, U.S. tech companies like Meta and Google are deploying AI at scale to moderate content and detect harassment, including gender-based hate, spurred by public pressure and their policies rather than law[51]. They report removing millions of posts for hate speech or harassment, and AI is often the first line of detection[26], [47], [53]. Yet advocacy groups note these systems still fail many women – for example, non-English abuse or coded misogyny often slips through[14], [36]. Canada, on the other hand, is more aligned with European thinking; it has proposed the Artificial Intelligence and Data Act and strong gender equality commitments[52]. Canadian initiatives have looked at AI for analyzing hotline data or for outreach to indigenous communities facing high GBV rates, emphasizing community consent and ethics[52]. Australia also has the e-safety act which is introduced to curb GBV. However, the lesson from North America is that innovation can outpace regulation, which can lead to both cutting-edge tools and high-profile missteps (e.g. an AI chatbot in a U.S. city that gave inappropriate responses to a teen user, triggering calls for better oversight) [54], [55]. The EU's more preemptive regulatory stance aims to avoid such missteps by setting standards upfront[8],[20].

B. Global South (Africa, Asia, Latin America):

In many developing or middle-income countries, GBV is high on the policy agenda, and there is interest in leveraging AI and mobile technology, given the rapid spread of smartphones[56]. However, gaps in infrastructure and governance often pose challenges. Several notable efforts exist: In West Africa, a youth-led NGO in Nigeria (Brain Builders Youth Development Initiative) launched HerSafeSpace [57], an AI-powered chatbot to tackle online GBV and provide a safe forum for young women. This underscores how NGOs fill gaps where government resources are thin. Similarly, in South Asia, Pakistan's Ministry of Human Rights introduced an AI chatbot for reporting harassment [58], aligning with a broader Women's Safety app. These chatbots often run on messaging apps (WhatsApp, Facebook Messenger) to maximize reach. The lesson from these contexts is the importance of low-bandwidth, anonymous access [30] – for example, the rAInbow [59] chatbot in South Africa was designed to work over basic SMS and Facebook Messenger to reach women with limited internet [59]. Policy in many of these regions is still catching up; few countries have AI-specific laws, though some have digital safety or cybercrime laws that criminalize online harassment and image-based abuse, for example, laws against "cyberstalking" in India [60] or Kenya's [61] Computer Misuse Act addressing intimate image abuse). International organizations provide guidance: UN Women has highlighted how AI can either bridge or widen the gender gap, noting that biased algorithms or the gender digital divide (with only twenty percent of women online in some low-income regions) can exacerbate inequalities [62]. The Generation Equality Forum 2021 launched a global Action Coalition on Gender-Based Violence and another on Technology and Innovation; through this, countries and companies pledged initiatives such as developing new tech tools and improving data systems on GBV. For example, an Africa-wide program was announced to use AI to better map GBV incident data and support hotlines, illustrating a growing political will to invest in tech solutions.

C. Key Lessons on Online Platforms and Generative AI:

One area that transcends borders is the challenge of technology-facilitated GBV, especially with the rise of generative AI. The misuse of deepfakes and AI-generated content is a global concern. A staggering 98% of deepfake pornographic content online is non-consensual and targets women, according to 2023 analyses [29]. This has prompted calls worldwide for stronger regulation of such practices. The EU's new directive explicitly criminalizes sharing deepfake intimate images without consent[35] , and some U.S. states have passed similar laws. Tech companies are developing AI tools to detect deepfakes – for example, Microsoft's Video Authenticator and initiatives under the Partnership on AI's Content Authenticity Initiative – but these tools need global adoption. Meanwhile, generative AI also enables new forms of harassment: "gendered disinformation" (using AI to generate false stories or images about women, often to silence women activists or politicians) has been documented as a rising threat [29]. International bodies like UNESCO have started addressing this; a 2023 UNESCO report [18]on technology-facilitated GBV in the era of generative AI underscores how new AI harms, such as fake nude images or AI-authored hate campaigns, are increasing the scale of abuse [64].[17], [18], [63], [65], [66] recommends multi-stakeholder action – AI developers building safety into their models, governments updating laws, and civil society raising awareness about these "newest forms" of GBV [29], [67]. A positive development is cross-regional collaboration; for example, the EU and UN convened a global advisory group on online violence in 2022, sharing insights with counterparts in Australia (which has an eSafety Commissioner actively tackling online abuse content), and companies like Meta have rolled out certain anti-harassment features (like "Women's Safety Hub") uniformly across countries, often after pilot testing in one region. One clear lesson is that no single country can solve online GBV alone, as AI platforms operate globally and abusers can target victims across borders. Thus, harmonizing policies [67] (for instance, agreeing on definitions of online stalking or standards for AI content moderation) and sharing successful tools, Such as open-source AI models for identifying abuse, are crucial. The EU's work can serve as a model, but it can also learn from others. New

Zealand's holistic approach to harmful digital communications and the grassroots digital literacy programs in Southeast Asia that teach girls how to outsmart online harassers can inform the already existing body of work. In conclusion, comparing regions reveals that the EU's approach characterized by strong regulatory frameworks and funding for ethical AI innovation – is somewhat unique. The U.S. illustrates the benefits and risks of a more decentralized, innovation-led path. Many Global South countries demonstrate ingenuity in adaptation but need supportive policies and safeguards. A key takeaway is that ethics and effectiveness must go hand in hand: tools developed in isolation from policy (or vice versa) tend to falter. International dialogue, facilitated through forums such as the UN and OECD, enables regions to learn from each other. For the EU, it is essential to remain vigilant about emerging threats (such as generative AI) and to ensure that its regulations are designed in a way that supports, rather than restricts, grassroots innovation in this domain. For other regions, adopting some of the EU's protective measures—such as requiring human-in-the-loop for high-stakes AI applications or mandating platform accountability for online abuse—could help mitigate potential harms while still leveraging the substantial benefits of AI.

VI. Open Data for Gender-Based Violence Prevention in the EU

GBV is a pervasive human rights issue, and data plays a critical role in understanding and preventing it. Open data freely accessible, reusable public data – can help reveal patterns of GBV, support evidence-based interventions, and hold institutions accountable [4], [68]. The European Union (EU) has emphasized data-driven approaches in its gender equality and anti-violence strategies. Data-driven approaches are a fundamental aspect of BUIS applications, leveraging open data to transform public resources into enterprise-specific assets through data processing and integration with other sources. This process ultimately enhances products, services, and processes [69]. This research examines how open data is applied to GBV prevention in the EU and reviews relevant EU policies and projects.

1) **EU Open Data Policies for GBV Prevention EU Policy Framework:** The EU's commitment to combating GBV is outlined in instruments like [32] and reinforced by international standards such as the [39]. Open data principles are embedded in the EU's broader digital strategy, notably the [70], which mandates open access to many public datasets. Crucially, in May 2024, the EU adopted its first comprehensive Directive on combating violence against women and domestic violence[71], which includes explicit data requirements. This new law obligates all member states to systematically collect, produce and disseminate statistics on violence against women (VAW) and domestic violence. At a minimum, countries must gather data on reported incidents, convictions, femicides (women killed due to gender violence), shelter capacity, and helpline calls, disaggregated by sex, age, and relationship to the victim [24]. To ensure consistency, states are urged to harmonize data with common standards developed by the European Institute for Gender Equality (EIGE) [6], [22], [35] Of importance, the Directive also requires that these statistics be made public in an easily accessible manner (with no personally identifiable data) [70]. This effectively mandates open data on GBV across the EU. This marks a significant policy step linking open data and GBV prevention.
2) **EU Data Collection Initiatives:** Even before the 2024 Directive, EU bodies recognized gaps in GBV data. The EU Fundamental Rights Agency's landmark 2014 survey (updated in 2024) [72] was the first EU-wide prevalence study, addressing the "lack of comprehensive and comparable data on violence against women" across Member States In 2020, EIGE coordinated the first-ever EU-level compilation of administrative GBV data (police and justice statistics on intimate partner violence, rape, femicide, etc.). This effort produced 13 harmonized indicators and highlighted the challenges of aligning legal definitions and recording practices across countries. EIGE's statistics on intimate partner violence mark the first time administrative data have been collected and released at the EU level, a critical step toward exposing data gaps and informing policy [73], [74]. The exercise revealed serious comparability issues – for example, not all countries recognize psychological or economic violence similarly, underscoring the need for common definitions and training for data providers. EIGE's report urged Member States to improve and openly publish their GBV data. It noted that few countries currently make detailed GBV data publicly available, and recommended investing in dynamic databases with standardized metadata to make such information accessible and useful.
   These EU-driven initiatives (surveys and administrative data harmonization) laid the groundwork that the 2024 Directive now builds into a binding obligation to open up GBV data. Table 2 summarizes key open data sources on GBV in the EU, categorized by type, with their sources and noted gaps in coverage.

Open data has emerged as a vital component in the fight against gender-based violence, offering transparency, evidence, and new avenues for innovation. In the EU, a robust policy framework is taking shape – from the Open Data Directive to the groundbreaking 2024 anti-GBV Directive – aiming to make GBV data more available and actionable than ever before. The benefits of open data in GBV prevention – identifying hidden trends, evaluating what works, empowering civil society – are profound. Yet, this must be balanced with careful governance to protect privacy, ensure security, and address ethical pitfalls in data-driven solutions. Bridging the remaining gaps will require continued investment in data infrastructure, cross-sector collaboration, and a steadfast commitment to the principle that open data for GBV prevention should ultimately serve and safeguard those whom GBV affects most.

TABLE II: Summary of Available Open Data and Gaps

Available Open Data and Gaps			
Data Type	***Examples of Open Data Sources***	***Coverage and Content***	***Key Gaps / Limitations***
Law Enforcement Data	Eurostat Crime Statistics (police-recorded offenses); National police statistics via NSOs.	EU-wide data on reported crimes (e.g., rape, sexual assault, homicide). Includes some sex-disaggregated info (e.g., female homicide victims by partner/family). Most countries report annually to Eurostat.	**-Incomplete reporting:** Some countries fail to report certain indicators, requiring estimates **-Inconsistent legal definitions:** Cross-country comparisons are difficult due to varying criminal codes **-Limited victim–perpetrator relationship data:** Only about half of EU countries record this, making intimate partner violence hard to identify **-Under-reporting:** Many GBV incidents never reach police, so the actual prevalence is higher than recorded
Victim Services Data	EIGE Victim Support Study (data on shelters and hotlines); National helpline statistics (ministries or NGOs); WAVE country reports.	Info on availability and use of support services. EIGE's study tallied shelter beds, counseling centers, etc. Some national data on hotline call volumes (spikes during COVID lockdowns).	**-No EU-wide data system:** Fragmented by country/organization; no single repository for helpline or shelter usage **-Uneven reporting:** Some Member States publish detailed stats, others do not **-Limited comparability:** Differing definitions of "GBV service" or service usage metrics **-Accessibility issues:** Much data is held by NGOs or local authorities and not publicly available
Legal/Judicial Data	Eurostat Justice Statistics (cases, prosecutions, convictions); National court statistics on DV/VAW cases; EIGE administrative data mapping	Some data on legal outcomes for GBV crimes. Eurostat collects number of persons prosecuted/convicted (by sex of offender/victim). A few countries track domestic violence cases separately.	**Sparse, non-specific data:** Many countries do not disaggregate by the victim's gender or relationship, making IPV cases harder to count **Missing prosecution/conviction numbers:** Not all Member States report these consistently to Eurostat **Legal differences:** "Domestic violence" is not a distinct criminal charge in all jurisdictions, complicating data collection **Invisible attrition:** Systematic data on case drop-off (from report to verdict) is rarely published
Social and Health Data	Health records (e.g., hospital admissions for assault injuries); Social services data (shelters, child protection); WHO/academic estimates.	Primarily national-level data; no unified EU dataset. Examples: hospital ERs noting domestic assault cases, surveys of health impacts, number of women receiving support. WHO publishes regional IPV estimates.	**-No routine EU-wide collection:** Hospitals and social services often lack standardized reporting on GBV **-Privacy barriers:** Confidentiality laws restrict sharing identifiable health data **-Absence of standard codes:** Many health systems do not label cases as "domestic violence," leading to under-capture **-Data quality issues:** Where social service data exists, it may not be open, standardized, or regularly updated
Surveys and Research	EU-GBV Survey (Eurostat, 2020s); FRA Violence Against Women Survey (2014); Eurobarometer on VAW (2016, 2023); National prevalence surveys	Broad coverage of GBV prevalence and attitudes. Large-scale surveys capturing self-reported experiences of physical, sexual, psychological violence, stalking, and harassment.	**-Infrequency:** EU-wide surveys can have a 7+ year gap, risking outdated data **-Partial, phased release:** Not all Member States' data becomes available simultaneously **-Methodological variation:** Earlier national surveys differ in approach; only recent EU-harmonized surveys enable robust cross-country comparison **-Scope limitations:** Some surveys focus only on women, leaving men and LGBTQ+ victims under-examined **-Need for updates:** Regular data collection is crucial, yet often hampered by resource constraints

VII. Critical Gaps and Future Directions in the EU Context

Despite progress in both technology and policy, there are still significant gaps in the EU's strategy to leverage AI for preventing GBV among youth. Identifying these gaps is crucial for informing future research, funding, and regulations:

- **Evidence of Effectiveness:** One fundamental gap is the lack of longitudinal evidence on what works. Many AI initiatives (chatbots, risk assessment tools, etc.) are in pilot phases or small-scale trials. There is "no clear evidence that these systems work" at scale to reduce violence. For example, does an AI risk score lead to measurable reductions in repeat offenses? Does a youth-focused chatbot demonstrably increase reporting or decrease victimization over time? These outcomes are hard to measure and often have not been rigorously evaluated. EU projects tend to publish deliverables and reports, but independent evaluations remain scarce. Without solid evidence, it's challenging for policymakers to justify rolling out these tools nationally or EU-wide. This points to a need for more research and data sharing – possibly creating EU-wide repositories of anonymized case outcomes to evaluate AI interventions. Additionally, any unintended negative effects (such as victims feeling discouraged by an automated system or perpetrators learning to game AI detection) need to be studied. Bridging this gap will require collaborative efforts between technologists, social scientists, and front-line service providers to monitor and assess AI tools in real-world settings.
- **Open Data, Integration and Scaling Issues:** Many current solutions exist in silos – a chatbot here, a police algorithm there – without integration into a coherent service ecosystem. A young GBV survivor in Europe might interact with multiple systems (social media reporting, school counselor, police, hotline), but if these systems' AI tools don't connect, information falls through the cracks. The EU has not yet developed standards for interoperability in this domain. For instance, could a risk assessment AI used by police be safely and ethically linked with a shelter's case management system to alert them of high-risk cases (with consent)? At present, such integration is rare. Scaling successful pilots across languages and cultures is another hurdle. A tool developed in English in one country may not easily transfer to another due to language nuances or different legal frameworks. The SHIELD project's tools will need adaptation as they move from development (with partners in Italy, Germany, Spain, Greece, and the Netherlands) to potential use by others.
- The gap here is a lack of a "playbook" or infrastructure for scaling up AI for social good. EU policy could encourage creating shared platforms or public-private partnerships so that effective AI tools (e.g. a proven harassment-detecting AI model) can be deployed widely, rather than each country reinventing the wheel or relying on small NGOs to sustain tools after project funding ends.
- **Policy-Technology Disconnect:** While EU policies are strong on paper, there can be a lag or disconnect in implementation on the ground. For instance, the new Directive (2024/1385) requires robust prevention measures, but many Member States might struggle to translate that into concrete tech-based services. There is a gap in guidance on how to use AI or data for prevention in practice. Policymakers may not be fully aware of the state-of-the-art in AI, while tech developers might not be versed in the legal duties and victim rights. This calls for interdisciplinary collaboration – bringing together youth advocates, legal experts, AI developers, and law enforcement to co-create solutions. The EU could facilitate this by establishing working groups or innovation sandboxes focused on "AI against GBV" to ensure policy and tech advance in step. Another facet of this disconnect is resource disparity: a well-resourced country like France or Germany might adopt advanced systems, whereas smaller countries or those with less digital capacity lag behind, creating uneven protection for EU citizens. The EU's principle of equality means all victims across Member States should benefit from innovations; addressing this may require dedicated EU funding to help under-resourced areas implement proven AI tools (similar to how the EU supports digitalization in justice systems).
- **Ethical and Legal Clarity:** Despite general ethical guidelines, there are still gray areas where practitioners might be unsure how to proceed. For example, if an AI detects a teen sending alarming messages (maybe indicating they're experiencing dating violence), at what point can/should that data be shared with authorities or parents under EU privacy laws? Youth advocates worry about confidentiality for minors, whereas child protection laws may mandate reporting imminent harm. Clear protocols balancing these concerns are not yet established for AI-mediated information. Similarly, the AI Act's requirements will need to be interpreted for specific GBV tools: What constitutes "human oversight" in a victim support chatbot? How can meaningful transparency for a complex model be achieved when explaining risk to a victim? These are gaps the regulatory technical standards (still to be developed after the AI Act) will need to fill. There's also the issue of liability – if an AI tool funded by the EU causes harm, current laws (like the Product Liability Directive) might not adequately cover software algorithms. The Commission has proposed an AI Liability Directive to ease the burden of proof for those harmed by AI, which will complement the AI Act. Stakeholders in the GBV sector should engage with these legal developments to ensure that, for instance, survivors can seek recourse if an AI failure contributed to them not getting timely help. Filling these gaps in ethical and legal

clarity will boost confidence among service providers to adopt AI tools, knowing they have a clear mandate and protection when using them responsibly.

- **Inclusion and Reach:** As noted, not all youth are being reached by current AI solutions. Marginalized groups – such as undocumented migrants, young people with disabilities, and those in very rural or economically disadvantaged settings – might not benefit from flashy new apps or internet-based tools. This is an equity gap. The EU's digital inclusion efforts (broadband expansion and digital skills training) need to intersect with GBV prevention so that no one is left behind. Moreover, including youth voices in the design of these AI systems is still not the norm. Youth advocacy groups often lack the funding or access to engage with AI research projects. One way to address this gap is through participatory design workshops and funding requirements: EU calls for proposals could require evidence of youth co-design or partnerships with survivor-led organizations. If the end-users feel a sense of ownership and trust in the tools (because they helped shape them), the uptake will be better. Additionally, as generative AI evolves, there is a gap in awareness among both youth and policymakers about its risks and opportunities. Many youths may not be aware of the concept of deepfakes or how AI could be misused against them (e.g., someone making a fake nude image to blackmail a classmate). Incorporating AI literacy into school curricula – as part of comprehensive sexuality and digital education – could empower young people to navigate these emerging threats. The EU could encourage Member States to integrate such content, aligning with the safe internet programs. On the flip side, youth should also learn how AI might be used for their safety, such as awareness that platforms are using algorithms to detect harmful content and how they can report issues effectively to trigger those systems. Bridging the inclusion and knowledge gap is essential to ensure AI tools are not just available but also accessible and trusted by those who need them most.

Despite the promise shown by AI tools (predictive analytics, chatbots, social media monitoring) in detecting risks and supporting victims (see Fig. 4), the EU still faces systemic gaps that hinder full-scale implementation for youth-focused GBV prevention. Most critically, there is minimal longitudinal evidence to prove whether these AI-driven solutions reduce repeat offenses or raise help-seeking behaviors. Open data practices remain fragmented, limiting interoperability across police, shelters, and educational settings, particularly for cross-border coordination. While progressive directives exist, policymakers often struggle to provide clear guidance for the real-world deployment of AI tools, resulting in a policy-technology disconnect. Ethical and legal uncertainties, such as safeguarding minors' privacy and ensuring liability coverage, further complicate adoption, and marginalized youth risk being excluded from digital solutions that assume robust infrastructure or high digital literacy. Strengthening evidence, establishing consistent data-sharing frameworks, and embedding youth perspectives in design can help the EU leverage AI effectively for GBV prevention without sacrificing rights or reach.

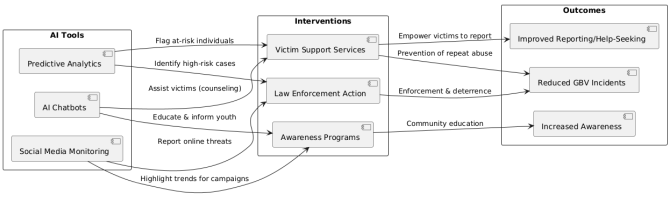

Fig. 4: AI Applications and Interventions in GBV Prevention

VIII. Limitations

Despite the comprehensive approach adopted in this study, several limitations should be acknowledged. First, the review is restricted to sources published between 2019 and 2024 and written in English; therefore, relevant research published outside this time frame or in other languages may not be represented. Second, although a systematic desk review was employed, the heterogeneity of the selected sources—in terms of methodologies, study contexts, and reporting styles—complicates direct comparisons between studies and may affect the generalizability of the synthesized findings. Third, many of the reviewed AI applications for GBV prevention are still in pilot phases or small-scale trials, resulting in limited longitudinal evidence to robustly assess long-term effectiveness or unintended negative consequences. Fourth, while open data practices formed a central criterion in the review, the quality and accessibility of open data differ significantly across EU Member States, and fragmented reporting standards may have resulted in the omission of potentially relevant data-driven studies. Finally, the inclusion of a range of grey literature and policy documents introduces variability in the rigor and peer-review status of the sources, which might impact the robustness of the conclusions drawn. These limitations underline the need for further research, especially studies that offer longitudinal evidence and more standardized methods of data collection and reporting across the EU.

IX. Conclusion

This review aimed to address three central research gaps regarding the use of AI in youth-focused gender-based violence (GBV) prevention within the EU:

1) **Technical Efficacy:**
   AI tools such as predictive analytics and social media monitoring offer promising capabilities in the early detection of GBV risk factors among youth. However, their current application is predominantly limited to pilot projects and small-scale deployments. The review reveals a significant shortage of longitudinal evidence to demonstrate the sustained effectiveness of these interventions.

In addition, the lack of interoperability among disparate systems (including law enforcement, social services, and Victim support services) and inconsistent data standards further constrain the scalability and replicability of these AI models across different Member States.

2) **Ethical Risks:**
The deployment of AI in sensitive, high-stakes contexts raises substantial ethical challenges. A primary issue is the tension between fostering open data transparency and protecting survivors' privacy. Many Member States publish only aggregated or heavily anonymised datasets, which limits the potential for detailed analysis and algorithmic refinement. Furthermore, algorithmic bias remains a critical concern—historical datasets often underrepresent vulnerable groups, leading to potential misclassification and inequitable intervention outcomes. Additionally, the processes for determining when and how to escalate sensitive information in automated systems remain underdefined, undermining youth agency despite existing mandates for human oversight.
3) **Policy Alignment:**
The EU's regulatory framework, embodied in instruments such as the GDPR, the forthcoming AI Act, and the 2024 GBV Directive, provides a robust foundation for balancing technological innovation with ethical safeguards. However, discrepancies persist in the practical implementation of these policies. Variability among Member States in translating high-level directives into cohesive, technology-driven services raises concerns regarding liability, oversight, and enforcement. Moreover, the interplay between open-data initiatives and privacy regulations continues to generate conflicts that hinder the development of standardized protocols for data sharing and effective use of AI in GBV prevention.Efforts such as the recently introduced Public consultation on draft Council of Europe Recommendation on accountability for technology-facilitated violence against women and girls is a move in the right direction to curb this [75].

In summary, while significant progress has been achieved in leveraging AI for GBV prevention among youth, additional longitudinal evaluations, improved open data integration, and standardized ethical and regulatory guidelines are needed to bridge the identified gaps. Future research should focus on establishing interoperable, evidence-based models and fostering interdisciplinary collaboration to ensure that AI-driven interventions are both effective and equitable. By addressing these gaps, the EU can better align technological innovation with comprehensive governance frameworks, ultimately enhancing the protection of vulnerable youth against GBV.

A. Moving Forwad

Going forward, these are the recommendations for AI in GBV prevention.

- Strengthen Evidence and Data Infrastructure: Expanding open, anonymized datasets on helpline usage, judicial outcomes, and long-term AI pilot results can shed light on what truly works. This requires both funding mechanisms for consistent data collection and robust safeguards to protect survivors' identities.
- Prioritize Ethical Design and Youth Participation: Placing youth at the center of AI development—via co-design workshops, iterative feedback, and dedicated funding—can reduce technology resistance and ensure tools genuinely meet young people's needs. AI risk assessments and chatbots should reflect ethical guidelines that weigh autonomy, privacy, and duty of care.
- Enhance Policy-Technology Synergy: EU institutions, Member States, and front-line practitioners must collaborate to close the "policy-technology disconnect." Clarifying how the AI Act and the GBV Directive intersect - particularly on issues such as mandatory human oversight, liability, and data-sharing protocols - will help align innovation with youth protection.
- Foster Interoperability and Scalability: Successful AI solutions remain scattered across member states and often stalled at the pilot stage. Encouraging standardized API frameworks, reference architectures, and cross-border partnerships can help replicate proven tools more broadly. This reduces duplication and ensures that smaller countries can benefit from the lessons learned by others.

By addressing these dimensions, technical, ethical, and policy-related AI can evolve into a more reliable, transparent, and youth-empowering instrument for GBV prevention. The EU, with its rights-based ethos and emerging open-data mandates, is well positioned to lead in harmonizing innovation with robust ethical safeguards. Doing so will accelerate progress toward a future where AI is not just a cutting-edge concept but a trusted ally in creating safer spaces for every young person across Europe.

ACKNOWLEDGMENT

This paper is prepared as part of the SHIELD – European AI-Powered Gender-Based Violence Reduction Initiative (Project No. 2024-1-IT03-KA220-YOU-000250850), funded by the European Union and implemented from 01/10/2024 to 30/09/2027. The findings and recommendations herein reflect the ongoing collaborative efforts of the project partners to harness AI responsibly in preventing and responding to gender-based violence among youth in the EU.

REFERENCES

[1] KMOP, "European AI-Powered Gender Based Violence Reduction Initiative," Accessed: Mar. 10, 2025. [Online]. Available: https://www.kmop.gr/projects-vf/news-shield/

[2] L. Fernández, P. Alvarez-Cueva, and M. J. Masanet, "From sexting to sexpreading: Trivialization of digital violence, gender differences and collective responsibilities," *Sex Cult*, pp. 1–33, Jan. 2025, doi: 10.1007/S12119-025-10316-5.

[3] E. Quilty and A. Flynn, "Technology-facilitated violence in the Indo-Pacific: A scoping review," *Trauma, Violence, & Abuse*, Feb. 2025, doi: 10.1177/15248380251323217.

[4] European Commission, *Regulation of the European Parliament and of the Council Laying Down Rules to Prevent and Combat Child Sexual Abuse*. Brussels: European Law, 2022.

[5] P. K. Trickett, J. G. Noll, and F. W. Putnam, "The impact of sexual abuse on female development: Lessons from a multigenerational, longitudinal research study," *Dev. Psychopathol.*, vol. 23, no. 2, pp. 453–476, May 2011, doi: 10.1017/S0954579411000174.

[6] European Commission, "Gender Equality Report Chapeau Communication," *Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions*, 2025.

[7] European Commission, "The EU's digital services act." Accessed: Mar. 10, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

[8] A. Valdivia, C. Hyde-Vaamonde, and J. García Marcos, "Judging the algorithm: Algorithmic accountability on the risk assessment tool for intimate partner violence in the Basque Country," *AI Soc.*, pp. 1–18, Jul. 2024, doi: 10.1007/S00146-024-02016-9.

[9] A. Wood, "Dark echoes: The exploitative potential of generative AI in online harassment," *Psybersecurity: Human Factors of Cyber Defence*, pp. 90–129, Sep. 2024, doi: 10.1201/9781032664859-5/DARK-ECHOES-ADRIAN-WOOD.

[10] S. Shashidhara, P. Mamidi, S. Vaidya, and I. Daral, "Using machine learning prediction to create a 15-question IPV measurement tool," *J. Interpers. Violence*, vol. 39, no. 1–2, pp. 11–34, Jan. 2024, doi: 10.1177/08862605231191187.

[11] J. Kim, "Ethical challenges in artificial intelligence deployment," *Int. J. Technol.*, vol. 10, no. 10, Jan. 2025, doi: 10.5281/TC7MZK60.

[12] I. R. Berson, M. J. Berson, and W. Luo, "Innovating responsibly: Ethical considerations for AI in early childhood education," *AI, Brain and Child*, vol. 1, no. 1, p. 2, Mar. 2025, doi: 10.1007/s44436-025-00003-5.

[13] European Commission, "Viogen 5.0: discovering Spain's risk assessment system of gender-based violence — Interoperable Europe Portal," *Public Sector Tech Watch - TECH*. Accessed: Mar. 13, 2025. [Online]. Available: https://interoperable-europe.ec.europa.eu/collection/public-sector-tech-watch/viogen-50-discovering-spains-risk-assessment-system-gender-based-violence

[14] N. Mylonas *et al.*, "Online child grooming detection: Challenges and future directions," presented at the *International Conference on ...*, 2025, pp. 237–247, doi: 10.1007/978-3-031-62083-6_19.

[15] European Union, *Regulation 2016/679 - General Data Protection Regulation (GDPR)*. 2016. Accessed: Mar. 14, 2025. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[16] European Union, *Regulation (EU) 2024/1689 - Artificial Intelligence Act*. 2024. Accessed: Mar. 14, 2025. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

[17] K. Brookfield, R. Fyson, and M. Goulden, "Technology-facilitated domestic abuse: An under-recognised safeguarding issue?," *Br. J. Soc. Work*, vol. 54, no. 1, pp. 419–436, Jan. 2024, doi: 10.1093/BJSW/BCAD206.

[18] UNESCO, "Fight against technology-facilitated gender-based violence — UNESCO." Accessed: Mar. 13, 2025. [Online]. Available: https://www.unesco.org/en/articles/fight-against-technology-facilitated-gender-based-violence

[19] United Nations, "How technology-facilitated gender-based violence impacts women and girls." Accessed: Mar. 13, 2025. [Online]. Available: https://unric.org/en/how-technology-facilitated-gender-based-violence-impacts-women-and-girls/

[20] European Union, "Ethics guidelines for trustworthy AI — Shaping Europe's digital future," 2019. Accessed: Mar. 13, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[21] E. Kelan, *Patterns of Inclusion: How Gender Matters for Automation, Artificial Intelligence and the Future of Work*. London, U.K.: Routledge, 2024, doi: 10.4324/9781003427100.

[22] European Commission, "Towards a gender-equal Europe," 2020. [Online]. Available: https://eige.europa.eu/thesaurus/terms/1263

[23] Europol, "How AI can strengthen law enforcement: Insights from Europol's new report," Accessed: Mar. 13, 2025. [Online]. Available: https://www.europol.europa.eu/media-press/newsroom/news/how-ai-can-strengthen-law-enforcement-insights-europols-new-report

[24] P. Office of the European Union L. and L. Luxembourg, "Directive (EU) 2024/1712 of the European Parliament and of the Council of 13 June 2024 amending Directive 2011/36/EU on preventing and combating trafficking in human beings and protecting its victims," 2024. [Online]. Available: http://data.europa.eu/eli/dir/2024/1260/oj

[25] European Union, *Charter of Fundamental Rights of the European Union*. 2017.

[26] A. Tang, *Safeguarding the Future: Security and Privacy by Design for AI, Metaverse, ...*. CRC Press, Accessed: Mar. 12, 2025. [Online]. Available: https://books.google.de/books?hl=en&lr=&id=xDVEEQAAQBAJ

[27] A. Thomaidou and K. Limniotis, "Navigating through human rights in AI: Exploring the interplay between GDPR and fundamental rights impact assessment," *J. Cybersecurity Privacy*, vol. 5, no. 1, p. 7, Feb. 2025, doi: 10.3390/JCP5010007.

[28] J. J. López-Ossorio, J. L. González-Álvarez, J. M. Muñoz Vicente, C. Urruela Cortés, and A. Andrés-Pueyo, "Validation and calibration of the Spanish police intimate partner violence risk assessment system (VioGén)," *J. Police Crim. Psychol.*, vol. 34, no. 4, pp. 439–449, Dec. 2019, doi: 10.1007/S11896-019-09322-9.

[29] T. Swan, "Bias in the bot: AI's relationship to gender-based violence," *NoMore.org*, Accessed: Mar. 13, 2025. [Online]. Available: https://www.nomore.org/bias-in-the-bot-ais-relationship-to-gender-based-violence/

[30] M. S. Khan, H. Umer, and F. Faruqe, "Artificial intelligence for low-income countries," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, pp. 1–13, Oct. 2024, doi: 10.1057/s41599-024-03947-w.

[31] CESIE ETS, "SHIELD – European AI-Powered Gender Based Violence Reduction Initiative," Accessed: Mar. 13, 2025. [Online]. Available: https://cesie.org/en/project/shield/

[32] European Commission, "Gender equality strategy," Accessed: Mar. 14, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en

[33] European Parliament, "AI technologies must prevent discrimination and protect diversity — News," Accessed: Mar. 13, 2025. [Online]. Available: https://www.europarl.europa.eu/news/en/press-room/20210311IPR99709/ai-technologies-must-prevent-discrimination-and-protect-diversity

[34] European Union, "Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment — Shaping Europe's digital future," 2020. Accessed: Mar. 13, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

[35] European Union, "Ending gender-based violence - European Commission," May 2024. Accessed: Mar. 13, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence/ending-gender-based-violence_en

[36] European Union Institute for Gender Equality, "Digitalisation and equal rights – the role of AI algorithms," *Gender Equality Index 2020: Digitalisation and the future of work*. Accessed: Mar. 13, 2025. [Online]. Available: https://eige.europa.eu/publications-resources/toolkits-guides/gender-equality-index-2020-report/digitalisation-and-equal-rights-role-ai-algorithms

[37] European Union, *Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on Combating Violence Against Women and Domestic Violence*. 2024. Accessed: Mar. 13, 2025. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2024/1385/oj/eng

[38] European Commission, "Elimination of violence against women: Tackling a global crisis - European Commission," 2024. Accessed: Mar. 13, 2025. [Online]. Available: https://rea.ec.europa.eu/news/elimination-violence-against-women-tackling-global-crisis-2024-11-22_en

[39] European Union, "Council of Europe convention on preventing and combating violence against women and domestic violence," 2011.

[40] I. Rodríguez-Rodríguez, J. V. Rodríguez, D. J. Pardo-Quiles, P. Heras-González, and I. Chatzigiannakis, "Modeling and forecasting gender-based violence through machine learning techniques," *Appl. Sci.*, vol. 10, no. 22, p. 8244, Nov. 2020, doi: 10.3390/APP10228244.

[41] European Commission, "The EU code of conduct on countering illegal hate speech online," Accessed: Mar. 13, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

[42] M. Heikkilä, "AI: Decoded: Spain's flawed domestic abuse algorithm — Ban debate heats up — Holding the police accountable," *Politico*, Accessed: Mar. 13, 2025. [Online]. Available: https://www.politico.eu/newsletter/ai-decoded/spains-flawed-\domestic-

abuse-algorithm-ban-debate-heats-\protect\discretionary{\char\
hyphenchar\font}{}{}up-holding-the-police-accountable-2/

[43] R. Lévy, "From tagging to tracking," in *Electronically Monitored Punishment: International and Critical Perspectives*, M. Nellis, K. Beyens, and D. Kaminski, Eds. New York, NY, USA: Routledge, 2013, pp. 100–114. Accessed: Mar. 13, 2025. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203103029-9/tagging-tracking-ren%C3%A9-l%C3%A9vy

[44] European Commission, "The EU strategy on the rights of the child and the European child guarantee - European Commission," 2021. Accessed: Mar. 13, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/rights-child/eu-strategy-rights-child-and-european-child-guarantee_en

[45] "AI is being used to assess which victims of domestic violence are most at risk – what's the risk? — AI Summer School," Accessed: Mar. 13, 2025. [Online]. Available: https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/ai-domestic-violence

[46] European Commission, "The EU's digital services act," Accessed: Mar. 13, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

[47] Y. Register, "The future of AI can be kind: Strategies for embedded ethics in AI education," unpublished manuscript.

[48] "Innovative solutions to eliminate domestic abuse," Sep. 2022, doi: 10.3030/101073922.

[49] "IMPRODOVA — Improving frontline responses to high impact domestic violence," Accessed: Mar. 13, 2025. [Online]. Available: https://www.improdova.eu/

[50] R. Ivandić, "Artificial intelligence could help protect victims of domestic violence," 2020. Accessed: Mar. 13, 2025. [Online]. Available: https://www.lse.ac.uk/News/Latest-news-from-LSE/2020/b-Feb-20/Artificial-intelligence-could-help-protect-victims-of-domestic-violence

[51] N. M. Navaneeth, "The need for a global regulatory framework for artificial intelligence: Implications of the European union's artificial intelligence act 2024," 2024.

[52] K. Thompson, "Gender-based violence (GBV) against immigrant women living in Canada: Blending big data and critical discourse approaches to news media representations," M.A. thesis, Saint Mary's University, Halifax, Canada, May 2022. Accessed: Mar. 13, 2025. [Online]. Available: https://library2.smu.ca/xmlui/handle/01/30953

[53] M. Nellis, K. Beyens, and D. Kaminski, *Electronically Monitored Punishment: International and Critical Perspectives*. New York, NY, USA: Routledge, May 2013, doi: 10.4324/9780203103029.

[54] N. G. Packin and K. Chagal-Feferkorn, "This is not a game: The addictive allure of digital companions," Feb. 2025, doi: 10.2139/SSRN.5133614.

[55] Q. Wong, "Teens are forming bonds with AI chatbots, raising concerns," *Los Angeles Times*, accessed Mar. 13, 2025. [Online]. Available: https://www.latimes.com/business/story/2025-02-25/teens-are-spilling-dark-thoughts-to-ai-chatbots-whos-to-blame-when-something-goes-wrong

[56] A. Karunwi, "Violence among Nigeria's female and predictors of help-seeking behavior using technology," Ph.D. dissertation, Walden Univ., Minneapolis, MN, USA, 2025. Accessed: Mar. 13, 2025. [Online]. Available: https://www.proquest.com/openview/8f02fe561aaaa05ece10881c3f2d2dd6/1?cbl=18750&diss=y&pq-origsite=gscholar

[57] I. Adam, "Brain Builders unveils AI tool against online GBV at Paris summit," *The Nation Newspaper*, Accessed: Mar. 13, 2025. [Online]. Available: https://thenationonlineng.net/brain-builders-unveils-ai-tool-against-online-gbv-at-paris-summit/

[58] S. A. Pasha, S. Ali, and R. Jeljeli, "Artificial intelligence implementation to counteract cybercrimes against children in Pakistan," *Human Arenas*, vol. 8, no. 1, pp. 79–97, Oct. 2022, doi: 10.1007/S42087-022-00312-8.

[59] World Justice Project, "rAInbow: Chatbot to support victims of domestic abuse — World Justice Project," Accessed: Mar. 13, 2025. [Online]. Available: https://worldjusticeproject.org/world-justice-challenge-2021/rainbow-chatbot-support-victims-domestic-abuse

[60] D. Manoj *et al.*, "Behind the screens: Understanding the gaps in India's fight against online child sexual abuse and exploitation," *Child Protection and Practice*, vol. 4, p. 100088, Apr. 2025, doi: 10.1016/J.CHIPRO.2024.100088.

[61] Kenyan Parliament, *The Computer Misuse and Cybercrime (Amendment) Bill*. 2024.

[62] UN Women, "Artificial intelligence and gender equality — UN Women," Accessed: Mar. 13, 2025. [Online]. Available: https://www.unwomen.org/en/articles/explainer/artificial-intelligence-and-gender-equality

[63] V. Mendizabal Joffre and P. Serafica, "Chatbots offer a new lifeline to address domestic violence," *Asian Development Blog*, Accessed: Mar. 13, 2025. [Online]. Available: https://blogs.adb.org/blog/chatbots-offer-new-lifeline-address-domestic-violence

[64] N. Koukopoulos, M. Janickyj, and L. M. Tanczer, "Defining and conceptualizing technology-facilitated abuse ('tech abuse'): Findings of a global delphi study," *J. Interpers. Violence*, 2025, doi: 10.1177/08862605241310465.

[65] J. Bailey, N. Henry, and A. Flynn, "Technology-facilitated violence and abuse: International perspectives and experiences," in *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*, Bingley, U.K.: Emerald, 2021, pp. 1–17, doi: 10.1108/978-1-83982-848-520211001.

[66] A. Bayne, E. A. Mumford, C. Lancaster, and J. Sheridan-Johnson, "Technology-facilitated abuse among Americans age 50 and older: A latent class analysis," *J. Elder Abuse Negl.*, vol. 35, no. 1, pp. 65–87, 2023, doi: 10.1080/08946566.2023.2197270.

[67] C. Brown and K. Hegarty, "Development and validation of the TAR scale: A measure of technology-facilitated abuse in relationships," *Comput. Hum. Behav. Rep.*, vol. 3, Jan. 2021, doi: 10.1016/J.CHBR.2021.100059.

[68] J. Goodey, "Violence against women: Placing evidence from a European Union-wide survey in a policy context," *J. Interpers. Violence*, vol. 32, no. 12, pp. 1760–1791, Jun. 2017, doi: 10.1177/0886260517698949.

[69] F. Wang, J. Jiang, and F. Cosenz, "Understanding data-driven business model innovation in complexity: A system dynamics approach," *J. Bus. Res.*, vol. 186, p. 114967, Jan. 2025, doi: 10.1016/J.JBUSRES.2024.114967.

[70] European Union, *Directive (EU) 2019/1024 of the European Parliament and of the Council on Open Data and the Re-Use of Public Sector Information*. 2019. Accessed: Mar. 14, 2025. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng

[71] P. Office of the European Union L. and L. Luxembourg, "Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence," 2024. [Online]. Available: http://data.europa.eu/eli/dir/2024/1385/oj

[72] European Union Agency for Fundamental Rights, *Violence against women: An EU-wide survey: Main results*. 2014, doi: 10.2811/98192.

[73] European Institute for Gender Equality, *Indicators on Intimate Partner Violence and Rape for the Police and Justice Sectors*. 2018, doi: 10.2839/252486.

[74] E. Kelan, *Patterns of Inclusion*. London, U.K.: Routledge, 2024, doi: 10.4324/9781003427100.

[75] Council of Europe, *"Public consultation on draft Council of Europe recommendation on accountability for technology-facilitated violence against women and girls is now open*. Public consultation, Council of Europe, accessed Apr. 23, 2025. [Online]. Available: https://www.coe.int/en/web/genderequality/-/public-consultation-on-draft-council-of-europe-recommendation-on-accountability-for-technology-facilitated-violence-against-women-and-girls-is-now-open