ACTION BRIEF

FROM WARFARE TO PEACEBUILDING: employing artificial intelligence for the women, peace and security agenda

EMPLOYING ARTIFICIAL INTELLIGENCE FOR THE WOMEN, PEACE AND SECURITY AGENDA

With increasing dependency on technology and a steadily rising number of Internet users across the Asia and the Pacific region, the stage for peacebuilding and security actors is changing. Artificial intelligence (AI) is becoming an integral element of information and communication technologies (ICT), from everyday digital tools to military technologies. While AI offers innovative solutions to countless issues, it also carries pitfalls that, if left unaddressed, may have serious implications on gender equality and peace and security.

AI AND GENDERED SECURITY RISKS

AI refers to the use of computer systems to mimic human thinking, behaviour and decision-making processes. This may include everyday technologies such as voice-powered virtual assistants, predictive text functions and search and recommendation algorithms on our personal devices. While the rise of AI has the potential to revolutionize numerous functions across societies, concerns are emerging as AI systems have been shown to perpetuate discriminatory practices with serious implications for gender equality, human rights as well as human and national security.

However, AI is an integral part of widely used online platforms, which are becoming increasingly important spaces for digital activism and civic engagement; these emerging technologies have the power to support and leverage such efforts. Against this backdrop, this action brief outlines the risks and opportunities that AI carries for the implementation of the Women, Peace and Security agenda.

Discriminatory gender biases across AI systems

Numerous studies have shown that many widely used AI-powered systems exhibit significant gender and racial biases. In these systems, women tend to be disadvantaged. For example, a 2018 study on facial recognition software found that darker-skinned women were the most misclassified group, with an error margin of up to 35 per cent. In contrast, lighter-skinned men were the least misclassified group, with an error margin of less than 1 per cent.[1] A 2019 study of nearly 200 facial recognition systems came to similar conclusions, indicating a higher incidence of false positives among women of colour.[2] Likewise, biases have been found in speech recognition software, where an analysed programme was 70 per cent more likely to correctly caption a man's voice compared to a woman's.[3]

Studies suggest that gender biases in AI result from existing and historical inequalities where women, people of colour and minorities have been underrepresented, stereotyped and discriminated against. With 78 per cent of AI developers being men, the algorithms that make up AI-powered systems are largely shaped by men's experiences and perceptions of the world and often overlook the experiences and needs of women users.[4] Stereotypes of women as submissive or sexualized are prevalent throughout these systems. Virtual assistants, for example, are overwhelmingly portrayed as docile women.[5]

Furthermore, many recruitment processes that have piloted AI-powered systems have been found to disadvantage women applicants. In one case, the algorithms used were based on recruitment data that had been collected during the 10 years preceding its development. Because women had been underrepresented in the related profession during this time, the system ended up classifying women's applications as less relevant to the advertised position, thus filtering them out.[6]

AI is also increasingly being used in criminal justice settings. If employed ethically, AI has the potential to upscale the provision of legal services and reduce costs, thereby making justice accessible to more people. However, it also carries risks. Algorithmic risk assessments use data on criminal history, drug use, mental health factors, education and employment to assess individuals' risk of recidivism. Such a system may disadvantage women if the systems do not include carefully analysed sex-disaggregated data and trends. Programmes piloted to make criminal justice risk assessments have tended to overpredict the risk of women reoffending, as they have not accounted for existing data that indicates that women are less likely to reoffend than men.[7]

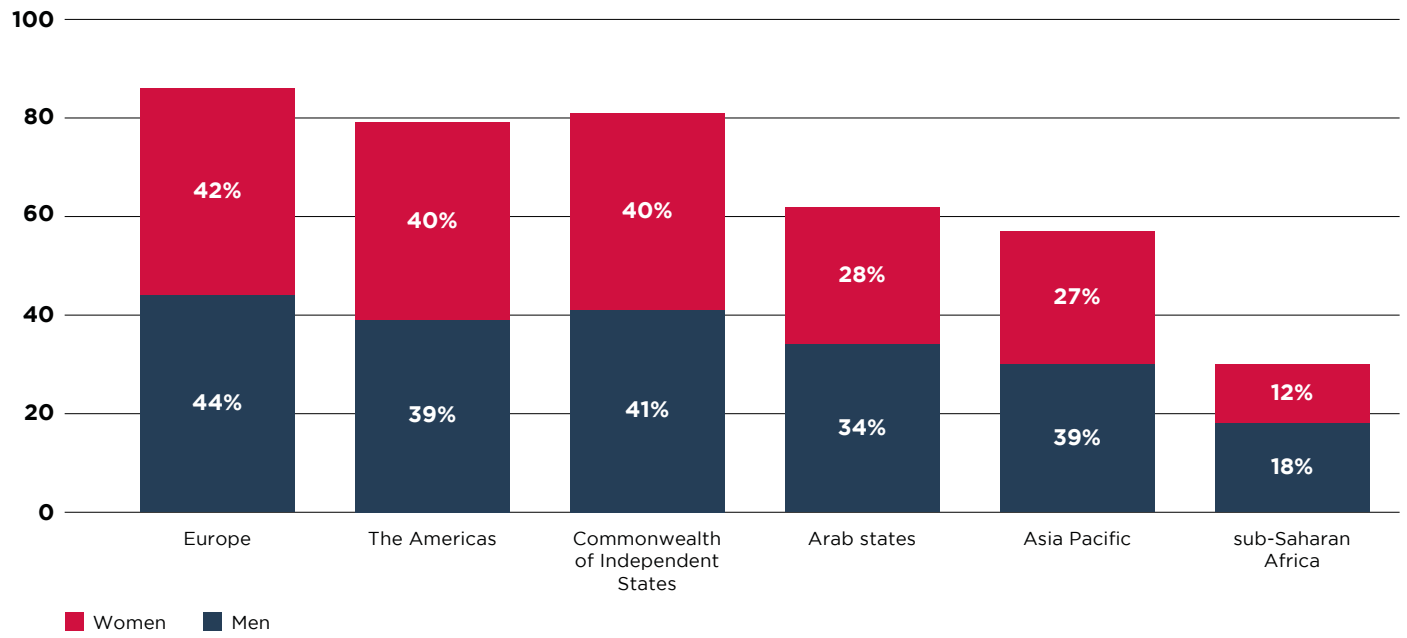
Hence, the lack of sex-disaggregated data and gender-sensitive algorithm programming contributes to bias and inaccuracy issues. The Judicial Integrity Network in ASEAN has highlighted the need for judicial participation in decisions concerning such technologies in order to ensure that they do not "perpetuate inequality, erode public confidence in the judiciary, and obscure critical information that judges require to make independent, transparent decisions."[8]

The digital gender gap further deepens bias issues

The overall demographics of the Internet play a significant role in AI creation and gender bias risks. This is particularly the case for machine learning technologies that develop their functional logic based on probability calculations of data and imagery readily available across the Internet. Approximately 40 per cent of the global population does not have Internet access; the two regions with the lowest access rates are the Asia and the Pacific and sub-Saharan Africa regions. Further, women tend to have less access to the Internet than men do. In the Asia and the Pacific region, the digital gender gap is 5 points, with 54 per cent of women and 59 per cent of men using the Internet.[9] These numbers foreshadow a reality of unequal Internet engagement, where men and persons in the Global North are more likely to inform and impact online content. **With fewer women and persons from the Global South leaving data points online, AI-powered systems are at risk of being skewed because they do not reflect the perspectives of underrepresented groups.**[10]

Without employing an informed and critical approach to AI development that addresses pre-existing biases, these technologies risk recreating and enforcing discriminatory practices across societies. As such, they dictate who will benefit from AI technologies and who will risk facing harm.

INTERNET USAGE RATE PER CAPITA PER REGION (2020), DISAGGREGATED BY SEX



Data: UN International Telecommunication Union Digital Development Estimate (2020) and UN Population Data (2020)

AI-powered military technologies carry risks

ICT and AI-powered systems are increasingly being used to facilitate military interventions, including in surveillance and intelligence operations, autonomous vehicles and lethal weapons. Biases across these systems are a serious concern to inclusive and gender-responsive peace processes across the Asia and the Pacific region and globally. A 2021 survey of 250 defence technology leaders of the North Atlantic Treaty Organization (NATO) concluded that all were considering adopting AI technologies to optimize their military interventions; 49 per cent were already implementing and optimizing such solutions.[11] As these technologies become further integrated into modern warfare, concerns are being voiced regarding the implications for armed conflict, national security and the implementation of the Women, Peace and Security agenda.

Because of its ability to identify and mimic patterns in large amounts of data, AI systems can be efficient when given organizational tasks. However, AI systems have yet to adequately capture the complexity of human decision-making or to account for contextual differences and sudden changes in data patterns.[12] This can have particularly severe ethical concerns for its use in armed conflict. For example, research on drone warfare has argued that civilian men of military-age may run a higher risk of being killed by automatised or semi-automatised drone strikes, as the systems are more likely to classify men as combatants based on behavioural indicators relating to their gender.[13] This illustrates how bias and algorithmic oversimplification may have a real risk to human lives. The United Nations Secretary-General and the Office of the United Nations High Commissioner for Human Rights have called for a ban on autonomous weapons and for moratoriums on the use and sale of military AI until adequate safeguards are in place.[14] **It is imperative to develop measures to assess how military AI technologies respond to gender bias and to integrate the Women, Peace and Security agenda into military AI strategies.**[15]

DEFINITIONS

a Digital echo chambers are online environments where users are only exposed to information or opinions that coincide with their own.

b Incel, short for 'involuntary celibate', is an online subculture for persons who consider themselves to be unable to find a romantic or sexual partner despite desiring one.

c Deepfakes are digitally manipulated video material that mimic real events.

AI-facilitated hate speech, misogyny and disinformation

AI and ICT can also be used to disrupt social cohesion and to incite conflict and violence. Disinformation campaigns are a growing issue on social media platforms and fuel conflict by inciting hate speech, spreading radicalism and encouraging violence. This was arguably the case in Myanmar, where, in 2017, social media hate speech was considered to have fuelled abuses of the Rohingya population.[16]

Algorithms that are designed to present users with tailored content based on personal interests are also creating *digital echo chambers*,[a] which are exacerbating these issues. These echo chambers have the potential to normalize, confirm and radicalize users' perceptions and world-views because they filter out content that may contradict or challenge users' opinions. This also applies to misogynistic attitudes, which highly correlate with support for violence against women and radical extremist ideals.[17] Echo chambers have been identified as one of the key drivers in the emergence of '*incel*' culture in the 2010s.[b] Incel culture, which has provoked violent hate crimes, including mass murder and sexual violence, is commonly characterized by misogyny (including the endorsement of violence against women), homophobia, racism and a sense of entitlement to sex.[18]

Although the extent to which AI is employed in disinformation campaigns remains debated, a 2020 study found that 81 countries have been involved in using social media to spread disinformation. 70 per cent of these countries used automated accounts ('bots'), which autonomously generate and publish content online.[19] Violent extremist groups are also known to widely use social media to recruit new members and to spread radical content and disinformation. Recruitment campaigns are tailored to fit their target audience, and gender stereotypes are widely used to manipulate gender norms.[20] For example, studies have shown that the Islamic State strategically uses notions of femininity and masculinity to create compelling messaging to recruit and govern its supporters.[21] **As AI technologies are becoming more user-friendly and accessible to a wider audience, there is a growing concern that these technologies will be used by extremist groups to facilitate and enhance their operations.**[22]

The potential to produce and spread *deepfakes* to dis-inform and harass has been outlined as a security risk.[c] The technology has been used to edit targets' faces over others' in pornographic videos, seamlessly making it appear as if the intended victim is the person depicted in the video. This tactic is particularly used against women, including against human rights defenders and journalists, as a means to discredit and silence them.[23] Deepfakes have also targeted political figures to distort or otherwise manipulate original

statements or messaging to either reduce or increase public support for political agendas.[24] Deepfakes have the potential to severely lower public trust in social media and online news, as it will be difficult to confirm or debunk their authenticity, thereby undermining security and political stability.[25]

Generic AI-powered filter mechanisms miss the mark

AI technology is being developed to counter the issue of disinformation and hate speech by automatically detecting, taking down or preventing such content from being posted online. However, a lot of work remains to make these protection mechanisms effective and appropriate for the diverse contexts in which they are being used. For example, women Internet users across Southeast Asia have highlighted that language barriers are an issue when reporting harmful and misleading information on social media platforms.

Automated content filters are not yet advanced enough to account for contextual differences, resulting in situations where content that may be harmful in a local context may not be detected by a generic filter mechanism. Activists have also criticized social media filter mechanisms for censoring non-explicit images of women's bodies (e.g. women breastfeeding) while failing to remove content depicting rape, hate speech and violence against women.[26] Further, recent cases have shown that algorithms used on social media platforms, which are designed to maximize user engagement, have promoted negative and destructive content ranging from ethnic hate speech to eating disorders. This approach has negatively affected the mental health and well-being of girls.[27]

OPPORTUNITIES FOR WOMEN PEACEBUILDERS

The COVID-19 pandemic has increased the extent to which women peacebuilders use ICT and digital solutions to support their work. Although gender biases in these technologies hinder equal and safe online engagement, digital peacebuilding and online civic engagement are venues for increased opportunities for women peacebuilders to advance their work. Digital solutions will play important roles in several key peacebuilding areas. [28]

Enhancing conflict analysis

ICTs are increasingly used to analyse conflict trends and stakeholder behaviours, thus informing mediation and advocacy efforts. This consists of manual and automated analyses of social media narratives, geographic information systems (e.g. satellite imagery), and outputs from big-data text mining. These technologies may provide new insights or bring support to existing analyses on conflict dynamics.[29] For instance, the analysis of satellite data shed light on the recent abuses of the Rohingya population in Myanmar, where imagery of cleared land across Rakhine state brought attention to conflict developments in 2017.[30]

ICTs can also be used to draw insights into the gender dynamics of conflict and insurgency. Analysis of social media content across Bangladesh, Indonesia, Malaysia and the Philippines showed that both men and women consumed gendered radical content. Radical narratives targeting women were shown to glorify motherhood and emphasize that true love could only be found through the liberation from secular 'restrictions' on religion. Men, on the other hand, were depicted as warriors and protectors of religion. **Insights into how gender norms are used to target women and men in different ways are useful in informing what can be done to efficiently counter recruitment and radicalization attempts.**[31]

Facilitating early warning and response

Peacekeeping missions and humanitarian actors are increasingly using advanced, computer-generated models to predict and forecast crises and conflicts; this has enabled faster and more efficient responses to arising needs. Simple indicators can also be used to alert peacebuilders and the public about emergent security risks. For example, a rise in hate speech towards a specific group, such as ethnic minorities or members of certain political factions, may indicate impending violence and conflict escalation. The increased use of misogynistic language is also an indicator of a rise in violent extremism, as misogyny is integral to the ideology of many extremist groups. Accounts of women's human rights and women human rights defenders being attacked are similarly an early warning sign for imminent violence.[32]

Online early warning systems are currently being piloted. These systems use AI to analyse the real-time spread of disinformation, hate speech and manipulated images. The outcomes are then used to inform journalists, election monitors and other human observers of developing trends, which facilitates rapid responses and countermeasures.[33]

Informing peaceful counter-narratives

Promoting peaceful narratives and evidence-based statements are effective means of countering hate speech and disinformation. A first step to developing peaceful counter-narratives is to understand the nature of hateful messaging and disinformation, how these are being spread and by whom. Increasing digital literacy is an important step in this process, as is strengthening capacities to use online verification tools that can help identify bots, certify content and fact-check sources.

Supporting women's civil society in putting pressure on social media and ICT companies to adopt context- and gender-sensitive policies to prevent the spread of disinformation and hate speech is also a crucial step forward. This includes developing and reforming methods of user engagement and algorithms that may give rise to harmful digital echo chambers and offer breeding grounds for violent and hateful narratives and action.[34]

Mobilizing engagement for peace and activism

Women peacebuilders widely use ICT tools, particularly social media, to connect, mobilize and disseminate information. For example, women across the Philippines have used online learning tools to support demobilization efforts by offering former female and male combatants' educational opportunities. Social media has also played an important role in the peace process in the Bangsamoro Autonomous Region of Muslim Mindanao in the Philippines; young women used online campaigns to encourage youth to vote and support the Bangsamoro Organic Law, which was pivotal in consolidating the peace agreement with the Government of the Philippines and the Moro Islamic Liberation Front.[35]

While AI at its early stages is giving rise to concerns, it is also opening up innovative solutions for online engagement that can be harnessed by the public and peacebuilders alike. Globally, initiatives such as chatbots are being developed to provide on-demand legal support to protestors and activists. Chatbots have also been used to facilitate political and peacebuilding dialogues with a broader group of stakeholders, including with the broader public who is affected by the conflict without being directly involved as combatants, who would normally struggle to gain access to and participate in such processes.[36]

RECOMMENDATIONS

Although mediators are increasingly using ICT and AI to inform their work, women peacebuilders are less likely to do so, particularly in the case of AI.[37] Because women's unequal engagement in digital peacebuilding is largely due to lower Internet access, digital literacy and compromised online security, steps to increase inclusivity in these processes include:

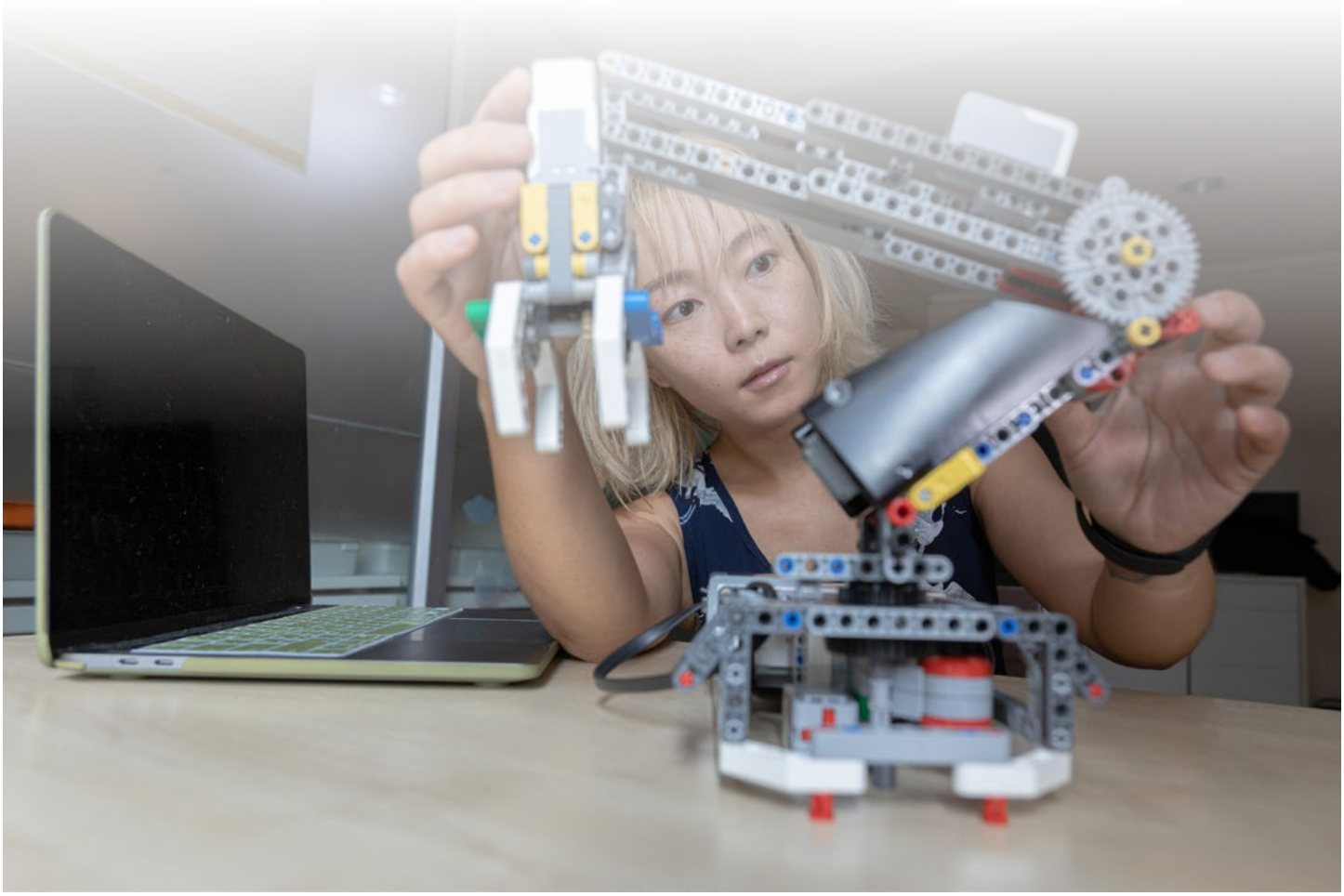
1. **Gathering more evidence, data and research on the implications of AI to support the implementation of the Women, Peace and Security agenda** across the Asia-Pacific region. This can inform evidence-based initiatives that promote gender-responsive peace online and offline.

2. **Supporting women's leadership and participation in the development and auditing of AI and ICT** for peace and social cohesion. This should also encompass related policies, laws and other decision-making processes.

3. **Advocating for the integration of a rights-based lifecycle approach in the design of AI technologies** that recognizes the diversity of contexts in which they function, has a clear, predefined intent and scope of functionality and whose design is gender-responsive and transparent to the user.

4. **Supporting women peacebuilders' enhancement of their conflict analysis, mediation efforts and advocacy** through digital means and online platforms by designing initiatives and programmes informed by the local needs of women peacebuilders and activists.

5. **Increasing digital literacy on ICT and AI with a specific focus on conflict contexts,** including the security implications of gender and racial biases and the threat of disinformation. This will empower women and girls to adopt secure online practices and to advocate for inclusive, safe and gender-responsive platforms.



Women learning robotics in Bangkok, Thailand, 02 April 2022.
©Photo: UN Women/Ana Norman Bermudez

ENDNOTES

1 Boulamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, Proceedings of Machine Learning Research, (81) pp. 1-15.

2 Grother, P., Ngan, M. and Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) – Part 3: Demographic Effects, National Institute of Standards and Technology, https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

3 R. Tatman, "*Gender and Dialect Bias in YouTube's Automatic Captions", Proceedings of the First Workshop on Ethics in Natural Language Processing, 2017*, pp. 53–59.

4 World Economic Forum (2018), "*Assessing Gender Gaps in Artificial Intelligence," The Global Gender Gap Report 2018*.

5 UNESCO, EQUALS Skills Coalition (2019). I'd Blush if I Could: Closing Gender Divides in Digital Skills through Education. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

6 Dastin. J. (2018). Amazon Scraps Secret AI Recruiting Tools that Showed Bias Against Women, *Reuters*, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

7 Hamilton, M. (2019). The Sexist Algorithm, Behavioral Sciences & the Law, 37(2), pp. 145-157, https://researchoutreach.org/articles/justice-served-discrimination-in-algorithmic-risk-assessment/

8 Judicial Integrity Network ASEAN (2021). Emerging Technologies and Judicial Integrity in ASEAN: Judicial perspectives on the use, opportunities, challenges and impact of technologies on the rule of law, access to justice and court administration. https://www.undp.org/publications/emerging-technologies-and-judicial-integrity-asean#modal-publication-download

9 ITU (2021), Measuring Digital Development: Facts and Figures 2020, https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

10 Chandler, K. (2021). Does Military AI Have Gender? Understanding Bias and Promoting Ethical Approaches in Military Applications of AI, *UNDIR Research Report.* https://unidir.org/publication/does-military-ai-have-gender-understanding-bias-and-promoting-ethical-approaches

11 D. Chenok, L. van Bochoven and D. Zaharchuk (2021). "Deploying AI in Defense Organizations", *IBM Center for the Business of Government,* https://www.ibm.com/downloads/cas/EJBREOMX

12 de Cremer, D. and Kasparov, G. (2021). AI Should Augment Human Intelligence, Not Replace it. Harvard Business Review, https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it

13 Shoker, S. (2021). Military-Age Males in Counterinsurgency and Drone Warfare, Palgrava Macmillan.

14 United Nations, "Amid Widening Fault Lines, Broken Relations among Great Powers, World 'in Turmoil', Secretary-General Tells Paris Peace Forum, Calling for New Social Contract", Press Release SG/SM/19852, 11 November 2019, https://www.un.org/press/en/2019/sgsm19852.doc.htm; "Urgent Action Needed over Artificial Intelligence Risks to Human Rights", 15 September 2021. https://news.un.org/en/story/2021/09/1099972.

15 Chandler, K. (2021). Does Military AI Have Gender? Understanding Bias and Promoting Ethical Approaches in Military Applications of AI, *UNDIR Research Report.* https://unidir.org/publication/does-military-ai-have-gender-understanding-bias-and-promoting-ethical-approaches

16 Choudhury, A. (2020). How Facebook is Complicit in Myanmar's Attacks on Minorities, *The Diplomat*, https://thediplomat.com/2020/08/how-facebook-is-complicit-in-myanmars-attacks-on-minorities/

17 Johnston, M. and True, J. (2019). Misogyny & Violent Extremism: Implications for Preventing Violent Extremism. https://asiapacific.unwomen.org/en/digital-library/publications/2019/10/misogyny-violent-extremism

18 Salojärvi, E., Rantanen, M., Neiminen, E. and Hanhela, H. (2020). The "Incel" Phenomenen in the Digital Era – How Echo Chambers have Fueled the Incel Movement in Amadae S.M. (eds.), Computional Transformation of the Public Sphere, pp 195-210, https://philarchive.org/archive/SALTIP

19 Bradshaw, S., Bailey, H. and Howard, P. N. (2021). Industrialized Disinformation: 2020 Global Inventory of Organised Social Media Manipulation. Working Paper 2021:1. Oxford, UK: Project on Computational Propaganda.

20 UN Women Regional Office for Asia and the Pacific (2019). Who's behind the keyboard? A gender analysis of terrorism and violent extremism in the online space in Bangladesh, Indonesia, Malaysia and the Philippines, https://asiapacific.unwomen.org/en/digital-library/publications/2019/03/whos-behind-the-keyboard

21 UN Women (2018). Empowerment or Subjugation: An analysis of ISIL's gendered messaging. https://arabstates.unwomen.org/en/digital-library/publications/2018/6/empowerment-or-subjugation

22 UNCCT and UNICRI (2021). Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes, https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf

23 Venema, A. E., (2020). Deepfakes as a Security Issue: Why Gender Matters, *Cybersecurity, Defense and Security Techologym WIIS Blog,* https://wiisglobal.org/deepfakes-as-a-security-issue-why-gender-matters/

24 Cook, E. (2019). Deepfakes Could Have Real Consequences for Southeast Asia, *The Interpreter,* https://www.lowyinstitute.org/the-interpreter/deep-fakes-could-have-real-consequences-southeast-asia

25 Vaccari, C. and Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty and Trust in News. *Social Media + Society.* January 2020.

26 Nurik, C. (2019). Self-Regulation, Hate Speech and Gender-Based Censorship on Facebook, *International Journal of Communication,* 13(2019), pp. 2878-2898.

27 https://www.wsj.com/articles/the-facebook-files-11631713039

28 Bunse, S. (2021). Social Media: A Tool for Peace or Conflict?, *Stockholm International Peace Research Institute Blog,* https://www.sipri.org/commentary/topical-backgrounder/2021/social-media-tool-peace-or-conflict

29 UNDPPA and UNCHD (2019), *Digital Technologies and Mediation in Armed Conflict,* https://peacemaker.un.org/sites/peacemaker.un.org/files/DigitalToolkitReport.pdf

30 BBC News (2017). Rohingya Crisis: Satellite Images of Myanmar Village Burning, https://www.bbc.com/news/world-asia-41270891

31 UN Women Regional Office for Asia and the Pacific (2019). Who's behind the keyboard? A gender analysis of terrorism and violent extremism in the online space in Bangladesh, Indonesia, Malaysia and the Philippines, https://asiapacific.unwomen.org/en/digital-library/publications/2019/03/whos-behind-the-keyboard

32 Johnston, M., True, J., Gordon, E., Chilmeran, Y and Riveros-Morales, Y. (2020). Building a Stronger Evidence Base: The Impact of Gender Identities, Norms and Relations on Violent Extremism (a case study of Bangladesh, Indonesia and the Philippines), *UN Women research report*, https://asiapacific.unwomen.org/en/digital-library/publications/2020/05/building-a-stronger-evidence-base-the-impact-of-gender-identities-norms

33 Yankoski, M., Weninger, T. and Scheirer, W. (2020). An AI Early Warning System to Monitor Online Disinformation, Stop Violence, and Protect Elections, *Bulletin of the Atomic Scientists,* 76(2), p. 85-90. https://www.tandfonline.com/doi/full/10.1080/00963402.2020.1728976

34 GNWP and ICT4Peace (2021). Women, Peace, and Security and Human Rights in the Digital Age: Opportunities and risks to advance women's meaningful participation and protect their rights, https://gnwp.org/wp-content/uploads/PolicyBriefGNWP-2021c.pdf

35 Ibid.

36 Office of the Special Envoy of the Secretary-General for Yemen (2020). Cutting-Edge Tech in the Service of Inclusive Peace in Yemen. https://osesgy.unmissions.org/cutting-edge-tech-service-inclusive-peace-yemen

37 GNWP and ICT4Peace (2021). Women, Peace, and Security and Human Rights in the Digital Age: Opportunities and risks to advance women's meaningful participation and protect their rights, https://gnwp.org/wp-content/uploads/PolicyBriefGNWP-2021c.pdf

Regional Office for Asia and the Pacific
UN Building, Rajdamnern Nok Avenue
Bangkok 10200, Thailand
Tel: +66 2 288 2093
info.bangkok@unwomen.org