

POLICY BRIEF N° IDB-PB-422

Innovations with Artificial Intelligence

An Exploratory Mapping in the Response to Violence Against Women

Agustina Suaya
Montserrat Bustelo
Karen Martínez

Inter-American Development Bank
Gender and Diversity Division
Department of Research and Chief Economist

November 2025



Innovations with Artificial Intelligence

An Exploratory Mapping in the Response to Violence Against Women

Agustina Suaya
Montserrat Bustelo
Karen Martínez

Inter-American Development Bank
Gender and Diversity Division
Department of Research and Chief Economist

November 2025



**Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library**

Bustelo, Monserrat.

Innovations with artificial intelligence: an exploratory Mapping in the response to violence against women / Monserrat Bustelo, Karen Martínez, Agustina Suaya.

p. cm. — (IDB Policy Brief ; 422)

Includes bibliographical references.

1. Women-Violence against-Latin America-Prevention. 2. Women-Violence against-Caribbean Area-Prevention. 3. Women-Crimes against-Latin America-Prevention. 4. Women-Crimes against-Caribbean Area-Prevention. 5. Digital communications-Latin America. 6. Digital communications-Caribbean Area. 7. Artificial intelligence-Social aspects-Latin America. 8. Artificial intelligence-Social aspects-Caribbean Area. I. Martínez, Karen. II. Suaya, Agustina. III. Inter-American Development Bank. Gender and Diversity Division. IV. Inter-American Development Bank. Department of Research and Chief Economist. V. Title. VI. Series.

IDB-PB-422

<http://www.iadb.org>

Copyright © 2025 Inter-American Development Bank ("IDB"). This work is subject to a Creative Commons license CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>). The terms and conditions indicated in the URL link must be met and the respective recognition must be granted to the IDB.

Further to section 8 of the above license, any mediation relating to disputes arising under such license shall be conducted in accordance with the WIPO Mediation Rules. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the United Nations Commission on International Trade Law (UNCITRAL) rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this license.

Note that the URL link includes terms and conditions that are an integral part of this license.

The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.





→ Innovations with Artificial Intelligence: An Exploratory Mapping in the Response to Violence Against Women

Artificial intelligence (AI) has emerged as a transformative tool for public services, with applications that promise to expand access, improve efficiency, and strengthen the State's capacity to respond (Cómez et al., 2020). In the field of violence against women, its use is still incipient but is beginning to be explored at different stages of the institutional response cycle, from initial guidance to post-intervention follow-up.

In Latin America and the Caribbean, where one in three women has experienced physical or sexual violence by an intimate partner (WHO, 2021), structural barriers continue to limit timely and adequate access to protection services. In this context, AI offers significant opportunities but also raises ethical dilemmas, technical challenges, and institutional risks that must be carefully considered.

This policy note presents a preliminary mapping of artificial intelligence applications in public policies aimed at preventing and responding to violence against women in the region. Based on an exploratory review of ongoing initiatives and available literature, it examines at which stages of the institutional response cycle AI-based tools are being incorporated, what functions they serve, and what evidence exists on their outcomes. While the use of these technologies is on the rise, empirical evidence on their effectiveness remains limited

AI offers significant opportunities but also raises ethical dilemmas, technical challenges, and institutional risks that must be carefully considered.

or nonexistent. The study identifies critical knowledge gaps and underscores the need to generate rigorous evaluations that guide implementation, provide a better understanding of their effectiveness, and ensure the protection of users' rights.

→ Preliminary Mapping of Processes with AI Potential

The use of artificial intelligence in public policies addressing violence against women has begun to be explored at different stages of the institutional response cycle, from initial guidance to post-intervention monitoring. Although these experiences are still incipient, they make it possible to identify critical functions where some degree of application or pilot testing of AI-based tools has been documented.

This section examines seven processes within prevention, response, protection, and justice systems, selected for their cross-cutting relevance and their potential to be strengthened through emerging technologies. Each subsection briefly describes the process, discusses possible ways in which artificial intelligence could contribute to its functioning, and presents examples of initiatives already implemented or at an experimental stage.

The selection of cases is neither exhaustive nor does it imply validation of their effectiveness. Its purpose is to provide concrete examples that serve as a basis for reflecting on potential uses, minimum conditions for implementation, and remaining questions regarding their effectiveness.

While priority is given to experiences from Latin America and the Caribbean, international examples that provide relevant lessons for the region are also included. In all cases, an exploratory approach is adopted, not with the aim of proposing definitive solutions, but rather to generate evidence for debate and for building an evidence-based research and public policy agenda.

The use of artificial intelligence in public policies addressing violence against women has begun to be explored at different stages of the institutional response cycle, from initial guidance to post-intervention monitoring.

→ Personalized Guidance and Information Dissemination

Access to clear, confidential, and timely information is a critical first step in the prevention and response pathway to violence against women. Whether for women directly affected or for people in their circles seeking to provide support, receiving proper guidance can facilitate decision-making and help-seeking. Traditionally, this guidance has been offered through telephone hotlines, care centers, or printed materials. However, its reach may be restricted by various personal, social, and structural barriers, such as fear of reprisals, mistrust of institutions, linguistic exclusion, or difficulties in access depending on time or place (Wieberneit et al., 2024; [Roza, Martín, 2021](#)). Even in services operating 24/7, human resources may prove insufficient to respond to demand in a timely manner (González et al., 2023; Garnelo et al., 2019).

In this context, some initiatives have explored the use of digital tools to expand access and improve the quality of initial guidance (López and Izaguirre, 2024). For example, chatbots with natural language processing have been developed to answer frequently asked questions in real time, guide users through the response system by adapting to their specific needs and circumstances, and even detect warning signs. These tools operate anonymously, without time restrictions, and in multiple languages. While they do not replace professional care, they can serve as an accessible entry point and—when integrated into institutional systems—facilitate the referral of more complex cases or contribute to the systematization of operational data.

Various initiatives along these lines have been documented in the region. One example is [Sara](#), an AI-based digital tool that provides confidential and free assistance to women, girls, and adolescents experiencing violence in Central American and Caribbean countries. This tool offers a range of resources such as legal counseling, safety plans, and access to shelters, adapting to local context and language thanks to training provided by specialists in each country (López and Izaguirre, 2024). In 2023, it was integrated into the protection system of the Dominican Republic through the Ministry of Women's [website](#). In 2024, under the leadership of the Secretariat of Security, it was adapted in Honduras under the name MARIA (Automated Information Assistance and Response Module) and made available to the public through the National Police [website](#). It has also been adapted in Belize, Costa Rica, El Salvador, and Guatemala (Universidad de los Andes, 2024).

Other experiences, although not part of a state policy, seek to complement the public offer. One of them is [Violetta](#), a chat available in Mexico through WhatsApp that provides psychoeducational and preventive



support on healthy, violence-free relationships. Combining supervised machine learning models with the support of a professional team (López and Izaguirre, 2024), it offers initial psychological and legal guidance and refers severe cases to human operators. Another tool is [OlimpiA](#), an automated chat designed to guide women and adolescents facing online gender-based violence. Also available via WhatsApp, it uses natural language processing and text and audio analysis algorithms to detect warning signs of crisis, and has been implemented in countries such as Mexico, Colombia, Ecuador, and Guatemala.

These tools are still in the early stages of implementation and, overall, there is little systematized information on their impact on the quality of response based on rigorous studies. Even so, advances in language models have improved their ability to generate empathetic and contextually relevant responses, opening opportunities for their use in initial guidance processes (Sanz, López-Belloso, and Izaguirre, 2025). However, their effectiveness will depend on multiple factors: from algorithm design, professional supervision, privacy protection, and ethical safeguards, to integration with existing services, alignment with institutional systems, and operational sustainability. Addressing these aspects in future research will be key to better understanding the potential contribution of these technologies to guidance processes and access to response services.

→ Case Registration and Referral

Effective response to situations of violence against women requires not only institutional commitment but also technical capacities to properly register cases, refer them in a timely manner, and ensure adequate follow-up throughout the protection system. In practice, survivors often interact with multiple actors—such as the police, social services, the justice system, and health services—and each of these links can introduce frictions, duplications, or gaps in response. In the absence of integrated systems, many women must repeat their story to various institutions, which increases the risk of revictimization and of cases being lost along the institutional journey.

The use of AI-based tools could help reduce these frictions. Natural language processing models, for example, can support the automatic transcription of interviews, structure the information generated during response, or assist in [assigning cases to specialized teams](#). However, more evidence is needed to determine whether AI is effective in these processes and whether its integration improves the efficiency and quality of the response provided.

In the region, some recent experiences have begun to explore these uses. In Mexico, the Secretariat for Substantive Equality between Women and Men of the state of Jalisco, in collaboration with Tecnológico de Monterrey and the Inter-American Development Bank (IDB), developed the tool [Olivia for Her](#). This application uses voice recognition to automatically record women's testimonies in care centers during psychological or legal interviews. The information is transcribed and classified in real time, which helps reduce the operational burden, avoid duplications, and facilitate case follow-up. The initiative was implemented in 2023 as a pilot and, according to official reports, enabled the automation of around [40% of case records](#) during its first year. Although it has not yet been formally evaluated, it stands out as a relevant example of AI application in sensitive contexts, complementing professional work (SISEMH, 2024).

Although these tools have been developed to improve case registration and referral, evidence on their effectiveness remains limited. Most initiatives are at the pilot stage, and there is little public documentation of their results. In addition, their long-term sustainability may be affected by factors such as reliance on external funding or the need for sustained institutional leadership. These conditions should be explored in greater depth in future research in order to better understand the contributions and limitations of using AI at this stage of the response system.

→ Risk Assessment and Revictimization in Registered Cases

Not all cases of violence against women present the same level of urgency or severity. Identifying those at higher risk of escalation—such as serious injuries or femicide—is a critical function of protection and justice systems, as it enables resources and institutional responses to be directed more effectively and in a timely manner (Arbach, 2023).

Historically, risk assessment has relied on the professional judgment of intervening teams. Over time, more structured approaches have been incorporated, such as actuarial and mixed models, which combine algorithms with professional criteria guided by protocols (Muñoz and López-Ossorio, 2016). However, these tools often depend on fixed rules and do not always incorporate information from more complex or dynamic systems.

In this context, some governments have begun to explore the use of AI to support risk assessment in cases already registered in response systems. These initiatives apply machine learning techniques to large volumes of administrative data—such as police, judicial, or clinical records—with the aim of identifying patterns associated with recidivism or lethality. However, their implementation raises ethical, methodological, and institutional challenges. Algorithm-supported decisions must be understandable, auditable, and accompanied by professional supervision, especially when they may affect fundamental rights ([Latorre, De Leo, and Rego, 2025](#); Snow, 2019; OECD, 2019).

The following section describes some documented experiences that illustrate different ways in which AI has begun to be applied to support risk assessment in cases registered in the system. Implemented in various contexts, these initiatives provide initial examples that help explore potential approaches, as well as the associated technical and ethical challenges. While they do not constitute validated models and have not all been thoroughly evaluated, they can provide relevant evidence for future research and policy discussions.

Since 2007, Spain has implemented the VioGén system to support risk assessment, combining administrative records with classification tools. This system centralizes information on survivors and perpetrators for use by security forces and combines structured forms with algorithmic models that guide personalized safety plans. In 2025, it was updated with a new protocol that introduced adjustments to the algorithms and to the categorization of risk levels. Although some initial analyses highlighted its predictive capacity, concerns have also been documented about possible overreliance on automated recommendations and their impact in cases with clear warning signs that were not detected in time (Éticas Foundation, 2024; Satariano and Roser, 2024).

In Taiwan, the Taipei Government developed a machine learning–based model to predict revictimization, with the aim of supporting social workers in resource allocation. The model was trained on data from more than 8,800 cases in the national protection system and, according to the available study, reached an accuracy level above 95% (Hsieh et al., 2018). However, limitations were also identified, such as dependence on historical data or the possibility of biases in reporting.



In Australia, the Queensland Police use a model that classifies perpetrators at high risk of recidivism based on police data. This type of tool has been used to prioritize preventive interventions, such as home visits or referrals to treatment programs. While some reports mention a reduction in incidents among prioritized groups, concerns have also been raised about possible unintended effects, such as increased tensions or discouraging survivors from seeking help (Douglas, 2021; Westendorf, 2022).

These experiences highlight both the potential of AI to support preventive decisions and its limitations. The available evidence does not yet allow for a clear determination of its impact on the effective protec-

tion of survivors. Predictive tools require ongoing validation, professional supervision, adaptation to local contexts, and ethical safeguards that ensure transparency, explainability, and equity ([Latorre et al., 2024](#); [2025](#); [Aránguiz Villagrán, 2022](#)).

→ Active Follow-up and Reminders

UA persistent challenge in responding to violence against women is maintaining contact with survivors over time. Factors such as fear of reprisals, family pressure, economic dependence, or frustration with the institutional response can lead to the abandonment of legal proceedings or the protection system.

Active follow-up is based on the premise that institutions must take on a proactive role, checking on the well-being of survivors, reminding them of appointments or procedures, and promoting their continued engagement in the service pathway. However, traditional mechanisms—such as phone calls or home visits—are resource-intensive, difficult to scale, and, in some cases, may pose additional risks if detected by the perpetrator (Petersen et al., 2022).

In response to these challenges, some initiatives have begun to explore the use of AI tools to complement these functions. For example, systems have been developed that automate the sending of personalized mes-

sages or use predictive models to identify people at higher risk of disengagement, allowing for more timely interventions. These solutions are still in the early stages, and there are no conclusive evaluations of their effectiveness. Nevertheless, they represent an emerging field for exploring new ways to strengthen continuity in access to protection.

An exploratory example in this area is Text-SAFE, an automated monitoring system developed for women who had received medical care following a sexual assault. For 37 days, participants received scheduled messages with appointment reminders, health information, and emotional monitoring. If responses indicated a high level of stress, the system generated an automatic alert for clinical staff. In a pilot with 25 users, 96% responded to at least one message and 84% completed the protocol (Downing et al., 2021). While it was a limited study, it suggested potential benefits in terms of sustained and discreet contact. However, broader reviews of SMS interventions in mental health have shown mixed results, suggesting that their effectiveness is not guaranteed and depends on multiple factors (Clough and Casey, 2014).

In another exploratory study, a research team from the University of Seville trained machine learning models with administrative data from court cases of intimate partner violence in Spain, with the aim of identifying factors associated with the abandonment of legal proceedings. The results suggest that variables such as the survivor's age, the duration of the violence, and the absence of support networks may be correlated with a higher probability of disengagement (Escobar-Linero et al., 2023). While it is a retrospective analysis without institutional validation, the findings could provide evidence for designing targeted follow-up strategies in the future.

These initiatives illustrate possible pathways for exploring how AI could complement institutional monitoring efforts through automated messages or prioritization tools. Their implementation, however, requires careful consideration of aspects such as informed consent, privacy, and the potential impact on survivors' autonomy. The available evidence is still limited, making further studies necessary to understand whether these tools can effectively contribute to improving continuity in access to protection and justice.

→ Early and Proactive Detection

One of the persistent challenges in addressing violence against women is that many cases are not reported in time or never reach institutional channels at all. As a result, a significant share of violence remains invisible or without response.

AI-based technologies have the potential to help detect early or indirect warning signs of violence, even before a formal complaint or explicit request for response is made. These tools can analyze unstructured data—such as clinical notes or online texts—to identify patterns that signal situations of risk. Their purpose is not to replace professional judgment but to complement the work of frontline teams, enabling efforts to be targeted more effectively and in a timely manner. Nevertheless, this is an emerging field with significant ethical, technical, and institutional challenges, requiring mechanisms for validation, auditing, and supervision ([Aránguiz Villagrán, 2022](#)).

The following section describes some exploratory experiences that illustrate different approaches to using AI to detect cases of violence that have not been formally reported. These initiatives are in the pilot phase and allow for discussion of both the potential and the limitations of such applications.

In Mexico, the United Nations Office on Drugs and Crime (UNODC), together with the National Institute of Statistics and local authorities, developed a pilot that applies natural language processing to the analysis of more than 60,000 calls to 911 in Mexico City. The model seeks to identify linguistic and paralinguistic patterns—such as content, rhythm, or tone of speech—that could indicate situations of violence, even if not explicitly verbalized (UNODC, 2024). The objective is to support call prioritization and activate follow-up protocols in the presence of warning signs. Although presented as an innovative experience, it has not yet undergone a systematic evaluation of its impact on improving the institutional response.

In other contexts, several studies have explored the analysis of clinical records. For example, in the United States, one study trained machine learning models with more than 34,000 radiology reports to identify patterns compatible with intimate partner violence in patients who later accessed specialized services. The results suggest that certain indicators present in the reports could anticipate cases of violence up to three years before formal entry into the system, although the model's performance varied by demographic group (Chen et al., 2020). Another study, based on unstructured clinical notes, used natural language processing to detect indirect descriptions of abuse. While the results were technically promising, the authors emphasize the need to adapt these models to the local context and ensure their application under professional supervision (Tabaie et al., 2022).

These initiatives reflect the growing interest in leveraging previously underutilized data to anticipate situations of violence. However, they also highlight the challenges of implementation. Detecting early indicators does not, in itself, guarantee an effective response, especially if alerts are not managed by trained teams or if clear response protocols are lacking. In addition, model training requires representative and contextualized data to avoid errors or biases that may exclude certain population groups (Snow, 2019). The available evidence remains limited, and further research is needed to understand under what conditions these tools can ethically, effectively, and usefully complement early detection efforts within response systems.



→ **Monitoring and Production of Statistics**

Designing effective public policies requires reliable, up-to-date, and disaggregated data. This includes incidence indicators—such as the number of cases, the femicide rate, or calls to helplines—as well as institutional response indicators, such as protection measures issued, average response time, or inter-institutional referrals. Traditionally, the production of these statistics has relied on administrative data collected manually and consolidated in periodic reports, often affected by underreporting, duplications, or lack of interoperability (European Institute for Gender Equality, 2016).

AI could provide useful tools to complement traditional monitoring mechanisms. For example, natural language processing could analyze unstructured records such as clinical notes or police reports, while automated models could classify cases, detect patterns, generate dynamic visualizations, or identify high-risk areas. However, the use of these tools in this field is still incipient, and their usefulness depends on multiple factors, such as the quality of source data, institutional capacity, and integration with existing systems.

Although implemented experiences remain limited, some initiatives make it possible to explore emerging approaches in the use of AI for monitoring and the production of statistics in the field of violence against women. In Colombia, a research team developed a machine learning model to classify the level of risk among women in situations of vulnerability in the municipality of Tumaco. Using data collected by local organizations, the system generated georeferenced neighborhood-level risk maps aimed at guiding resource allocation and targeting preventive strategies. However, no published information is available on its impact on institutional decision-making (Bernal-Monroy et al., 2025).

In Brazil, the FrameNet Brasil project, in collaboration with Vital Strategies, designed a tool to improve the quality of data on violence recorded in the health system. Through the analysis of clinical forms using natural language processing techniques, the system identifies indicators of violence in records that were not explicitly coded as such. The objective of this pilot is not to intervene on a case-by-case basis, but rather to improve aggregate statistics by correcting underreporting and ensuring terminological harmonization, thereby contributing to the construction of more comparable and useful indicators for national monitoring.

In the United Kingdom, an exploratory study applied unsupervised clustering algorithms to more than 40,000 police records from Essex County with the aim of identifying perpetrator profiles in cases of domestic violence. Using variables such as the perpetrator's age, relationship with the survivor, type of violence, and prior history, the analysis generated five distinct clusters. According to the authors, these profiles could help to better understand patterns of perpetration and to design more targeted interventions. The study highlights its preliminary nature and does not propose direct implementation of the results, but rather explores the usefulness of this type of technique for guiding future research (Hadjimatheou et al., 2024).

These experiences reflect initial ways in which AI could improve the availability and usefulness of data. However, their application poses significant challenges. Statistical automation does not replace the need to strengthen registration infrastructure or coordination among institutional actors. Without clear data governance frameworks, common standards, and technical capacities, there is a risk of producing sophisticated evidence

that is disconnected from public management. Further research is needed to assess under what conditions these tools can effectively support more informed and timely institutional decisions.

→ Institutional Learning

Beyond addressing individual cases or producing aggregate statistics, governments also need to develop the capacity to systematically learn from their own practice. Each administrative decision—whether a protection measure, a subsidy granted, a case referral, or a sentence—leaves traces that, when properly analyzed, can reveal patterns, recurring errors, unintended effects, or replicable good practices. This type of institutional learning is key to improving the quality of public responses and strengthening accountability (Gertler et al., 2016; Newcomer, Haltry, and Wholey, 2015).

In this context, AI could contribute to institutional learning through the analysis of large volumes of administrative or textual data. By processing dispersed or unstructured records, AI can help identify trends or relationships that might go unnoticed by operational teams. While documented experiences are still limited, research is beginning to emerge in sectors such as justice and health that illustrates this potential.

In Brazil, a study developed a machine learning model to identify gender bias in judicial rulings, focusing on cases of violence against women (Benatti et al., 2024). Using a database of more than 6,000 judgments classified by legal experts as “biased” or “not biased” according to criteria such as victim-blaming or justification of the perpetrator, the model achieved 81% accuracy in classifying rulings. Similar initiatives have begun to be explored in other countries, such as the herEthical AI project in England and Wales, which analyzes family court rulings to identify language that blames victims.

In Argentina, the AymurAI prototype seeks to partially automate the process of opening judicial rulings related to violence against women in criminal courts. Using natural language processing techniques, it extracts relevant information (type of violence, relationships between the parties, socio-economic context) and anonymizes the documents before uploading them to a public database. This tool does not replace human review but complements it, facilitating the structuring of data for safe publication.

In the United States, a study of more than 28,000 court hearings in Philadelphia trained a model to simulate counterfactual scenarios in decisions regarding release or detention of defendants in domestic violence



cases (Berk, Sorenson, and Barnes, 2016). The objective was to estimate how many cases of recidivism could have been avoided if the decisions had been based on the model's recommendations. The results suggested that, in that simulated scenario, recidivism could have been significantly reduced. This exercise was not implemented in judicial practice but sought to generate empirical evidence to reflect on the potential effects of different decision-making strategies.

In Spain, an interdisciplinary team applied supervised machine learning techniques to 491 judicial rulings on violence in intimate partner relationships, in order to explore whether the textual content of these judgments makes it possible to distinguish between lethal cases (femicides) and non-lethal cases. Using variables such as criminal records, frequency and severity of the violence, sanctions imposed, and contextual factors, the models correctly classified more than 75% of the cases. Although the study does not propose immediate operational application, it suggests that this information could be used to strengthen detection and prevention systems, as well as to inform risk assessment protocols adapted to the judicial context (García-Vergara et al., 2023).

These experiences show how AI can be used to analyze institutional decisions, reveal biases or risk patterns, and support the systematization of data. Although practical applications and rigorous evaluations are still incipient, they provide evidence to explore how AI could contribute to strengthening organizational learning in justice and protection systems.



HOW TO MOVE TOWARD RESPONSIBLE USE OF AI IN SENSITIVE CONTEXTS?

The use of artificial intelligence in guidance and response services for women experiencing violence can offer benefits but also poses significant risks. Organizations such as the Inter-American Development Bank and the Organisation for Economic Co-operation and Development have identified fundamental principles to guide its responsible implementation:

- **Data protection and informed consent:** The information processed by these services is often highly sensitive. International recommendations are to limit data collection to what is strictly necessary, apply anonymization techniques, and ensure secure storage. Consent must be clear, specific, and revocable ([Buenadicha et al., 2021](#); OCDE, 2019; [Latorre, Muro y Rego, 2024](#)).
- **Bias mitigation and explainability:** Algorithms can amplify historical inequalities if not properly audited. For this reason, experts promote equity audits and the development of explainable systems, especially in critical decisions such as case prioritization or the assignment of risk levels (OECD, 2024; EDPB, 2024; [Aránguiz Villagrán, 2022](#)).
- **Human oversight and professional training:** Automated decisions must be interpreted and adjusted by trained operators. Their participation from the design stages improves practical integration and strengthens trust in the system (Snow, 2019; Herm, 2023).

→ How to Move Forward? Evaluate, Document, and Learn

The experiences reviewed show that Latin America and the Caribbean are active participants in the global conversation on AI applied to the prevention and response to violence against women. Governments and organizations are exploring innovative uses of these technologies—some already implemented, others still in the pilot phase—with the goal of improving the accessibility, coordination, and effectiveness of public systems. This dynamism should be valued.

However, moving forward does not mean rushing processes. Structural challenges in the digitalization of services persist, including infrastructure gaps, deficits in data governance, low interoperability, and the absence of solid protocols for safeguarding sensitive information (Bustelo et al., 2025). Artificial intelligence cannot replace foundations that have not yet been built.

A significant evidence gap also persists. Most of the initiatives identified lack rigorous evaluations of their implementation, outcomes, and differential effects. Little is known about how these tools impact the quality of the response, the workload of teams, the continuity of processes, or users' trust. Nor are there systematic studies that examine their actual contribution to the protection of rights or their suitability for diverse institutional contexts.

In this scenario, it is a priority to invest in research that not only measures technical performance but also critically and contextually examines the ethical, institutional, and equity implications of AI. Rigorous evaluations, with an intersectional perspective, make it possible to define the conditions under which these technologies could be integrated usefully and responsibly into response systems. Only with solid evidence will it be possible to move toward an integration of AI that enhances institutional capacity for the prevention of and response to violence against women. Useful AI is not only the one that makes accurate predictions, but the one that contributes to more accessible, sensitive, and just public systems.

It is a priority to invest in research that not only measures technical performance but also critically and contextually examines the ethical, institutional, and equity implications of AI.

References

- [1] Arbach, K. (2023). Final advisory report on the Program to Strengthen the Uruguayan Prison System to Reduce the Risk of Recidivism. Internal document no. 2. Inter-American Development Bank.
- [2] Benatti, R., Severi, F., Avila, S., and Colombini, E. L. (2024). Gender bias detection in court decisions: A Brazilian case study. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 746–763).
- [3] Berk, R. A., Sorenson, S. B., and Barnes, G. (2016). Forecasting domestic violence: A machine learning approach to help inform arraignment decisions. *Journal of Empirical Legal Studies*, 13(1), 94–115.
- [4] Bernal-Monroy, E. R., Castañeda-Monroy, E. D., Rentería-Ramos, R. R., Campaña-Bastidas, S. E., Barrera, J., Palacios-Yampuezan, T. M., González Gustin, O. L., Tobar-Torres, C. F., and Ceballos-Villada, Z. R. (2025). Detection of victimization patterns and risk of gender violence through machine learning algorithms. *Informatics*, 12(1), 21. <https://doi.org/10.3390/informatics12010021>
- [5] Buenadicha, C., Galdon, G., Hermosilla, M. P., Loewe, D., and Pombo, C. (2019). The ethical management of data: Why it matters and how to make fair use of data in a digital world. Inter-American Development Bank. https://publications.iadb.org/es/publications/spanish/viewer/La_Gesti%C3%B3n_%C3%89tica_de_los_Datos.pdf
- [6] Chen, I. Y., Alsentzer, E., Park, H., Thomas, R., Gosangi, B., Gujrathi, R., and Khurana, B. (2020). Intimate partner violence and injury prediction from radiology reports. In *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium* (pp. 55–66).
- [7] Clough, B. A., and Casey, L. M. (2014). Using SMS reminders in psychology clinics: A cautionary tale. *Behavioural and Cognitive Psychotherapy*, 42(3), 257–268. <https://doi.org/10.1017/S1352465813001173>
- [8] Douglas, H. (2021). QLD police will use AI to “predict” domestic violence before it happens. Beware the unintended consequences. University of Melbourne. <https://findanexpert.unimelb.edu.au/news/30734-qld-police-will-use-ai-to-%27predict%27-domestic-violence-before-it-happens-beware-the-unintended-consequences>
- [9] Downing, N. R., Bogue, R. J., Terrill, P., and Tucker, S. (2021). Development and test of a text-messaging follow-up program after sexual assault. *Violence Against Women*, 27(11), 2111–2128. <https://doi.org/10.1177/10778012211014567>
- [10] Ellery, B. (2024). AI can help police predict if someone is at risk of domestic abuse. *The Times*. <https://www.thetimes.com/uk/crime/article/ai-can-help-police-predict-if-someone-is-at-risk-of-domestic-abuse-jbbrbgt9m>

- [11] Escobar-Linero, E., García-Jiménez, M., Trigo-Sánchez, M. E., Cala-Carrillo, M. J., Sevillano, J. L., and Domínguez-Morales, M. (2023). Using machine learning-based systems to help predict disengagement from legal proceedings by women victims of intimate partner violence in Spain. *PLoS One*, 18(6), e0276032. <https://doi.org/10.1371/journal.pone.0276032>
- [12] Éticas Foundation. (2024). Audit of VioGén: Assessing Spain's gender-based violence predictive system. https://eticasfoundation.org/wp-content/uploads/2024/12/Eticas_Audit_of_VioGen.pdf
- [13] European Data Protection Board (EDPB). (2024). AI auditing checklist for AI auditing. Support Pool of Experts (SPE) Programme. https://www.edpb.europa.eu/system/files/2024-06/ai-auditing_checklist-for-ai-auditing-scores_edpb-spe-programme_en.pdf
- [14] European Institute for Gender Equality (EIGE). (2016). Administrative data collection on violence against women: Good practices. <https://eige.europa.eu/sites/default/files/documents/mh0416095enn.pdf>
- [15] García-Vergara, E., Almeda, N., Fernández-Navarro, F., and Becerra-Alonso, D. (2023). Artificial intelligence extracts key insights from legal documents to predict intimate partner femicide. *Scientific Reports*, 13(1), 18212.
- [16] Garnelo, M., Bustin, C., Duryea, S., and Morrison, A. (2019). Applying behavioral sciences to intimate partner violence: Improving services for survivors in Latin America and the Caribbean. Inter-American Development Bank. <https://doi.org/10.18235/0001980>
- [17] Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. (2016). Impact evaluation in practice. World Bank Publications.
- [18] Gómez Mont, C., Del Pozo, C. M., Martínez Pinto, C., and Martín del Campo Alcocer, A. V. (2020). Artificial intelligence at the service of social good in Latin America and the Caribbean: Regional overview and snapshots from twelve countries. Inter-American Development Bank. <http://dx.doi.org/10.18235/0002393>
- [19] González Álvarez, J. L., López Ossorio, J. J., Urruela, C., and Rodríguez Díaz, M. (2018). Integral Monitoring System in cases of gender violence: VioGén System. *Behavior & Law Journal*, 4(1). <https://doi.org/10.47442/blj.v4.i1.56>
- [20] González, P., Peroni, A., Bonilla, A., Contreras, F., Guzmán, F., Heiss, J., Leyton, C., Soto, P., Lecourt, J., and Piras, C. (2023). Improvement of the 1455 Hotline Service of the Program for Care, Protection, and Reparation in Violence Against Women of SernamEG (Study conducted between January and July 2023). <https://minmujeryeg.gob.cl/wp-content/uploads/2025/01/Mejoramiento-del-servicio-de-atenci%C3%B3n-y-derivaci%C3%B3n-del-fono-1455-del-Programa-Atenci%C3%B3n-Protecci%C3%B3n-y-Reparaci%C3%B3n-en-Violencia-contras-las-Mujeres-del-SernamEG.-BID-2023.pdf>

- [21] Hadjimatheou, K., Quiroz Flores, A., Weir, R., and Skevington, T. (2024). Using unsupervised machine learning to find profiles of domestic abuse perpetrators. *Policing: A Journal of Policy and Practice*, 18, paae092. <https://doi.org/10.1093/police/paae092>
- [22] Herm, L. V. (2023). Impact of explainable AI on cognitive load: Insights from an empirical study. arXiv preprint, arXiv:2304.08861. <https://arxiv.org/pdf/2304.08861>
- [23] Hsieh, T. C., Wang, Y. H., Hsieh, Y. S., Ke, J. T., Liu, C. K., and Chen, S. C. (2018). Measuring the unmeasurable—A study of domestic violence risk prediction and management. *Journal of Technology in Human Services*, 36(1), 56–68. <https://doi.org/10.1080/15228835.2017.1417953>
- [24] Latorre, L., Muro, V., Rego, E., Gutiérrez, M., Cerrato, I., and Zárate, J. D. (2024). Tech report: Artificial intelligence. Inter-American Development Bank. <https://publications.iadb.org/en/tech-report-artificial-intelligence>
- [25] Latorre, L., De Leo, L., and Rego, E. (2025). Tech report: Digital ethics. Inter-American Development Bank. <https://publications.iadb.org/en/tech-report-digital-ethics>
- [26] López Belloso, M., and Izaguirre Choperena, A. (2024). New forms of support for situations of gender-based violence: The irruption of artificial intelligence in support services for women victims. In D. B. Benito Sánchez (ed.), *The protection of victims of gender-based violence: Legal and care aspects* (pp. 47–86). Bilbao: University of Deusto.
- [27] Microsoft News Center LATAM. (2022). The city incorporates artificial intelligence to strengthen the operation of Line 144. <https://news.microsoft.com/es-xl/la-ciudad-incorpora-inteligencia-artificial-para-for-talecer-el-funcionamiento-de-la-linea-144/>
- [28] Municipality of Lomas de Zamora. (2020). Integrated System for Supporting Victims of Gender-Based Violence Using Artificial Intelligence (SIAVIGia). <http://www.alertalomas.com.ar/niunamenos/>
- [29] Muñoz, V., and López-Ossorio, J. J. (2016). Psychological risk assessment of violence: Scope and limitations for forensic use. *Anuario de Psicología Jurídica*, 26(1), 130–140. <https://doi.org/10.1016/j.apj.2016.04.004>
- [30] Newcomer, K. E., Hatry, H. P., and Wholey, J. S. (eds.). (2015). *Handbook of practical program evaluation* (Vol. 864). Hoboken, NJ: John Wiley & Sons.
- [31] OECD (Organisation for Economic Co-operation and Development). (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- [32] ---- 2024. Assessing potential future artificial intelligence risks, benefits and policy imperatives. Paris: OECD. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/11/assessing-potential-future-artificial-intelligence-risks-benefits-and-policy-imperatives_8a491447/3f4e3dfb-en.pdf

- [33] WHO (World Health Organization). (2021). Gender-based violence prevalence estimates, 2018: Global and regional estimates of the prevalence of gender-based violence. Geneva: WHO.
- [34] Petersen, K., Davis, R. C., Weisburd, D., and Taylor, B. (2022). Effects of second responder programs on repeat incidents of family abuse: An updated systematic review and meta-analysis. *Campbell Systematic Reviews*, 18(1), e1217. <https://doi.org/10.1002/cl2.1217>
- [35] Roza, V., and Martín, C. (2021). Sexual and gender-based violence: Roadmap for prevention and care in Latin America and the Caribbean. <https://doi.org/10.18235/0003819>
- [36] Sanz, B., López-Belloso, M., and Izaguirre Choperena, A. (2025). Empathy, bias, and data responsibility: Evaluating AI chatbots for gender-based violence support. *Frontiers in Political Science*, 7, 1631881. <https://doi.org/10.3389/fpos.2025.1631881>.
- [37] Satariano, A., and Tool, R. (2024, July 18). An algorithm told police she was safe. Then her husband killed her. *The New York Times*. <https://www.nytimes.com/interactive/2024/07/18/technology/spain-domestic-violence-viogen-algorithm.html>
- [38] SISEMH (Secretariat for Substantive Equality Between Women and Men). (2024). Final report: OlivIA for Her. <https://igualdad.jalisco.gob.mx/wp-content/uploads/2024/12/reporte-final-Olivia.pdf>
- [39] Smee, B. (2021, September 14). Queensland police to trial AI tool designed to predict and prevent domestic violence incidents. *The Guardian*. <https://www.theguardian.com/australia-news/2021/sep/14/queensland-police-to-trial-ai-tool-designed-to-predict-and-prevent-domestic-violence-incidents>
- [40] Snow, T. (2019). Three key principles to help public sector organizations make the most of AI tools. Nesta. https://media.nesta.org.uk/documents/Decision-making_in_the_age_of_the_algorithm.pdf
- [41] Tabaie, A., Zeidan, A. J., Evans, D. P., Smith, R. N., and Kamaleswaran, R. (2022). A novel technique to identify intimate partner violence in a hospital setting. *The Western Journal of Emergency Medicine*, 23(5), 781–788. <https://doi.org/10.5811/westjem.2022.7.56726>
- [42] Trafford, L., Le, B. L., Firtala, S., and Williams, B. (2024). Lizzy – Building an AI-powered domestic abuse risk assessment tool based on nationally representative online survey data. <https://doi.org/10.31219/osf.io/shpjm>
- [43] UNODC (United Nations Office on Drugs and Crime). (2024, September). 911 calls: AI flags violence against women. https://www.unodc.org/unodc/en/frontpage/2024/September/911-calls_-ai-flags-violence-against-women.html
- [44] University of the Andes. (2024). SARA: AI to combat gender-based violence in Central America. Public Algorithms Systems. <https://sistemaspublicos.tech/sara-ia-para-combatir-la-violencia-de-genero-en-centroamerica/>

- [45] Westendorf, T. (2022). AI and policing: What a Queensland case study tells us. Australian Strategic Policy Institute. <https://www.aspistrategist.org.au/ai-and-policing-what-a-queensland-case-study-tells-us/>
- [46] Wieberneit, M., Thal, S., Clare, J., Notebaert, L., and Tubex, H. (2024). Silenced survivors: A systematic review of the barriers to reporting, investigating, prosecuting, and sentencing of adult female rape and sexual assault. *Trauma, Violence, & Abuse*, 25(5), 3742–3757.

This material was developed by:

Agustina Suaya, Monserrat Bustelo and Karen Martinez

Copyright © 2022 Inter-American Development Bank ("IDB"). This work is subject to a Creative Commons license CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>). The terms and conditions indicated in the URL link must be met and the respective recognition must be granted to the IDB.

Further to section 8 of the above license, any mediation relating to disputes arising under such license shall be conducted in accordance with the WIPO Mediation Rules. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the United Nations Commission on International Trade Law (UNCITRAL) rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this license.

Note that the URL link includes terms and conditions that are an integral part of this license.

The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent..

 www.iadb.org/gdlab

 gdlab@iadb.org



