GBV AoR HELPDESK

Gender-Based Violence in Emergencies

Gender-Based Violence and Artificial Intelligence (AI): Opportunities and Risks for Women and Girls in Humanitarian Settings

Jeanne Ward, with Sarah Spencer and Kavita Kalsi | August 2023

Introduction

Gender-based violence (GBV) affects women and girls everywhere in the world. In the 21st century, new technological developments—from the rise of mobile phones to the use of advanced technologies, like artificial intelligence (AI) and machine learning (ML)—have both facilitated and been used to prevent and respond to acts of GBV. As highlighted by the GBV AoR Helpdesk's Learning Series on Technology-Facilitated Gender-Based Violence and Guidance Note on Harnessing Technology to Prevent, Mitigate, and Respond to GBViE, the role of technology has dramatically altered the landscape for humanitarian actors working to address GBV in situations of conflict and crisis.

AI represents the next chapter in the future of technology and GBV. The recent release of ChatGPT—an AI chatbot developed by OpenAI and launched in November 2022—has drawn increased attention and public scrutiny to the ways in which AI is transforming daily life. As this technology and other forms of AI continue to develop, it is important for those working to address GBV in humanitarian settings to be aware of the role various forms of AI can play in relation to GBV.

This learning brief provides an initial introduction to AI and its links to GBV. It begins with an overview of key terms associated with AI relevant to GBV actors and summarizes current learning around how AI can exacerbate GBV in humanitarian settings. It then considers the ways in which AI may be used to address GBV, as well as the risks associated with the use of AI in GBV prevention and response. It concludes with several key takeaways for humanitarian actors working in and around GBV and AI and notes key areas for further learning and research.

What is artificial intelligence?

Artificial intelligence is broadly defined as the application of computer science methods such as machine learning and deep learning to large datasets to enable problem-solving through intelligent processes. It encompasses a range of powerful computer science tools and techniques to identify patterns, analyze text and images, and generate human-like speech (IBM, n.d.-a). It uses advanced **algorithms,** or sets of well-defined rules which can be executed by a machine, and draws on other capabilities, including **machine learning (ML), natural language processing (NLP)**, **knowledge representation**, and **automated**

reasoning, which enable AI to exhibit human-like behaviors. Some forms of AI like chatbots exist only in the virtual world while others are integrated into physical devices which comprise the **internet-of-things (IoT).** Figure 1 defines these and other key terms for GBV actors working with AI.

Figure 1. Key Terms for GBV Actors Working with Artificial Intelligence.	
Algorithm	In computer science: a set of well-defined rules to find the solution to a problem in a finite number of steps. (IBM, n.d.-b)
Artificial intelligence (AI)	The application of computer science methods such as machine learning and deep learning to large datasets to enable problem-solving through intelligent processes. (IBM, n.d.-a)
Automated reasoning	A specific discipline of artificial intelligence (AI) that applies logical deduction to computer systems.
Bots and Chatbots	A bot is an application that performs an automated task, such as when a computer rather than a human facilitates a food order. Bots are an example of narrow AI, as they possess some degree of human intelligence to carry out tasks. A chatbot is a computer program that uses AI and NLP to understand customer questions and automate responses to them, simulating human conversation. (IBM, n.d-f)
Data scraping	AI-powered Web scraping can be defined as using advanced techniques, such as artificial intelligence, machine learning algorithms, natural language processing (NLP), and computer vision, to automate data extraction from various websites.
Deepfakes and Disinformation	Deepfakes are false images or videos (synthetic media) created using AI to produce realistic (but entirely fake) representations of people and things. They are a type of disinformation that AI can produce. AI can not only develop false content, it can support the proliferation and dissemination of false content.
Internet-of-things (IoT)	Refers to the billions of physical devices around the world that are now connected to the internet, all collecting and sharing data (IBM, n.d.-e).
Knowledge representation	A fundamental concept in AI that involves creating models and structures to represent information and knowledge in a way that intelligent systems can use.
Machine learning (ML)	Advanced technology which uses data and algorithms to imitate the way that humans learn, gradually improving its accuracy (IBM, n.d.-c). Deep learning is an example of a machine learning technique that teaches computers to learn by example, such as with voice control on phones, tablets and other consumer devices.
Natural language processing (NLP)	The branch of artificial intelligence concerned with giving computers the ability to understand text and spoken word in much the same way human beings can. (IBM, n.d.-d)

How can artificial intelligence exacerbate GBV in humanitarian settings?

The rapid rise of AI presents new risks for women and girls living in humanitarian settings. As discussed further below, examples of these risks include AI-generated and AI-distributed disinformation, which can contribute to increased armed conflict and conflict-related sexual violence (CRSV), as well as direct tracking and targeting of female activists and vulnerable women and girls through misuse of AI-powered advanced biometrics and facial recognition. Given the rapid pace of technological advancement, many policy makers remain ill-equipped to handle these new risks, including governments, UN agencies, and NGOs working in the humanitarian space. According to Pauwels (2022),

> "The United Nations is not set up to address the spread of AI and data capture technologies in the hands of corporations and violent nonstate actors, nor is it able to deter its members [states] consistently and effectively from using such technologies for unlawful purposes" (p. 1).

Disinformation

A key risk for AI-related GBV in humanitarian settings stems from disinformation. Defined by the European Commission as "false or misleading content that is spread with an intention to deceive or secure economic or political gain," AI fuels disinformation in two key ways: AI-created false content, and AI-enabled proliferation, distribution and/or promotion of false content. The creation and spread of online disinformation can be used to incite violence and enflame racial, ethnic, and religious tensions. Disinformation has been used in multiple countries by both state and non-state actors to spark violence. For instance, in Myanmar, disinformation spread on Facebook and amplified by the social media site's AI-powered algorithms helped incite mob violence against the Rohingya minority, including widespread rape and sexual assault carried out as a part of a campaign of ethnic cleansing (Amnesty International, 2022; Human Rights Watch, 2017). AI-generated and AI-amplified disinformation increases the risk of GBV when it leads to mob violence and armed conflict. As with the crisis in Myanmar, women and girls may be vulnerable to CRSV or forced to flee, increasing the risks of many forms of GBV while displaced, including intimate partner violence (IPV), child marriage, sexual assault, and sexual exploitation.

Deepfakes

A related type of AI-facilitated GBV is deepfakes. Deepfakes are false images or videos (synthetic media) created using AI to produce realistic (but entirely fake) representations of people, often targeting celebrities, politicians, journalists, and others in the public domain. According to Pauwels (2022), researchers in Israel recently produced a new method for making deepfakes in real time which require no extensive facial data training, providing "a powerful tool kit to create realistic video forgeries at scale and with minimal know-how" (p. 13).

While the issue of deepfakes has received more attention in high-income countries, the phenomenon is increasingly prevalent across low- and middle-income countries (LMICs) and has been used to target female public figures such as Indian journalist Rana Ayyub, who was the victim of deepfake attacks on social media (Compton, 2021). Deepfakes can be used as a form of disinformation—such as through the dissemination of false violent propaganda by the Islamic State of Iraq and the Levant (ISIL) on social media—and can constitute GBV itself when used to create pornographic videos of women and girls which are shared without the individual's consent or used for blackmail (Pauwels, 2020).

Notably, disinformation generated by deepfakes is not limited to people; deepfakes can be used to lend credibility to and spread false information about many things. In humanitarian settings, this might include narratives about which areas are safe, where to safely cross borders, where to access support, etc.

Targeting of Female Activists and Vulnerable Women and Girls

One increasingly common way of tracking and targeting female activists through AI is with online bot campaigns, which can be used to harass and discredit women activists through social media. These bots can deluge accounts much faster and in much greater number than humans. In Iran, for example, accounts of female activists' and rights organizations' have been targeted through a flood of fake followers that impact activists' ability to spread their message to those they are trying to serve (Newman, 2022).

Another form of AI-facilitated GBV is the surveilling of female activists and vulnerable women and girls through facial and biometric recognition, (Pauwels, 2020). AI and biometrics can precisely verify the identity of individuals based on their physiological and behavioral traits. AI-powered biometrics offer relatively secure authentication protocols and have been used to better facilitate the distribution of humanitarian aid and to track highly-mobile populations in humanitarian settings (including in Afghanistan, Bangladesh, Jordan, Uganda, and Yemen). However, they can also increase the risks faced by women and girls if the data generated from AI-powered biometrics is not handled according to safety and ethical protocols (Peszkowska, 2022; Raftree & Steinacker, 2019).

Biometric data has been collected in humanitarian settings for a while. As AI is increasingly used to facilitate improved collection and analysis of biometric data, caution is essential. In one concerning example from a humanitarian setting, since 2018 UNHCR has shared biometric data collected as a part of aid distributions to Rohingya refugees in Bangladesh with the Government of Myanmar without the consent of affected populations (HRW, 2021). This breach of data privacy protections risks exposing returnees to further violence, including GBV, insofar as they can be specifically targeted by the government on their return.

Even beyond the risks associated with sharing data with governments, the existence and use of increasingly sophisticated and reliable AI-powered biometric data may expose women and girls to additional risks of GBV, as anyone with access to this technology (including aid workers) can immediately ascertain detailed information about an individual, such as their age, gender, location, and family structure.

How can artificial intelligence be used to address GBV in humanitarian settings?

While AI can be used to perpetrate existing as well as new forms of GBV, it can also be used by humanitarian actors to prevent, mitigate, and respond to GBV. It can facilitate research on GBV incidents, support case management, and be used to disseminate information and learning and conduct advocacy. Most of the learning around GBV and AI thus far has come from development settings (apart from an online GBV detection algorithm developed for use in Iraq); however, much of this innovation is applicable to humanitarian contexts. The following section provides an overview of five specific ways in which AI can be or has been used to address GBV: (1) sharing information about GBV services, (2) facilitating GBV risk assessment, (3) identifying GBV online, (4) prevention of (or rapid response to) GBV, and (5) GBV data collection.

Sharing information about GBV services (Chatbots)

Chatbots have been used in Central America, Thailand, and Mongolia to provide information and resources to GBV survivors. In Thailand, Sis Bot can be accessed 24/7 via Facebook Messenger and provides information to users about how to report incidents of GBV to the police, how to preserve forensic evidence, and what additional resources they are entitled to as a GBV survivor under the law (UN Women, 2019). Sara—the UNDP GBV chatbot for Central America—provides legal advice and helps GBV survivors make safety plans anonymously (UNDP, 2023). In Mongolia, GBV chatbots have begun integrating text message functions to reach people in rural areas and others who lack regular internet access (Ramamurthy et al., 2022). Importantly, these GBV chatbots were each developed by a team of GBV, legal, communication, and information technology specialists to ensure that their responses would be evidence-based and not put survivors in further jeopardy.

Facilitating GBV risk assessment

Another potential use of AI can be found in the facilitation of GBV risk assessments. Hunt et al. (2020) discusses the potential use of AI to predict the risk of domestic and intimate partner violence against women and children. These AI tools can be used either (a) to identify individuals who are at high risk of violence perpetration or victimization (cf., Petering et al., 2018), or (b) once an instance of GBV has been reported, to help case workers accurately assess ongoing risks faced by survivors and design effective safety plans. These AI tools frequently take as inputs a variety of personal data such as medical records from public health institutions and child welfare records as well as surveillance data from public locations (Rodríguez et al., 2021). In one example, an algorithm designed to identify instances of IPV successfully identified facial injuries caused by physical violence with an accuracy of 80 percent (Rodríguez et al., 2021).

Even as AI-based GBV risk assessment tools have the potential to ease the workload of case managers and other GBV care providers, they present a number of ethical concerns. High among these is the potential for violation of individual rights to privacy when a client has not consented to sharing personal data through AI technology. The only way these predictive tools work is with very large datasets. Predicting the likelihood of repeat violence will be based on previous events and incidents with hundreds of thousands of other clients. Serious ethical issues arise if this involves accessing records that go back years or even decades, when consent for sharing information to AI technology was not a consideration.

Another ethical concern is the risk of algorithmic bias in the technology itself, in which minority racial, ethnic, religious, or cultural groups are inaccurately profiled by the algorithm (Cockerill, 2020). Most existing AI algorithms are trained on so-called "WEIRD" datasets—originating from Western, Educated, Industrialized, Rich, and Democratic countries (Edit, n.d.), such that even if the data sets used are local, the interpretations of the data may result in bias. This risk of bias extends not only to survivors, but also to analysis of alleged or potential perpetrators.

Yet another risk with automating GBV risks assessments is that providers may over time increasingly defer to the automated outputs of risk assessments, without properly reviewing or interrogating the algorithm or the data outputs. This means that decisions are a result more of the machine than of the individual operating the machine. According to Spencer (2021),

> "Most technologists argue that AI should be designed and deployed to enable and support human decision-making, not make decisions themselves. But, human oversight is far from guaranteed. Humans are subject to a range of cognitive biases and prejudices that affect the way in which they make decisions. Without the right training and where models are poorly designed, humans could increasingly defer to the recommendations made by AI/ML systems." (p 23).

Identifying GBV online

AI can also be used to identify online GBV (e.g., non-consensual sharing of intimate images, online sexual harassment) as well as to flag individuals who are at high risk for perpetrating GBV offline based on social media activity. NLP algorithms have been used to successfully data-scrape websites to identify abusive language and expressions of violent intent on social media and could be used as the basis for targeted GBV prevention (Hunt et al., 2020).[1] Rodríguez et al. (2021) classifies online detection algorithms into those designed to identify (1) online misogyny, (2) sexism,[2] (3) child grooming, (4) reports of abuse, (5) child sexual abuse media, and (6) peer violence on online primary- and secondary-level school education platforms, reflecting the diversity of the types of GBV which may be addressed through online, AI-powered tools.

---

[1] Activist organizations are also spearheading efforts to address online GBV. One example is https://myimagemychoice.org, which monitors abuse and advocates for increased protections.

[2] The authors distinguish misogyny from sexism. They define misogyny as "hate speech that is targeted towards women" (p 5) and sexism as "any expression based on the idea that certain people are inferior based on their sex or gender" (p 5).

It is important to note, however, that among existing algorithms identified by Rodríguez et al.'s systematic review, most were more accurate for English-language users. For example, the best performing machine learning model for identifying online misogyny had a 91 percent accuracy rate in English compared to 85 percent for Italian and 81 percent for Spanish. While there have some efforts to expand these detection algorithms to LMICs, most research remains concentrated in developed countries. One notable exception is Abdulkareem and Karan's (2022) artificial neural network model which attempts to identify instances of online GBV in tweets from Iraqi users; however, this model was only designed to input Roman alphanumeric characters, excluding all Arabic-language tweets.

Prevention of (or rapid response to) GBV (Internet-of-Things)

One potentially promising avenue for AI-based GBV prevention which could be adapted for use in humanitarian settings is internet-of-things (IoT)-based devices. Once such example is Bindi, a wearable pendant and bracelet designed to detect physiological and auditory signs of fear from the user (e.g., heart rate, breathing patterns, etc.) and alert emergency contacts and local law enforcement (see Figure 2). Bindi has demonstrated 64 percent accuracy based on a small test data set collected by the developers (Miranda et al., 2022).

Figure 2. Design plan for Bindi, a wearable GBV prevention device, reproduced from Miranda et al. (2022).

GBV data collection and classification

Lastly, data scientists have been able to use machine learning to improve the collection of data on GBV, developing a model to identify cases of femicide from media reports from Latin America and the Caribbean. This effort is rooted in the theory of data feminism, a way of thinking about data science which asserts the importance of analyzing power hierarchies through an intersectional feminist lens, and is designed to counteract the absence of accurate government data on GBV (D'Ignazio et al., 2020). The resulting model was able to accurately classify media reports as femicide 81 percent of the time, representing a promising solution for reducing the labor required to identify and document femicides in low-resource settings.

Box 1: A Proposal for an AI Information Bot for GBV in Emergencies

Using AI in emergency settings is still in the very first stages. However, there are some exciting possibilities in the pipeline. In one example from the GBV Sub-Sector in Turkey, the GBV information manager is seeking funding to develop a system that will support GBV specialists in accessing and utilizing the information available in various GBV guidelines, manuals, and protocols. By leveraging AI-powered generative agents, the proposed system will support the information needs of GBV actors working in emergency settings, giving them convenient access to a centralized repository of GBV guidelines, manuals, and protocols, and enabling them to quickly retrieve relevant information to support their work. By streamlining information retrieval processes through a search engine dedicated to GBV tools and guidance, the generative agents will reduce time spent by GBV actors in accessing relevant information, allowing personnel to focus more on providing support and assistance to survivors. Another potential advantage to this system is that the information bot will promote consistent understanding and application of GBV guidelines, manuals, and protocols across different personnel, supporting a

Core Issues to Consider When Scaling Up Use of Artificial Intelligence to Address GBV

While AI can be used to address GBV, the rapid growth of AI technology and its ongoing limitations present a unique set of practical and ethical concerns for humanitarian actors. Issues discussed further below include: (1) reliance of private sector actors for AI technology, (2) concerns about data privacy and human rights, (3) AI "hallucinations", and (4) potential inequities in access among service-delivery organizations as well as GBV survivors.

(1) Reliance on the private sector

Currently, most AI technology is owned by a limited number of private sector companies which operate under limited government regulation. Thus, GBV actors which choose to integrate these tools into existing response and prevention programs become subject to private regulation, external market forces, potential supply chain disruptions, and company-specific data privacy policies. AI technology previously available at no or low cost could suddenly become prohibitively expensive or, even when the tools have been vetted by GBV actors for their ethical use, private (non-regulated) companies could change their privacy policies in the future in a way that conflicts with existing standards for the confidentiality of GBV survivors. According to one recent review,

> "Without adequate foresight, risk assessment, and normative leadership, governmental and international conflict prevention efforts may gradually rely on new, enhanced forms of behavioral surveillance driven by technologies fully or partially made by private sector actors in weakly regulated supply chains." (Pauwels, 2020, p. 1)

(2) Data privacy and human rights

A second serious concern with the use of AI already noted above is data privacy and the risks AI poses to the basic rights of self-determination, free expression, and non-discrimination. AI technologies typically rely on large datasets, which may include highly personal information about GBV survivors and other vulnerable populations. These databases are subject to misuse by the private sector companies which own them or hacking by external malicious actors, as in the November 2021 cyberattacks against Red Cross and Red Crescent Societies which compromised data on more than 515,000 individuals receiving services from the International Red Cross and Red Crescent Movement (ICRC, 2022).

Moreover, AI programs often function as "black boxes" even to their developers, meaning that it can be difficult or impossible to understand how a given AI program functions. This presents serious risks for discrimination and bias in AI-facilitated GBV response, prevention, and risk mitigation. Existing AI programs have been shown to have less accurate facial recognition for darker skin tones; discriminate against women in hiring decisions; and function poorly with non-native English speakers (Ohlheiser, 2023). As mentioned above, these risks are likely to be particularly acute in humanitarian settings, given that most training datasets come from populations living in the Global North.

AI and automation risk removing the 'human factor' that is critical to effective GBV response. Therefore, if AI is increasingly utilized in care for survivors, the aspect of human engagement that is critical to supporting the recovery of survivors may be compromised.

Where humans are engaged to identify, filter and label/moderate toxic AI content (e.g. sexual abuse or racist language and imagery) there is also a risk of harm and trauma to them. Time magazine reported on some of these forms of harm, including financial exploitation, experienced by human moderators in Kenya (Perrigo B, 2023). There may also be other risks such as economic exploitation. These risks must be acknowledged, and steps taken by technology companies and other actors to adequately mitigate them.

(3) AI Algorithmic Failures

AI-based chatbots have been shown to "hallucinate"—a term used to describe AI-generated responses which are factually incorrect and based on no real-world inputs. While this problem affects all currently available generative AI (such as OpenAI's ChatGPT, Google's Bard, and Microsoft's Bing), it is especially dangerous for chatbots (Weise & Metz, 2023). For example, it is possible that a generative AI could "hallucinate" the name and location of a GBV service provider, directing a survivor to a non-existent resource.

(4) The ability of NGOs and CSOs, as well as survivors, to access and leverage the potential benefits of AI

Finally, the use of AI presents important equity concerns for local service providers as well as for women and girls seeking GBV services in humanitarian settings. Most if not all current AI requires users to access computer technology and devices, which local organizations may not have. According to a Humanitarian Practice Network Paper on Humanitarian AI, few NGOs and CSOs are unable to fully leverage the potential benefits of AI,

> "…unless they have the networks, reputation, or operational reach to successfully broker pro-bono relationships with corporate entities. And the organisational structure of some agencies and the low number of AI/ML experts or data scientists employed by aid organisations further inhibits the design and uptake of AI/ML." (Spencer, 2021. p. 10).

Clients may also be affected, if a mobile phone and/or stable internet connection is required. In their systematic review of ethical challenges and opportunities for addressing IPV with AI-technologies, Novitzkey et al. (2023) caution that not all GBV survivors have access to personal mobile phones, which have been the predominate focus of AI-based GBV services. This is likely to be especially true in humanitarian settings where forced displacement, lack of access to electricity, and poor internet connectivity may impede access to online services, or where women only have access to technology (phones, tablets, computers) owned by their male partners and relatives.

Key Legal and Ethical Frameworks and Approaches When Using AI to Address GBV

In order to counter some of the safety and ethical concerns identified above, GBV actors in humanitarian contexts should always consider the following frameworks and approaches to promote safe and ethical use of AI in addressing GBV.[3]

1. **Data protection and privacy laws:** AI systems dealing with GBV may process sensitive personal data. Compliance with data protection and accountability principles is paramount in data processing. (See Box 1.)

Box 2: Data protection and accountability principles

The European Union's General Data Protection Regulation (GDPR), which is reportedly among the toughest privacy and security laws in the world, outlines seven protection and accountability principles for processing data.

1. Lawfulness, fairness, and transparency — Processing must be lawful, fair, and transparent to the data subject.
2. Purpose limitation — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
3. Data minimization — You should collect and process only as much data as necessary for the purposes specified.
4. Accuracy — You must keep personal data accurate and up to date.
5. Storage limitation — You may only store personally identifying data for as long as necessary for the specified purpose.
6. Integrity and confidentiality — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g., by using encryption).
7. Accountability — The data controller is responsible for being able to demonstrate GDPR compliance with all these principles.

2. **Human rights frameworks:** The Universal Declaration of Human Rights (UDHR) and other international agreements play a significant role in guiding the use of AI in addressing GBV. Respect for human dignity, privacy, non-discrimination, and access to justice are key principles to consider.

3. **User-centered consent and autonomy:** Involving survivors in decision-making regarding the use of AI technology is crucial. Ensuring informed consent, autonomy, and empowerment of survivors and those at risk of GBV should be prioritized when using AI tools for GBV interventions.

4. **Bias and fairness:** Efforts should be made to ensure that AI systems are fair and unbiased. Algorithms used should be regularly audited and tested for biases related to gender, race, or any other protected characteristic. **Every effort should be made to** prevent marginalizing vulnerable groups even more (e.g., if data is taken from the Global North, are those models reflected in the Global South and used to make decisions there?)

5. **Accountability and liability:** There should be clear accountability and liability mechanisms in place. Developers, organizations, and users of AI systems addressing GBV should take responsibility for any

[3] These frameworks and approaches were largely generated using CHATGPT (Oct 2023) in response to a user question and reviewed for alignment with other materials and guidance accessed in the desk review for this paper.

potential harm or misuse arising from the technology.

6. **Transparency:** AI systems used in GBV interventions should be transparent and provide explanations for their decisions. This helps promote trust and enables individuals to understand how decisions about their safety and well-being are being made.

7. **Multi-stakeholder engagement and multidisciplinary approaches:** Development and use of AI to address GBV should be a collaboration with survivors, relevant communities, civil society organizations, and experts from various domains (law, psychology, etc.) to ensure that AI solutions are developed, deployed, and monitored in an inclusive, ethical and comprehensive manner.

8. **Ongoing assessment and regulation:** Given the rapidly evolving nature of AI technology, there is a need for continuous assessment, monitoring, and regulation of its use in addressing GBV. Governments, organizations, and relevant bodies should regularly update and adapt legal and ethical frameworks as new challenges and risks emerge.

Conclusion

Like any technology, AI has the power both to help stop GBV and to perpetuate it. In communities already affected by conflict and crisis, new AI technologies from social media algorithms to online chatbots can increase the risks faced by women and girls. AI can create and amplify highly sophisticated disinformation that incites violence and worsens underlying racial, ethnic, religious, or political tensions, leading to forced displacement and increased risks of sexual violence. It can also be used to create synthetic media, such as deepfake videos, which can be used to harass or blackmail vulnerable populations both off and online. New advances in biometric recognition and online surveillance can help both state and non-state actors identify and target female activists, journalists, and vulnerable women and girls.

Box 3: Useful online resources with basic information on AI

- https://www.edx.org/course/humanitarian-action-in-the-digital-age/
- https://www.udacity.com/course/intro-to-artificial-intelligence--cs271
- https://www.coursera.org/learn/introduction-to-ai
- https://www.socialworkers.org/About/Ethics/AI-and-Social-Work
- https://www.radicalai.org/feminist-ai

On the other hand, AI can also be used in a variety of ways to help humanitarian actors address GBV. AI-powered chatbots can help GBV survivors access services anonymously, and AI risk assessment tools can help GBV specialists identify at-risk women and girls as well as potential perpetrators of GBV before it happens. Online detection algorithms can help weed out online misogyny and sexual harassment while wearable devices like Bindi can facilitate real-time rapid response and GBV prevention. Finally, AI can be used to scrape data from online sources to provide more accurate information on and reporting about GBV in a given community.

While these new ways of tackling GBV are promising tools for GBV actors in humanitarian settings, they also present risks which must be addressed. AI-powered tools are still largely the purview of the private sector, meaning that GBV actors who choose to use or develop these tools must consider both the benefits and risks of private sector involvement in the humanitarian space. Moreover, AI magnifies existing challenges with data privacy and confidentiality that exist in the GBV sector, and some AI-powered tools like chatbots can provide incorrect information about survivors, as well as about alleged or potential perpetrators, in the absence of extensive testing and safeguards. INGOs and NGOs have a responsibility to protect the data sets they are holding and to recognize the risk of these falling into the hands of hacker and in AI gaining access to this data and/or manipulating it. Finally, as with all new technological developments, GBV actors must ensure that any AI-based resources are equally available to local women's organizations, and that survivors have the tools (such as cellphones) to benefit from the resource.

There are conversations globally about whether and to what extent AI can and should be regulated. AI is part of a new wave of technological innovation which promises to reshape how billions of people live their daily lives, including those living and working in humanitarian crises. While more research is needed to understand the true impacts of this nascent field on GBV in emergencies, GBV actors must be proactive in learning about AI, considering its role in shaping the lives of women and girls affected by humanitarian emergencies, and joining in discussions and debates about the benefits and challenges of AI globally. Now is the time for GBV advocates, coordinators, program specialists and donors to forge alliances with AI developers to contribute expertise and insights to the safe development and utilization of AI to address GBV in humanitarian contexts.

References

Abdulkareem, L. R., & Karan, O. (2022). Using ANN to Predict Gender-Based Violence in Iraq: How AI and data mining technologies revolutionized social networks to make a safer world. In 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 298-302). IEEE. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9932831.

Amnesty International. (2022). The social atrocity: Meta and the right to remedy for the Rohingya. https://www.amnesty.org/en/documents/ASA16/5933/2022/en/.

Cockerill, R. G. (2020). Ethics Implications of the Use of Artificial Intelligence in Violence Risk Assessment. The journal of the American Academy of Psychiatry and the Law, 48(3), 345-349. https://jaapl.org/content/48/3/345.long.

Compton, S. (2021). More and more women are facing the scary reality of deepfakes. Vogue. https://www.vogue.com/article/scary-reality-of-deepfakes-online-abuse.

D'Ignazio, C., Val, H. S., Fumega, S., Suresh, H., & Cruxên, I. (2020). Feminicide & machine learning: detecting gender-based violence to strengthen civil sector activism. http://hdl.handle.net/10625/60535.

Edit. (n.d.). AI has a WEIRD problem. https://edit.co.uk/blog/ai-has-a-weird-problem/.

European Commission. (n.d.). Tackling online disinformation. https://digital-strategy.ec.europa.eu/en/policies/online-disinformation.

Human Rights Watch. (2017). "All of my body was pain": Sexual violence against Rohingya women and girls in Burma. https://www.hrw.org/report/2017/11/16/all-my-body-was-pain/sexual-violence-against-rohingya-women-and-girls-burma.

Human Rights Watch. (2021). UN shared Rohingya data without informed consent. https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent.

Hunt, X., Tomlinson, M., Sikander, S., Skeen, S., Marlow, M., du Toit, S., & Eisner, M. (2020). Artificial intelligence, big data, and mHealth: The frontiers of the prevention of violence against children. Frontiers in artificial intelligence, 3, 543305. https://doi.org/10.3389/frai.2020.543305.

IBM. (n.d.-a). What is artificial intelligence? https://www.ibm.com/topics/artificial-intelligence.

IBM. (n.d.-b). Algorithm. https://www.ibm.com/docs/en/iad/7.2.1?topic=algorithm

IBM. (n.d.-c). What is machine learning? https://www.ibm.com/topics/machine-learning#:~:text=the%20next%20step-,What%20is%20machine%20learning%3F,learn%2C%20gradually%20improving%20its%20accuracy.

IBM. (n.d.-d). What is natural language processing? https://www.ibm.com/topics/natural-language-processing.

IBM. (n.d.-e). IoT solutions. https://www.ibm.com/cloud/internet-of-things#:~:text=The%20Internet%20of%20Things%20(IoT,all%20collecting%20and%20sharing%20data.

IBM. (n.d.-f). What is a chatbot? https://www.ibm.com/topics/chatbots.

ICRC. (2022). Cyber attack on ICRC: What we know. https://www.icrc.org/en/document/cyber-attack-icrc-what-we-know.

Miranda, J.A., Rituerto-González, E., Luis-Mingueza, C., Canabal, M.F., Bárcenas, A.R., Lanza-Gutiérrez, J.M., Peláez-Moreno, C., & López-Ongil, C. (2022). Bindi: Affective Internet of Things to Combat Gender-Based Violence. IEEE Internet of Things Journal, 9, 21174-21193. https://ieeexplore.ieee.org/document/9780201.

Novitzky, P., Janssen, J., & Kokkeler, B. (2023). A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools. Heliyon. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10277589/.

Ohlheiser, A. (2023). AI automated discrimination. Here's how to spot it. Vox. https://www.vox.com/technology/23738987/racism-ai-automated-bias-discrimination-algorithm.

Pauwels, E. (2020). Artificial Intelligence and Data Capture Technologies in Violence and Conflict Prevention. Global Centre on Cooperative Security. Accessed on, 10(07), 2022. https://www.jstor.org/stable/resrep27551.

Perrigo, B. (2023). OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. TIME. https://time.com/6247678/openai-chatgpt-kenya-workers/

Petering, R., Um, M. Y., Fard, N. A., Tavabi, N., Kumari, R., & Gilani, S. N. (2018). Artificial intelligence to predict intimate partner violence perpetration. Artificial intelligence and social work, 195. https://www.researchgate.net/publication/329281099_Artificial_intelligence_to_predict_intimate_partner

_violence_perpetration.
Peszkowska, A. (2022). How the use of biometrics in the humanitarian sector has the potential to put people at risk – an interview with Belkis Wille. Responsible Data. https://responsibledata.io/2022/05/24/how-the-use-of-biometrics-in-the-humanitarian-sector-has-the-potential-to-put-people-at-risk-an-interview-with-belkis-wille/.
Raftree, L., & Steinacker, K. (2019). Head to head: Biometrics and aid. The New Humanitarian. https://www.thenewhumanitarian.org/opinion/2019/07/17/head-head-biometrics-and-aid.
Ramamurthy, A., Serafica, P., & Joffre, V.M. (2022). Chatbots offer a new lifeline to address domestic violence. Asian Development Blog. https://blogs.adb.org/blog/chatbots-offer-new-lifeline-address-domestic-violence.
Rodríguez, D. A., Díaz-Ramírez, A., Miranda-Vega, J. E., Trujillo, L., & Mejia-Alvarez, P. (2021). A systematic review of computer science solutions for addressing violence against women and children. IEEE Access, 9, 114622-114639. https://www.researchgate.net/publication/353792843_A_Systematic_Review_of_Computer_Science_Solutions_for_Addressing_Violence_Against_Women_and_Children.
Spencer, S. (2021) Humanitarian AI: The hype, the hope and the future. Humanitarian Practice Network, Network Paper Number 85. https://odihpn.org/publication/humanitarian-artificial-intelligence-the-hype-the-hope-and-the-future/
UNICEF. (n.d.) Safer Chat Bots: Six Steps to Make your Chatbot Safer for Children and Young People, https://www.unicef.org/media/114671/file/6-Steps-to-make-your-chatbot-safer-children-young-people-2022.pdf
UNICEF. (n.d.) Safeer Chat Bots Implementation Guide, https://www.unicef.org/media/114681/file/Safer-Chatbots-Implementation-Guide-2022.pdf
UN Women. (2019). Using AI in accessing justice for survivors of violence. https://www.unwomen.org/en/news/stories/2019/5/feature-using-ai-in-accessing-justice-for-survivors-of-violence.
UNDP. (2023). Sara: The new artificial intelligence tool to tackle gender violence in Central America. https://www.undp.org/latin-america/press-releases/sara-new-artificial-intelligence-tool-tackle-gender-violence-central-america.
Weise, K., & Metz, C. (2023). When A.I. chatbots hallucinate. The New York Times. https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html.

The GBV AoR Help Desk

The GBV AoR Helpdesk is a unique research and technical advice service which aims to inspire and support humanitarian actors to help prevent, mitigate and respond to violence against women and girls in emergencies. Managed by Social Development Direct, the GBV AoR Helpdesk is staffed by a global roster of senior Gender and GBV Experts who are on standby to help guide frontline humanitarian actors on GBV prevention, risk mitigation and response measures in line with international standards, guidelines and best practice. Views or opinions expressed in GBV AoR Helpdesk Products do not necessarily reflect those of all members of the GBV AoR, nor of all the experts of SDDirect's Helpdesk roster.

The GBV AoR Helpdesk

You can contact the GBV AoR Helpdesk by emailing us at:
enquiries@gbviehelpdesk.org.uk

The Helpdesk is available 09.00 to 17.30 GMT Monday to Friday.

Our services are free and confidential.