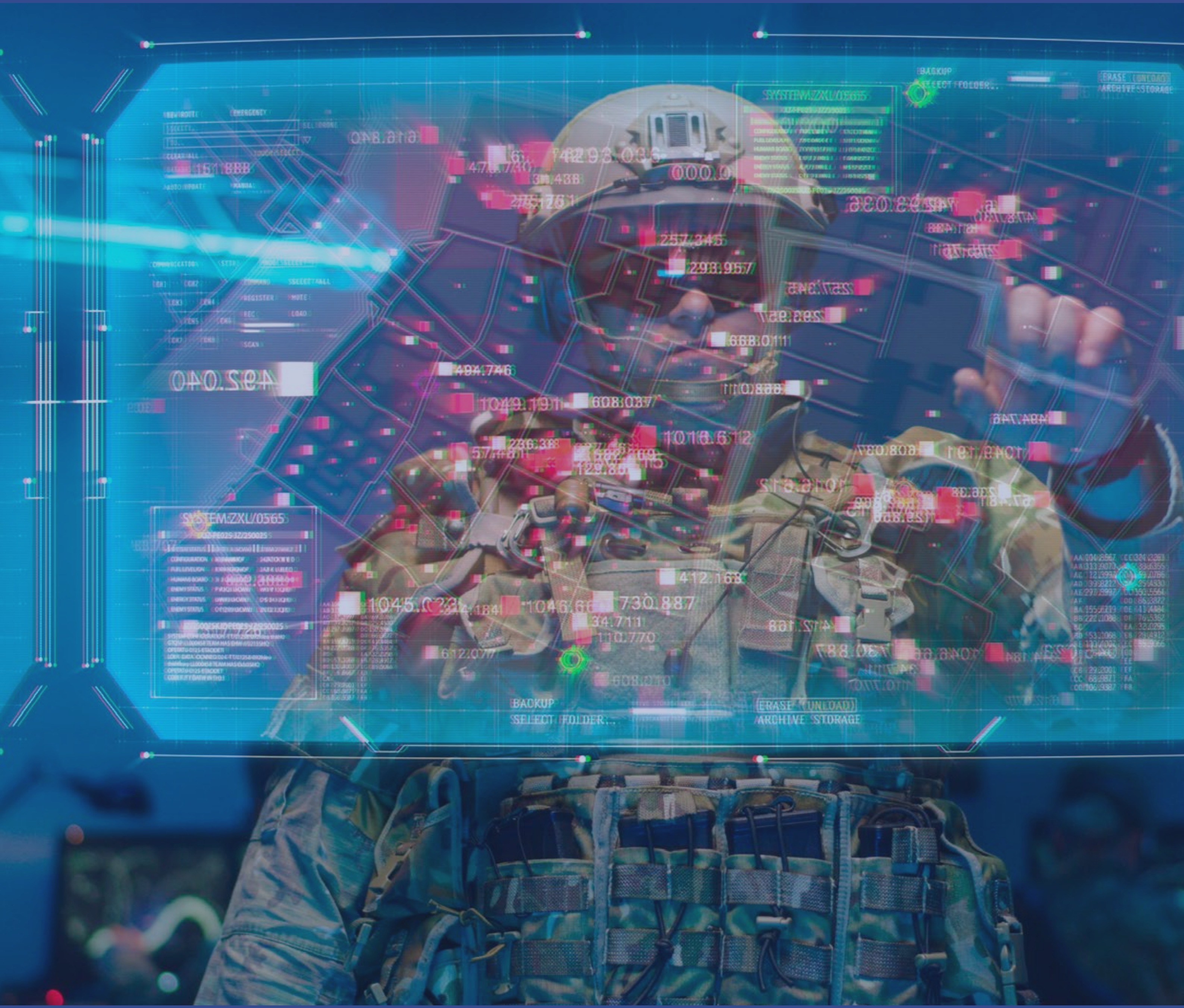
UNIDIR UNITED NATIONS INSTITUTE FOR DISARMAMENT RESEARCH



Does Military AI Have Gender?

Understanding bias and promoting ethical approaches in military applications of AI

KATHERINE CHANDLER

ACKNOWLEDGEMENTS

Support from UNIDIR core funders provides the foundation for all of the Institute's activities. The Gender and Disarmament Programme is supported by the governments of Canada, Ireland, Norway, Sweden and the United Kingdom. In addition, the author wishes to thank Renata H. Dalaqua, Arthur Holland, Shimona Mohan, Giacomo Persi Paoli and Lucy Suchman for their thoughtful comments and suggestions, as well as numerous subject matter experts who contributed to the research.

NOTES

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. The views expressed in the publication are the sole responsibility of the individual authors. They do not necessarily reflect the views or opinions of the United Nations, UNIDIR, its staff members or sponsors.

CITATION

Chandler, Katherine, *Does Military AI Have Gender? Understanding bias and promoting ethical approaches in military applications of AI*, UNIDIR, Geneva, 2021. https://doi.org/10.37559/GEN/2021/04.

ABOUT UNIDIR

UNIDIR is a voluntarily funded, autonomous institute within the United Nations. One of the few policy institutes worldwide focusing on disarmament, UNIDIR generates knowledge and promotes dialogue and action on disarmament and security. Based in Geneva, UNIDIR assists the international community to develop the practical, innovative ideas needed to find solutions to critical security problems.

ABOUT THE GENDER AND DISARMAMENT PROGRAMME

The Gender and Disarmament Programme seeks to contribute to the strategic goals of achieving gender equality in disarmament forums and effectively applying gender perspectives disarmament processes. It encompasses original research, outreach activities and resource tools to support disarmament stakeholders in translating gender awareness into practical action.

ABOUT THE AUTHOR

Dr. Katherine Chandler is an Assistant Professor in the Walsh School of Foreign Service, Georgetown University, affiliated with the Culture and Politics Program and Science, Technology and International Affairs. She studies the intersection of technology, media, gender and politics. Her first book, *Unmanning: How Humans, Machines, and Media Perform Drone Warfare*, was published in 2019. Her current research studies how a gender, peace and security agenda can be used to address digital war.

ABBREVIATIONS

AI	Artificial intelligence
CCW	Certain Conventional Weapons (Convention)
GGE	Group of Governmental Experts
LAWS	Lethal autonomous weapon systems

Photos: www.shutterstock.com: Front cover: Frame Stock Footage, p6: Konstantin Faraktinov, p18: Ihor 2020, p24: ktsdesign, www.pexels.com: p5 Pixabay

Table of contents

Executive summary

Although evidence of bias in civilian applications of artificial intelligence (AI) is easy to find, less research exists on how military applications of AI may draw on and reproduce inequalities. To overcome this gap, this report examines how gender norms can be implicitly and explicitly encoded in machine learning processes and assesses the potential consequences for military applications of AI. It builds on evaluations of civilian applications of AI that underline gender bias and applies these insights to military AI.

Military applications of AI are still in the initial stages of research and development. This is both an opportunity and a challenge for policymakers. Although regulators have the chance to have a substantial impact on how the technology develops, it is difficult to know how untried systems will be used in the future. Nonetheless, it is clear that military applications of AI that fail to account for gender difference or that do so in a way that is culturally specific has the potential to limit human rights and to turn back advances made by United Nations Security resolution 1325 of 2000 on Women, Peace and Security.

Below are some of the main findings and recommendations.

- Gender norms and bias can be introduced into machine learning throughout its life cycle, which includes data collection, the training of algorithmic models, their evaluation, their use, and their archiving or disposal. Within these systems, harms can be amplified through connections between gender and other identity markers, including race, age and ability.
- The challenges associated with gender and military applications of AI are not reducible to a single cause: gender norms are not constant, nor are they consistent across cultures; rather, they reflect the specific ways in which gender is interwoven with society, politics and economics and normalized as inherent characteristics.
- The limitations of modelling particular people as universal can be seen in applications of AI as diverse as voice recognition, image detection and machine translation, all of which currently recognize men at higher rates than women. Such machine learning applications contribute to proposed uses for military applications of AI, which extend beyond autonomous weapons and vehicles to include human resource management; intelligence, surveillance and reconnaissance (ISR); cyberspace operations; command and control; and military logistics.
- To overcome these problems, a gender-based approach to human–machine interactions is needed. This framework would consider how the development of military applications of AI reflect the roles, behaviours, activities and attributes that a given society at a given time considers to be appropriate or the “norm” for women, men, girls and boys, as well as non-binary or gender-fluid people.
- If AI is to replicate human intelligence, a narrow understanding of what is human must instead be replaced by a more complex model that includes the range of bodies, abilities and emotions that are all part of human experience. Thus, a wider range of experts, including scholars of gender, race and ability, should be included in debates on military applications of AI, as a way to ensure that policies regarding the military applications of AI are informed by diverse perspectives.

- Rather than rely on a neutral category of "human", developers of military applications of AI should make transparent how their systems respond to and reflect the diversity of humanity. A gender-based review of military applications of AI should make explicit how the system represents and responds to gender and how harmful effects have been mitigated.
- Gender-based considerations must be supported by regulatory policies, which include holding relevant parties responsible for gender inequities or violence that result from military applications of AI. The report affirms that military applications of AI that have the potential to continue or exacerbate gender-based harms should be subject to moratoriums or banned.

1. Introduction

In recent multilateral discussions, governmental experts have reiterated the significance of understanding artificial intelligence (AI) as being the result of human and machine synthesis.[1] This is an important step in recognizing how technological advance is not separate from human actions and how the development of AI is tied to social, economic, legal and political decisions. However, gender – an integral category to humans – remains underdeveloped in debates on military applications of AI. Although research in artificial intelligence seems to primarily entail replicating human processes through computation, AI systems also do the reverse: they normalize a particular "human" through their models. Machine learning, the dominant model of AI in use today, relies on massive data sets to train algorithms to recognize patterns. This data, and the algorithmic models that are created from it, implicitly and explicitly reproduce gender norms, often under the guise of neutral, smachine models.

This report reviews gender and machine learning research conducted over the past 10 years. It builds on evaluations of AI that underline gender bias and applies these insights to military applications of AI. Artificial intelligence refers to computer systems that aim to replicate human processes, though the term is constantly evolving.[2] The report follows the current convention of referring to machine learning as AI, although machine learning processes are typically understood by computer scientists as a specific application of AI.[3] While future military applications of AI systems may not rely on machine learning, they will continue to recreate "human" intelligence and, in so doing, reference gender norms. This close analysis of machine learning illustrates how gender becomes part of ostensibly non-gendered systems.

Military AI is defined in this report as applications of artificial intelligence developed for national security.[4] This definition points out how the same AI systems can be used for both military and non-military purposes. This research addresses the diversity of systems encompassed by military AI by stepping back to ask three questions: Who builds the technology? For whom is it intended? And with whose interests in mind?[5] It shows how gender norms shape understandings of military applications of AI and uncovers how gender bias – which refers to the ways in which one gender is privileged over others – occurs within machine learning.[6]

[1] CCW Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Report of the 2019 Session, CCW/GGE.1/2019/3, 25 September 2019, https://undocs.org/CCW/GGE.1/2019/3, p. 3.

[2] R.W. Button, "Artificial Intelligence and the Military", RAND Blog, 7 September 2017, https://www.rand.org/blog/2017/09/artificial-intelligence-and-the-military.html.

[3] S. Brown, "Machine Learning Explained", Ideas Made to Matter, MIT Sloan School of Management, 21 April 2021, https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

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

[5] C. D'Ignazio and L.F. Klein, Data Feminism, 2020, https://doi.org/10.7551/mitpress/11805.001.0001.

[6] This definition differs from the meaning of bias in other fields, notably statistics and law, and includes bias drawn from existing gender norms. In the field of statistics, bias is a specific problem that indicates systemic differences between a sample and the population. This contrasts with the meaning of bias in the field of law, for example, where the term refers to a judgment based on prejudices rather than on fact. An algorithm with no statistical bias – that is, when the sample accurately models the population – could still contain gender bias, given that gender inequalities persist within the broader population. See K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021, https://www.degruyter.com/document/doi/10.12987/9780300252392/html, p. 135.

Gender norms and bias can be introduced into machine learning throughout its life cycle. The machine learning life cycle includes data collection, the training of algorithmic models, their evaluation, user interactions, and the archiving or disposal of the systems.[7] Harms can be amplified in the machine learning life cycle through connections between gender and other identity markers, including race, age and ability.[8] Some of these limitations can be overcome through gender-based approaches to machine learning research and development, while policies can be adopted to limit or ban AI that perpetuates harm or fails to be equitable in its uses.

Errors are introduced into machine learning by assuming that a specific subset of human beings stands for all. In her 2019 book *Invisible Women*, gender researcher Caroline Criado Perez marshals hundreds of studies to show the "widespread societal bias that frames men as the default, neutral humans" and "how this default male bias has led to data gaps that, at their most serious, can prove fatal for women".[9] The limitations of modelling particular people as universal can be seen succinctly in applications of AI as diverse as voice recognition,[10] image detection[11] and machine translation,[12] all of which currently successfully recognize men at higher rates than women. Such machine learning models contribute to proposed military applications of AI, which encompass autonomous weapons and vehicles; human resource management; intelligence, surveillance and reconnaissance (ISR); cyberspace operations; command and control; and military logistics.[13]

The challenges associated with gender and military applications of AI are not reducible to a single cause: gender norms are not constant, nor are they consistent across cultures; rather, they reflect the specific ways in which gender is interwoven with society, politics and economics and normalized as inherent characteristics.[14] Gender, moreover, does not only refer to women; indeed, the existing norms of war mean that algorithmic models might be less likely to identify civilian men as non-combatants, raising concerns about gender bias against men.[15]

Chapter 2 of this report explains how gender is the backdrop to the human–machine interactions that create military applications of AI. It illustrates how gender is implicit in the three factors that make machine learning possible: data collection, algorithms and computer

[7] H. Suresh and J. Guttag, "Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle", MIT Case Studies in Social and Ethical Responsibilities of Computing, summer 2021, https://doi.org/10.21428/2c646de5.c16a07bb.

[8] S.M. West, M. Whittaker and K. Crawford, Discriminating Systems: Gender, Race and Power in AI, AI Now Institute, April 2019, https://ainowinstitute.org/discriminatingsystems.pdf.

[9] C. Criado Perez, Invisible Women: Exposing Data Bias in a World Designed for Men, 2019, p. 319.

[10] J. Palmiter Bajorek, "Voice Recognition Still Has Significant Race and Gender Biases", Harvard Business Review, 10 May 2019, https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases.

[11] K. Yang et al., "Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy", in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 547–558, https://doi.org/10.1145/3351095.3375709.

[12] Gendered Innovations, "Machine Translation: Analyzing Gender", Stanford University, n.d., https://genderedinnovations.stanford.edu/case-studies/nlp.html#tabs-2.

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

[14] See, for example, S. Harding, The Science Question in Feminism, 1986; D. Haraway, Simians, Cyborgs, and Women: The Reinvention of Nature, 1991; C. Cohn, "Sex and Death in the Rational World of Defense Intellectuals", Signs, vol. 12, no. 4, 1987, pp. 687–718, https://doi.org/10.1086/494362; D. MacKenzie and J. Wajcman (eds.), The Social Shaping of Technology, 2nd ed., 1999); and C. Enloe, Bananas, Beaches and Bases: Making Feminist Sense of International Politics, 2014.

[15] N. Linos, "Rethinking Gender-Based Violence during War: Is Violence against Civilian Men a Problem Worth Addressing?", Social Science & Medicine, vol. 68, no. 8, 2009, pp. 1548–1551, https://doi.org/10.1016/j.socscimed.2009.02.001.

processing. Chapter 3 provides an overview of three case studies of AI in practice, exemplifying some of the challenges associated with the development and deployment of these systems in specific contexts. Chapter 4 then considers how best practices from research in ethical AI can be applied to military applications of AI. Finally, chapter 5 outlines a gender-based approach to military applications of AI: an approach that would account for the roles, behaviours, activities and attributes that a given society at a given time considers to be appropriate or the "norm" for women, men, girls and boys, as well as non-binary or gender-fluid people.[16] Ultimately, military AI that fails to address gender or does so in a way that is culturally specific risks reproducing and exacerbating existing inequalities.[17]

[16] M. Mikkola, "Feminist Perspectives on Sex and Gender", Stanford Encyclopedia of Philosophy, 25 October 2017, https://plato.stanford.edu/entries/feminism-gender/#SexDis. Non-binary categories of gender introduce significant concerns for AI that are largely unaddressed in this report. See K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021, https://www.degruyter.com/document/doi/10.12987/9780300252392/html, pp. 131–133.

[17] Leverhulme Centre for the Future of Intelligence, "Global AI Narratives", https://www.ainarratives.com.

2. Gender norms and military applications of AI

2.1. Human–machine interaction

Artificial intelligence describes the interdisciplinary field that aims to replicate human cognition through machines. As such, AI cannot be uncoupled from human actions. Rather, AI, and disputes regarding the concept, are embedded in political, social, legal and economic relations.[18] Governmental experts evaluate and regulate how machine-led systems can retain meaningful human control.[19] Computer programmers decide how to fine-tune a facial-recognition algorithm.[20] Commercial machine learning products integrate human workers to overcome the limitations of artificial neural networks. Labourers in warehouses fulfil AI-directed logistics and human moderators review explicit content flagged by AI.[21] In short, people direct how AI is funded, engineered, used, regulated and disposed of, while a vast network of users and labourers interact with and carry out the instructions associated with the programs. The limitations associated with AI are not only technological but also emerge through inter-sections with political, social and economic conditions.

A brief history of artificial intelligence in the United States of America and the current shift to machine learning algorithms illustrates how AI models of human processes are culturally specific and tied to strategic and economic goals.[22] US military budgets from the 1960s to the 1990s funded research in AI with the long-term aim of providing "decision support" to the armed forces through computation. Early AI formalized human cognitive processes through symbolic logic, programming these rules into computer models.[23]

Despite the resources allocated to these projects, there were few substantial breakthroughs during this period, in part due to the symbolic approach adopted by researchers. Standard rules and definitions meant that nuance easily captured by people – for example, the multiple meanings of a single word – resulted in machine error. At the end of the Cold War, funding was scaled back substantially and many researchers in the field questioned whether "expert" computer systems would ever be able to replicate human intelligence.[24] That debate continues today.

[18] See, for example, H.L. Dreyfus, What Computers Still Can't Do: A Critique of Artificial Reason, 1992; and a review of the book by J. McCarthy, in Artificial Intelligence, vol. 80, no. 1, 1996, pp. 143–150, https://doi.org/10.1016/0004-3702(95)00086-0; L. Suchman, "Feminist STS and the Sciences of the Artificial", in E.J. Hackett et al. (eds.), The Handbook of Science and Technology Studies, 2008, https://doi.org/10.4135/9781412990127; and H. Collins, Artifictional Intelligence: Against Humanity's Surrender to Computers, 2018.

[19] United Nations Office for Disarmament Affairs, "Background on LAWS in the CCW", n.d., https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw.

[20] Gender Shades, MIT Media Lab, http://gendershades.org.

[21] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021, https://www.degruyter.com/document/doi/10.12987/9780300252392/html.

[22] This overview points out how the initial wave of AI research was shaped and limited by the US military, even as subsequent sections of this report indicates how AI should incorporate a far more diverse field of practices from around the globe. Leverhulme Centre for the Future of Intelligence, "Global AI Narratives", https://www.ainarratives.com.

[23] National Research Council, "Developments in Artificial Intelligence", in Funding a Revolution: Government Support for Computing, 1999; and P.N. Edwards, The Closed World: Computers and the Politics of Discourse in Cold War America, 1996.

[24] G. Lewis-Kraus, "The Great AI Awakening", New York Times, 14 December 2016, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html; G. Allen, "Understanding AI Technology", US Department of Defense, 2020, https://www.ai.mil/docs/Understanding AI Technology.pdf.

Recent advances in machine learning use an alternative approach to developing AI. Instead of programming a set of rules, machine learning replicates human cognition through algorithms. These models learn relationships from data that are correlated statistically through trial and error. Research in this field is led by scientists in the United States, China and Europe.[25] This second wave of AI research was not initially tied to national security; rather, it emerged from research conducted at major Internet companies in the 2000s. Earlier attempts to build machine learning models had been limited by the available data and hardware, leading to the field's marginalization. The massive expansion of the Internet in the 2000s resulted in an exponential increase in data collection and computer processing. Advances in machine learning since 2010 have created AI systems that can translate between languages, respond to voice commands and identify photographs.[26] However, the most effective models have clear, easy-to-identify parameters, and experts emphasize that algorithms are limited in their capacity to "generalize or adapt to conditions outside a narrow set of assumptions".[27]

Today, machine learning has become largely synonymous with artificial intelligence, and its applications are being tested by militaries worldwide. Promoters of AI suggest that algorithms will now be able to provide the "decision support" capabilities long theorized for them by the military. In a recent survey of 250 defence technology leaders for allied forces of the North Atlantic Treaty Organization (NATO), all indicated that they were considering machine learning solutions for their armed forces, while 49 per cent had already tested AI in some aspect of defence. To harness these capacities, militaries are engaged in new efforts to collect data relevant to military applications of AI and secure in advance the cloud-computing platforms necessary for machine learning.[28] These initiatives will come not only as the result of technological innovation but will also require substantial investment and infrastructural transformation. The consequences of these changes will emerge in tandem with the technology. As a 2019 report for a United Nations conference states, "Implications for international peace and security of AI's integration into national militaries remains to a large extent unclear."[29]

2.2 Gender norms and technology development

Military applications of artificial intelligence are an acute example of how gender norms can be built into and reinforced by technological systems. Gender norms and other social categories are not stable across cultures; identities respond to conditions that transect local, national and global scales and fit with race, ethnicity, age and ability.[30] As Judy Wajcman explains, "both technology

[25] D. Castro, M. McLaughlin and E. Chivot, "Who Is Winning the AI Race: China, the EU or the United States?", Center for Data Innovation, 19 August 2019, https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states.

[26] G. Lewis-Kraus, "The Great AI Awakening", New York Times, 14 December 2016, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

[27] M. L. Cummings, "The Surprising Brittleness of AI", WomenCorporateDirectors, 2020, https://www.womencorporatedirectors.org/WCD/News/JAN-Feb2020/Reality Light.pdf, p. 1.

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

[29] M. Sisson, "Multistakeholder Perspectives on the Potential Benefits, Risks, and Governance Options for Military Applications of Artificial Intelligence", in B. Finlay, B. Loehrke and C. King, The Militarization of Artificial Intelligence, Workshop report, United Nations, New York, August 2019, https://www.stimson.org/wp-content/uploads/2020/06/TheMilitarization-ArtificialIntelligence.pdf, p. 3.

[30] UN Women, "Intersectional Feminism: What It Means and Why It Matters", 1 July 2020, https://www.unwomen.org/en/news/stories/2020/6/explainer-intersectional-feminism-what-it-means-and-why-it-matters.

and gender are products of a moving relational processes, emerging from collective and individual acts of interpretation".[31] As such, gender and technology are mutually shaping. AI draws on forms of rationality that are typically coded as masculine, found both in war and computing. Feminist critiques of symbolic AI in the 1990s noted how the field was built on a model of intelligence that dissociated cognition from the body.[32]

Today, attempts to replicate human processes through mechanical means continue to privilege logic and gaming. "Hard" data, for example, is commonly opposed to "soft" intelligence, which is associated with empathy, creative problem solving and persuasion – traits associated with femininity.[33] Humanoid robotics uphold gender norms through appearances, voices, mannerisms and movements that imitate differences between men and women.[34] Sex robots, for example, are overwhelmingly figured as female and researchers have pointed to their potential to exacerbate gender-based violence.[35] Meanwhile, robots associated with military applications take on masculine attributes, and are promoted as "super soldiers" for conflict.[36] These stereotypes of female submission and male superiority are reflected in the history of computation itself: the first computer programmers were women, and early coding was described as a secondary, derivative task. It is only much more recently that the field of computer science became a male-dominated profession, now credited with transforming society.[37]

The associations that link together technology, manliness and superiority need not link them in this way, however. Gender-based approaches to technology underline the importance of diversifying the people involved and expanding the framework of expertise. For military AI systems, this means assessing who is involved in their development, implementation, evaluation and regulation and then increasing gender parity. This does not just mean including more women engineers, computer scientists and military commanders in the process, although this is an important step. It also means recognizing a broader range of experts, including scholars of gender and identity, who can speak to the complexities and limitations of mimicking human processes through machine models. If AI is to replicate human intelligence, a narrow understanding of what is human must be replaced by a more complex model that includes the range of bodies, abilities and emotions that are all part of human experience.

[31] J. Wajcman, "Feminist Theories of Technology", Cambridge Journal of Economics, vol. 34, no. 1, 2010, https://doi.org/10.1093/cje/ben057, p. 150.

[32] A. Adam, Artificial Knowing: Gender and the Thinking Machine, 1998, https://doi.org/10.4324/9780203005057.

[33] C. Collett and S. Dillon, AI and Gender: Four Proposals for Future Research, Leverhulme Centre for the Future of Intelligence, 2019, https://www.repository.cam.ac.uk/handle/1810/294360, p. 8.

[34] Ibid., p. 9; N. Ni Loideain and R. Adams, "From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law", King's College London, Dickson Poon School of Law Legal Studies Research Paper Series, 9 November 2018, https://doi.org/10.2139/ssrn.3281807; and A. LaFrance, "Why Do So Many Digital Assistants Have Feminine Names?", The Atlantic, 30 March 2016, https://www.theatlantic.com/technology/archive/2016/03/why-do-so-many-digital-assistants-have-feminine-names/475884.

[35] F.R. Udwadia and J. Illes, "Sex Robots Increase the Potential for Gender-Based Violence", The Conversation, 27 August 2019, http://theconversation.com/sex-robots-increase-the-potential-for-gender-based-violence-122361. See also Stop Killer Robots, "Gender and Killer Robots", 2021, shttps://www.stopkillerrobots.org/gender-and-killer-robots.

[36] S. Cave, K. Coughlan and K. Dihal, "'Scary Robots': Examining Public Responses to AI", AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, January 2019, pp. 331–337, https://doi.org/10.1145/3306618.3314232.

[37] M. Hicks, Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing, 2017; and N.L. Ensmenger, The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise, 2010.

2.3 How is data used in machine learning?

Conventional accounts of the rise of machine learning link its recent success with the mass collection of data that has been enabled by the Internet. Machine learning developed for military applications draws on these existent models and data. On an average day in 2019, approximately 350 million photographs were uploaded to Facebook and 500 million tweets were posted.[38] Proprietary data sets owned by major AI companies include texts, voice recordings and images scraped from the Internet, mobile phones and other sources. Machine learning finds patterns in this data through statistical correlation. Outputs based on these relations are used to chat with customers, respond to voice commands and recognize faces.

Data-driven processes purport to be neutral and objective. Yet, data for a machine learning model must fit a complex and varied world into a set of discrete classifications and individual data points.[39] Data is collected, labelled and organized by teams of engineers and piecework digital labourers; it is not simply extracted from the Internet, nor is it static. Data training sets can be flawed due to incomplete data, low-quality data, incorrect or false data, or discrepant data.[40] These limitations are all at play, often in overlapping ways, in considerations of gender and machine learning.

A review of publicly available information on 133 biased AI systems, deployed across different economic sectors from 1988 to 2021, found that 44 per cent (59 systems) exhibited gender bias, including 26 per cent (34 systems) that exhibited both gender and racial biases.[41] A 2016 study of speech-recognition software found that the program was 70 per cent more likely to accurately recognize men’s speech than women’s speech.[42] In a 2017 evaluation of three commercially available facial-recognition algorithms, researchers found that the maximum error rates for lighter-skinned men was less than 1 per cent, while the misclassification rates for darker-skinned women went up to 35 per cent.[43] A more recent evaluation of commercial facial recognition continued with the finding that “all tested gender classifiers still favor the male category” and “dark-skinned females tend to yield higher classification error rates”, although error rates were lower.[44]

Best practices for AI rely on a careful review of data sets and their limitations. Developers make highly consequential decisions in determining which sets of data should be used to train even the most complex deep-learning algorithms. For example, the “Rosetta Stone” for Google Translate was the complete bilingual records of Canadian Parliament.[45] This choice meant that machine

[38] Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021, https://www.degruyter.com/document/doi/10.12987/9780300252392/html, p. 106.

[39] Ibid., p. 135.

[40] Ibid., p. 4.

[41] G. Smith and I. Rustagi, “When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity”, Stanford Social Innovation Review, 31 March 2021, https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity.

[42] 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, https://doi.org/10.18653/v1/W17-1606.

[43] J. Buolamwini and T. Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, Proceedings of Machine Learning Research, no. 81 (2018), pp. 77–91, http://gendershades.org/docs/ibm.pdf.

[44] K. Karkkainen and J. Joo, “FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation”, In 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2021, pp. 1547–57, https://doi.org/10.1109/WACV48630.2021.00159, p. 1555.

[45] G. Lewis-Kraus, “The Great AI Awakening”, New York Times, 14 December 2016, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

translation through machine learning was improved by using a data set that itself modelled translation, not just by the quantity of information. This example points out how choices made by developers about what data to use in a machine learning model have an impact on outcomes. It also suggests the importance of including diverse viewpoints in the design process.

Military AI requires particularly close attention to the data being used to model war. The challenges of data-collection practices for military purposes can be seen in new efforts to mine data globally. In 2020, Nand Mulchandani, then Chief Technology Officer of the US Department of Defense's Joint Artificial Intelligence Center, observed that the data set for current AI systems

> is not representative of say, global terrain, or global information, or even things like faces. So when you think of the diversity of … humankind out there … if you're doing something like facial recognition or something, the training data set from a testing and representative perspective is so important.[46]

Given that conflict environments are harsh, dynamic and adversarial, there will always be more variability on the battlefield than in the limited sample of data that will be used to develop military applications of AI.[47] This complexity includes the individuals on the battlefield and whether they or not are combatants – determinations that are often linked to, but are not reducible to, gender.

2.4 How are social biases reinforced through data, algorithms and machine learning processes?

Machine learning algorithms use correlations found in data sets to model human processes, which are then used to generate, for example, chat messages, image descriptions and responses to voice commands. Algorithms act as a set of instructions. During the training process, the computer creates a statistical model to accomplish a specific task. Algorithms called "learners" are trained on labelled data examples. These processes inform algorithms known as "classifiers" about how to best analyse the relation between new inputs and the desired outputs or predictions for the machine learning model. Instrumental decisions are made by engineers in the evaluation of the algorithm to determine whether outcomes are accurate.[48]

Artificial intelligence encodes the patterns found in the data it is trained on. A machine learning model trained on Google News articles, for example, exhibited disturbing patterns of female/male gender stereotypes, reproducing historical bias. The model replicated the existent associations between "computer scientist" and "man", for example, and these connections were augmented by the artificial neural network.[49] These patterns also appear in machine translation,

[46] C. Todd Lopez, "Artificial Intelligence Deployment Requires Diverse Image Data", US Department of Defense, 20 July 2020, https://www.defense.gov/Explore/News/Article/Article/2280560/artificial-intelligence-deployment-requires-diverse-image-data (ellipses in original).

[47] A. Holland, Known Unknowns: Data Issues and Military Autonomous Systems, UNIDIR, 2021, https://doi.org/10.37559/SecTec/21/AI1, p. 1.

[48] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021, https://doi.org/10.12987/9780300252392, pp. 96–97. See also, Sara Hooker, "Moving beyond 'Algorithmic Bias Is a Data Problem'", Patterns, vol. 2, no. 4, April 2021, 100241, https://doi.org/10.1016/j.patter.2021.100241.

[49] T. Bolukbasi et al., "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings", arXiv 1607.06520, 21 July 2016, https://arxiv.org/abs/1607.06520.

which default to "he said", for example, when translating into English from languages without gendered pronouns.[50] Meanwhile, social media platforms show more highly paid job advertisements to men than to women.[51] Such examples expose the circularity of data-driven models, which take current conditions as a given. These relationships can be amplified by aggregate bias, observed when a specific subset of samples does not fit with the dominant model – this is found, for example, when natural language processing is applied to specific dialects.[52]

Rather than providing an objective corrective to human prejudice, machine learning models can replicate and amplify systemic inequalities. A 2016 study of a computer program designed to evaluate potential for recidivism within the criminal justice system found that Black minorities in the United States were twice as likely to be categorized as high risk.[53] The analysis included another, less-mentioned detail: the system "unevenly predicts recidivism between genders", which makes women appear to be a higher risk than they are.[54] The company that built the model nevertheless justified its product, indicating that the statistical method was accurate and that the differences noted by the study instead reflected social inequities.[55] Follow-up research indicated that the algorithmic model used proxy variables that had substantially different meanings depending on the race or gender of the person being evaluated. The algorithm not only reflected historical bias; it amplified these outcomes based upon the measurements it used.[56]

The consequences of bias in machine learning are augmented in a military context. Consider, for example, a machine translation program used by military intelligence that would assign "male" as the gender of a person of unspecified gender. Or an algorithm designed to recruit military personnel that might pass over qualified women candidates given their historically low levels of participation in the armed forces. Or a voice-control system that does not recognize the voice of a woman pilot. Or an automated system designed to provide emergency relief that does not include provisions specific to women and girls. On top of these considerations, one must contemplate the potential consequences of gender and racial biases in autonomous weapon systems. The criteria that will inform who is and is not a combatant – and, therefore, a target – will be likely to involve gender, age, race and ability. Assumptions about men's roles, for example, may miscategorize civilian men as combatants due to encoded gender biases among human operators as well as within the data-driven process itself.

[50] Gendered Innovations, "Machine Translation: Analyzing Gender", Stanford University, n.d., https://genderedinnovations.stanford.edu/case-studies/nlp.html#tabs-2.

[51] B. Imana, A. Korolova and J. Heidemann, "Auditing for Discrimination in Algorithms Delivering Job Ads", Proceedings of the Web Conference 2021, April 2021, pp. 3767–3778, https://doi.org/10.1145/3442381.3450077; and A. Kofman and A. Tobin, "Facebook Ads Can Still Discriminate against Women and Older Workers, Despite a Civil Rights Settlement", ProPublica, 13 December 2019, https://www.propublica.org/article/facebook-ads-can-still-discriminate-against-women-and-older-workers-despite-a-civil-rights-settlement.

[52] H. Suresh and J. Guttag, "Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle", MIT Case Studies in Social and Ethical Responsibilities of Computing, summer 2021, https://doi.org/10.21428/2c646de5.c16a07bb.

[53] J. Angwin et al., "Machine Bias", ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. See also K. Lum and W. Isaac, "To Predict and Serve?", Significance, vol. 13, no. 5, 2016, pp. 14–19, https://doi.org/10.1111/j.1740-9713.2016.00960.x; and B.J. Jefferson, Digitize and Punish: Racial Criminalization in the Digital Age, 2020.

[54] J. Larson et al., "How We Analyzed the COMPAS Recidivism Algorithm", ProPublica, 23 May 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[55] W. Dieterich, C. Mendoza and T. Brennan, "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", Northpointe, 8 July 2016, http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf.

[56] H. Suresh and J. Guttag, "Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle", MIT Case Studies in Social and Ethical Responsibilities of Computing, summer 2021,

2.5 How do global inequalities have an impact on AI development and deployment?

Theorists studying the future of military AI suggest that computers have the potential to replace people on the battlefield and imagine a conflict environment that is composed primarily of machines. Internet wars are depicted as incorporeal and automatic, but military applications of AI require substantial research and resources, again stressing the vast network of human actors necessary to build, implement and maintain machine learning systems.[57] Importantly, these factors are unevenly distributed worldwide. A widely cited index evaluating global competitiveness in AI emphasizes how machine learning is predicated on the cultivation of a talented pool of computer scientists, access to key hardware components and the extraction of data. Eighty-five per cent of the world's fastest computers are located in the United States, Europe and China, and these three regions account for 77 per cent of the world's PhDs in fields associated with machine learning. China leads the world in data collection, drawing on more than 394 million broadband subscriptions and more than 525 million individuals using mobile payments.[58]

Gender inequalities are woven into this data, given that women comprise only 24 per cent of the world's computer scientists and account for just 12 per cent of publications on machine learning.[59] Global divides further compound these inequalities, as many of the support workers – such as the people hired to label the data sets described above – are located in the Global South.[60] No statistics exist for Internet contract workers, but the available research underlines how the intersection of race, age and gender has an impact on this work and draws from historically high numbers of women in service-related jobs.[61]

Men overwhelmingly lead the high-level implementation of AI projects, however. These trends are visible in the global employment data of Amazon, one of the largest employers in the world (which, to the company's credit, tracks gender parity in its workforce and publishes its progress). Women comprise 45 per cent of its global workforce, and they proportionally hold lower-level jobs, accounting for only 31 per cent of global corporate positions and 22 per cent of the senior leadership.[62] In the case of military applications of AI, these patterns would intersect with gender inequities within the defence industry, where, for example, in the United States, women accounted for 25 per cent of the employees and 22 per cent of the senior leadership in 2019.[63] More research needs to be done in the field of artificial intelligence, and for military applications of AI in particular, to establish baseline gender representation and address imbalances.

Turning to the hardware and data that are necessary to run AI systems further emphasizes the vast resources necessary for

[57] M.C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power", Texas National Security Review, vol. 1, no. 3, May 2018, https://doi.org/10.15781/T2639KP49.

[58] D. Castro, M. McLaughlin and E. Chivot, "Who Is Winning the AI Race: China, the EU or the United States?", Center for Data Innovation, 19 August 2019, https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states.

[59] Y. Yuan, "Exploring Gender Imbalance in AI: Numbers, Trends, and Discussions", Synced, 13 March 2020, https://syncedreview.com/2020/03/13/exploring-gender-imbalance-in-ai-numbers-trends-and-discussions.

[60] L.C. Irani and M.S. Silberman, "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk", in CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 2013, pp. 611–620, https://doi.org/10.1145/2470654.2470742; and C. Metz, "A.I. Is Learning from Humans. Many Humans", New York Times, 16 August 2019, https://www.nytimes.com/2019/08/16/technology/ai-humans.html.

[61] S.T. Roberts, "Commercial Content Moderation: Digital Laborers' Dirty Work'", Media Studies Publications, vol. 12, 2016, https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1012&context=commpub.

[62] Amazon, "Our Workforce Data", 31 December 2020, https://www.aboutamazon.com/news/workplace/our-workforce-data.

[63] Al Root, "The Captains of Industry Are Still Mostly Men – but Not in Defense", Barron's, 20 June 2019, https://www.barrons.com/articles/defense-industry-women-ceos-51560974063.

machine learning. Research published in 2019 explains that new methodologies and hardware have enabled significant advances in natural language processing. However, these improvements depend on the availability of exceptionally large computational resources that necessitate substantial energy consumption. The models are costly to train and develop "both financially, due to the cost of hardware and electricity or cloud compute time, and environmentally, due to the carbon footprint required to fuel modern tensor processing hardware".[64] The high costs of machine learning and its dependence on data to carry out natural language processing suggest, for example, that these tools will never be available in thousands of languages. The risk is not only that just a few countries will be able to develop comprehensive military AI, but also that specific cultural assumptions – including norms associated with gender roles and threat detection – will have undue influence in AI products. This potential is already suggested in the worldwide use of AI for surveillance: "At least seventy-five out of 176 countries globally are actively using AI technologies for surveillance purposes. This includes: smart city/ safe city platforms (fifty-six countries), facial recognition systems (sixty-four countries), and smart policing (fifty-two countries)."[65]

Finally, military applications of AI raises important concerns about data collection, privacy and sovereignty worldwide. Approximately 40 per cent of the world's population does not have access to the Internet, and the majority of the world's unconnected people are women and girls.[66] These statistics underline the limits of relying on data from the Internet as being representative of "humankind" writ large. As a recent report on the digital economy observes, "Africa and Latin America together account for less than 5 per cent of the world's colocation data centres. If left unaddressed, these divides will exacerbate existing income inequalities."[67] The observation also highlights how countries in the Global South may not control data sets tied to their own populations. This affects not only economic development but also privacy and national security.

These global inequalities have an impact on who will use, benefit from and be harmed by military applications of AI. Apparently benign practices of data collection – for example, digital photographs uploaded to the Internet – could become the basis for lethal weapons and are likely to already be part of mass-surveillance programmes. Concerns about gender-based online violence indicate how algorithmic models can perpetuate harms, for example, through AI fakes and other cybersecurity concerns.[68] For many people worldwide, inclusion in algorithmic systems will ultimately result in the development of technologies that are designed to surveil, criminalize and control them.[69]

[64] E. Strubell, A. Ganesh and A. McCallum, "Energy and Policy Considerations for Deep Learning in NLP", arXiv 1906.02243, 5 June 2019, https://arxiv.org/pdf/1906.02243.pdf, p. 1.

[65] S. Feldstein, The Global Expansion of AI Surveillance, Carnegie Endowment for International Peace, September 2019, https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf, p.7.

[66] International Telecommunication Union, "Bridging the Gender Divide", July 2021, https://www.itu.int/en/mediacentre/backgrounders/Pages/bridging-the-gender-divide.aspx.

[67] United Nations Conference on Trade and Development (UNCTAD), Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries, 2019, https://unctad.org/system/files/official-document/der2019_en.pdf, p. xvi.

[68] K. Millar, J. Shires and T. Tropina, "Gender Approaches to Cybersecurity", UNIDIR, 2021, https://doi.org/10.37559/GEN/21/01; and A. Sey and N. Hafkin, "Taking Stock: Data and Evidence on Gender Equality in Digital Access, Skills, and Leadership", United Nations University, 2019, https://i.unu.edu/media/cs.unu.edu/attachment/4040/EQUALS-Research-Report-2019.pdf.

[69] C. Barabas, "Beyond Bias: Contextualizing 'Ethical AI' Within the History of Exploitation and Innovation in Medical Research", Medium, 8 January 2020, https://medium.com/mit-media-lab/beyond-bias-contextualizing-ethical-ai-within-the-history-of-exploitation-and-innovation-in-d522b8ccc40c.

3. Case studies: challenges in deploying AI systems

3.1. Finding bias in ImageNet

One of the largest publicly available databases for images is ImageNet, which has been instrumental in advancing computer vision and deep learning research.[70] As its authors explain, “The dataset was created to benchmark object recognition – at a time when it barely worked. … An emerging problem now is how to make sure computer vision is fair and preserves people’s privacy.”[71] ImageNet serves as a case study of how multiple layers of human actions shape and limit a machine learning model. ImageNet harvested more than 14 million images from the Internet and organized them into over 20,000 categories. Piecework labourers were paid per image to populate the database. A review in 2019 revealed that the “Person” heading included such offensive categories as “alcoholic”, “ape-man”, “hooker” and “slant eye”.[72] Images uploaded to these subcategories contained stereotypes, errors and absurdities.[73] More than 600,000 images within the “Person” category were removed after the review, along with over 1,000 specific labels that were deemed inherently offensive or inappropriate.[74]

Embedded within ImageNet are culturally specific gender stereotypes, found in the images and in the way in which they are coded through language and translated. With machine learning tools trained from ImageNet, “A white woman wearing a white wedding dress is labelled as a bride, whereas a North Indian woman wearing a wedding sari …. is labelled as performance art.”[75] Researchers found that over half the images from ImageNet were from the United States and the United Kingdom, while images from China and India, the two most populous countries in the world, accounted for less than 3 per cent of the images.[76] Another study found that only 66 per cent of the labels from ImageNet generated through machine translation in Arabic were accurate.[77]

ImageNet indicates how training data and instructions shape the outcomes of machine learning models. AI researchers have since developed a distinct tool to test object recognition, called ObjectNet. It uses photos taken by paid freelancers and shows objects tipped on their side, shot at odd angles and displayed in clutter-strewn rooms. The accuracy of images tested on ImageNet fell from

[70] ImageNet, https://image-net.org.

[71] K. Yang et al., “Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy”, in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 547–558, https://doi.org/10.1145/3351095.3375709.

[72] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021, https://www.degruyter.com/document/doi/10.12987/9780300252392/html, p. 109.

[73] Ibid.

[74] K. Yang et al., “Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy”, in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, pp. 547–558, https://doi.org/10.1145/3351095.3375709.

[75] C. Collett and S. Dillon, AI and Gender: Four Proposals for Future Research, Leverhulme Centre for the Future of Intelligence, 2019, https://www.repository.cam.ac.uk/handle/1810/294360, p. 19.

[76] S. Shankar et al., “No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World”, ArXiv 1711.08536, 22 November 2017, http://arxiv.org/abs/1711.08536.

[77] A. Alsudais, “Image Classification in Arabic: Exploring Direct English to Arabic Translations”, IEEE Access, vol. 7, 2019, https://doi.org/10.1109/ACCESS.2019.2926924.

97 per cent to 50–55 per cent when they were evaluated against ObjectNet.[78] These limitations point out how the accuracy of machine learning models is tied to their evaluation, and they illuminate the need for military AI to develop robust methods to test and review machine learning models. This concern is amplified when people, not just objects, are identified by artificial intelligence.

3.2 Using data sets and algorithms to distinguish between civilians and combatants

A key point of consensus among the Group of Governmental Experts (GGE) convened to develop a normative and operational framework for lethal autonomous weapon systems (LAWS) within the 1981 Certain Conventional Weapons (CCW) Convention is that human–machine interactions established by the weapon system are crucial to ensure compliance with international humanitarian law.[79] While popular depictions of military applications of AI often oppose human and machine, the discussions within the GGE on LAWS indicate agreement on the requirement to maintain a close association between the two in the development of emergent technologies.[80] According to the Guiding Principles drawn up by the GGE and agreed upon by the CCW States parties in 2019, autonomous weapons must fit within a responsible chain of human command and control.[81] They should also be governed by rules on the conduct of hostilities, including distinction, proportionality and precaution in attack; they should retain the ability to determine between civilians and combatants, as well as civilian objects and military objects; and they should make context-based decisions about the potential impact of an attack on civilians.[82]

A recent training exercise for military cadets aimed to model this problem. Students guided a small robotic tank armed with a spear against balloon targets. One colour of balloon – red – indicated enemy fighters, while the other balloons represented civilians. Students reacted in a range of ways: some taught their miniature tanks to turn away from civilians, while others "program[med] their tanks with a more gung-ho approach, sometimes leading the machines to slay balloons – including 'civilians' – with abandon".[83] This exercise suggests how outcomes associated with autonomous weapon systems remain tied to human decisions. Actions by users– and by extension, regulatory bodies – about how to train robots result in significantly different outcomes. Yet, one might question whether this exercise serves as an accurate model of the problems that future soldiers will confront using AI.

In the simulation, combatants were clearly designated by a colour. While military battlefields have long relied on uniforms to designate soldiers from civilians, contemporary conflict environments are far more complex. One might instead ask: What data set would be used to determine who is a soldier and

[78] Martineau, "This Object-Recognition Dataset Stumped the World's Best Computer Vision Models", MIT News, 10 December 2019, https://news.mit.edu/2019/object-recognition-dataset-stumped-worlds-best-computer-vision-models-1210.

[79] CCW Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Report of the 2019 Session, CCW/GGE.1/2019/3, 25 September 2019, https://undocs.org/CCW/GGE.1/2019/3.

[80] R. Slayton, "The Promise and Risks of Artificial Intelligence: A Brief History", War on the Rocks, 8 June 2020, https://warontherocks.com/2020/06/the-promise-and-risks-of-artificial-intelligence-a-brief-history.

[81] CCW Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Report of the 2019 Session, CCW/GGE.1/2019/3, 25 September 2019, https://undocs.org/CCW/GGE.1/2019/3, p. 13.

[82] V. Boulanin et al., Limits on Autonomy in Weapons Systems: Identifying Practical Elements of Human Control, SIPRI, June 2020, https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy.pdf.

[83] Z. Fryer-Biggs, "Can Computer Algorithms Learn to Fight Wars Ethically?", Washington Post, 17 February 2021, https://www.washingtonpost.com/magazine/2021/02/17/pentagon-funds-killer-robots-but-ethics-are-under-debate.

who is a civilian in non-conventional conflict scenarios? How would this determination be made by an algorithm, given that markers that distinguish combatants and non-combatants are not always predictable? Who decides whether a machine learning model is sufficiently accurate? The previous sections cover the limits of machine learning data and the persistence of human bias in the outputs of AI programs. If the machine learning model does not accurately distinguish between combatant and non-combatant, the programmed model will not function as expected by users. It would not matter if the users chose to protect civilians if they are not adequately represented through algorithmic processes.

If recent conflicts were to be used as sources of data sets, that would have several implications. In the case of the war in Afghanistan, for instance, gender and age have been deciding factors in whether a person is an acceptable military target or not.[84] Training data would be likely to identify military-aged men as potential targets and women and children as non-combatants, as this is the pattern provided. The data would fix a contingent determination as a rule. Civilian men would be included in this pattern. Moreover, a data set from the war in Afghanistan would rely on culturally specific patterns, potentially using dress and head covering, for example, to make decisions. Yet, material indicators of gender and age are not universal, and even the use of uniforms and weapons vary from region to region, while domain adaptability – which refers to an AI's capacity to adjust between two settings that are not the same – is limited in algorithmic processes.[85] Today's AI relies on previous data to make predictions, and its outputs are limited to recognizable patterns. Even if gender and age are not explicit in the machine learning model, patterns drawn from neutral characteristics, such as uniforms or evidence of weapons, could still implicitly incorporate gender norms.

3.3 Autonomous weapons and humanitarian impacts

On 27 March 2020, during a military skirmish between the United Nations-recognized Government of Libya and armed groups affiliated with the opposition, logistics convoys and retreating forces of the opposition were potentially "hunted down and remotely engaged" by uncrewed combat aerial vehicles (UCAVs) equipped with autonomous target recognition.[86] The weapon in question is described as "capable of selecting and engaging human targets based on machine-learning object classification".[87] It operates as a so-called kamikaze, which detonates in the proximity of a person through remote control or an autonomous modality.[88] The weapon created confusion: it is unclear whether there was a remote operator or not. There is no information about the machine learning model used by the weapon. An online video promoting the system's capabilities shows the weapon's munitions exploding in front of a square target.[89] This test scenario – solving a simple problem of static object recognition of a square – is utterly distinct from tracking a combatant's face in a war environment.

[84] S. Laastad Dyvik, "Women as 'Practitioners' and 'Targets': Gender and Counterinsurgency in Afghanistan", International Feminist Journal of Politics, vol. 16, no. 3, 2014, pp. 410–429, https://doi.org/10.1080/14616742.2013.779139.

[85] D.S. Hoadley and N.J. Lucas, "Artificial Intelligence and National Security", Congressional Research Service, US Congress, 16 April 2018, https://crsreports.congress.gov/product/pdf/R/R45178/3, p. 30.

[86] Security Council, S/2021/229, 8 March 2021, https://undocs.org/S/2021/229, p. 17.

[87] H. Nasu, "The Kargu-2 Autonomous Attack Drone: Legal and Ethical Dimensions", Articles of War, Lieber Institute, 10 June 2021, https://lieber.westpoint.edu/kargu-2-autonomous-attack-drone-legal-ethical.

[88] Ibid.

[89] STM, "Kargu Rotary Wing Attack UAV", n.d., https://www.stm.com.tr/en/kargu-autonomous-tactical-multi-rotor-attack-uav.

This example indicates how users of LAWS can draw on uncertainty to deflect criticism, even as the systems contribute to and potentially amplify existing patterns of war.[90] While there is no indication that the autonomous weapon violated wartime rules of engagement,[91] its deployment in Libya is concerning nonetheless.[92] Its use, moreover, suggests how battlefield experiments with autonomous weapons overlay significant humanitarian crises. Notably, the deployment of these weapons appears in a 500-page report by a panel of experts on Libya, which observes that "[b]oth parties to the conflict have committed acts that violate the applicable legal framework" and underlines how "[c]ivilian casualties increased owing to the escalation in hostilities during the first half of 2020 and are attributable mainly to ground fighting, explosive remnants of war, targeted killings and air strikes, the first two being the leading causes of death".[93]

[90] F. Slijper, Where to Draw the Line: Increasing Autonomy in Weapon Systems ¬ – Technology and Trends, PAX, November 2017, https://paxforpeace.nl/media/download/pax-report-where-to-draw-the-line.pdf.

[91] H. Nasu, "The Kargu-2 Autonomous Attack Drone: Legal and Ethical Dimensions", Articles of War, Lieber Institute, 10 June 2021, https://lieber.westpoint.edu/kargu-2-autonomous-attack-drone-legal-ethical

[92] Airforce Technology, "Kargu Rotary-Wing Attack Drone", https://www.airforce-technology.com/projects/kargu-rotary-wing-attack-drone.

[93] Security Council, S/2021/229, 8 March 2021, https://undocs.org/S/2021/229, p. 10.

4. Countering bias and promoting ethical approaches to military applications of AI

4.1. Countering bias in machine learning

Researchers addressing algorithmic fairness underline how no single approach to bias is effective, and bias can be introduced at multiple stages in the machine learning life cycle (see figure 1).[94] Three such researchers explain that "we need moral reasoning and domain-specific considerations to determine which test(s) are appropriate, how to apply them, determine whether the findings indicate wrongful discrimination, and whether an intervention is called for".[95]

The most effective practice for addressing the limitations of machine learning is to make explicit the suppositions of data and algorithmic models. "Datasheets for Datasets", authored by a team of ethical AI researchers led by Timnit Gebru, proposes "that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses, and so on".[96] The proposal draws on standards from the electronics industry; the aim is to facilitate communication between the creators and consumers of data sets, as well as to prioritize transparency and accountability in machine learning models.[97] "Model Cards for Model Reporting", authored by an associated team of ethical AI researchers led by Margaret Mitchell, aims to evaluate trained machine learning models. A model card would document the system's performance "in a variety of conditions, such as across different cultural, demographic, or phenotypic groups . . . and intersectional groups … that are relevant to the intended application domains".[98] Gender is a key category in the benchmark. The authors apply their proposal to an algorithm designed to determine whether a person is smiling. The model card revealed that "the false discovery rate on older men is much higher than that for other groups", which "means that many predictions incorrectly classify older men as smiling when they are not".[99] They suggest how the findings might be corrected through additional fine-tuning of the algorithm on images of older men.

While making apparent the assumptions and limitations of machine learning models is a necessary starting point in order to counter bias, these measures alone will not prevent the harmful use of artificial intelligence. Rather, transparency and independent evaluations like those proposed in "Datasheets for Datasets" and "Model Cards for Model Reporting" should be backed by industry standards and government regulations that outline clear standards to limit bias and ensure accountability to ethical standards. A recent study of AI principles in Ireland and the United Kingdom found that "ethics guidelines are generically performative, operating at a

[94] S. Barocas, M. Hardt and A. Narayanan, Fairness and Machine Learning: Limitations and Opportunities, 2019, https://fairmlbook.org, chapter 5. See also UNIDIR, Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies, 22 August 2018, https://unidir.org/publication/algorithmic-bias-and-weaponization-increasingly-autonomous-technologies.

[95] Ibid., chapter 5, p. 1.

[96] T. Gebru et al., "Datasheets for Datasets", arXiv 1803.09010, 19 March 2020, https://arxiv.org/abs/1803.09010, p. 2.

[97] Ibid.

[98] M. Mitchell et al., "Model Cards for Model Reporting", in Proceedings of the Conference on Fairness, Accountability, and Transparency, January 2019, pp. 220–229, https://doi.org/10.1145/3287560.3287596.

[99] Ibid.

linguistic level to assuage and deflect critique and regulation".[100] The researchers found that ethics discourses and solutions, rather than serving as an effective tool for governance, served as assurance for investors and the general public. In the case of military AI, bias must be addressed in all stages – the conceptualization and design of the machine learning model, data-collection and labelling process, testing and evaluation, and disposal of the system (see figure 1).

Figure 1.
Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle[101]

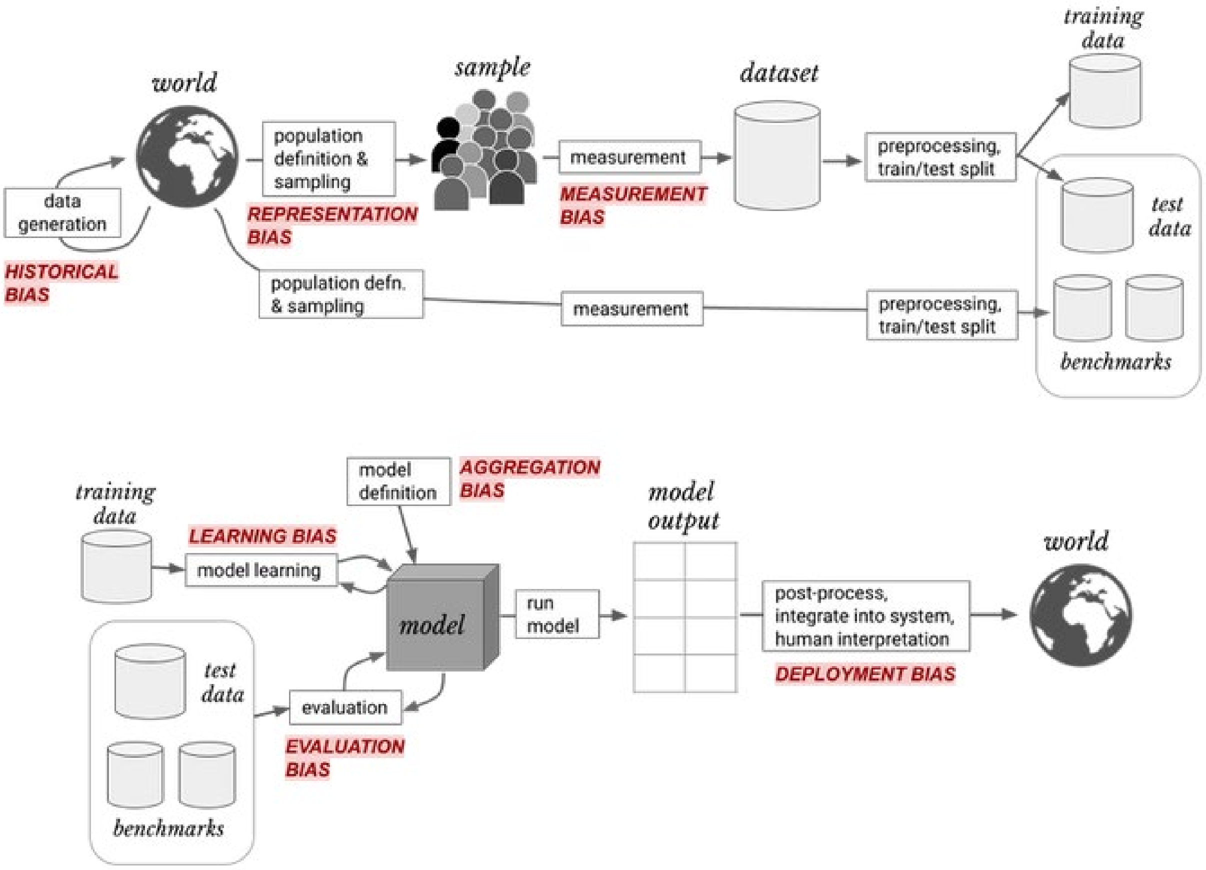


Figure 1. *(Top)* The data generation process begins with data collection. This process involves defining a target population and sampling from it, as well as identifying and measuring features and labels. This data set is split into training and test sets. Data is also collected (perhaps by a different process) into benchmark data sets. *(Bottom)* A model is defined and optimized on the training data. Test and benchmark data are used to evaluate it, and the final model is then integrated into a real-world context. This process is naturally cyclic, and decisions influenced by models affect the state of the world that exists the next time the data is collected or decisions are applied. In red, we indicate where in this pipeline different sources of downstream harm might arise.

[100] A. Kerr, M. Barry and J.D. Kelleher, "Expectations of Artificial Intelligence and the Performativity of Ethics: Implications for Communication Governance", Big Data & Society, vol. 7, no. 1, 2020, https://doi.org/10.1177/2053951720915939, p. 10.

[101] H. Suresh and J. Guttag, "Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle", MIT Case Studies in Social and Ethical Responsibilities of Computing, summer 2021, https://doi.org/10.21428/2c646de5.c16a07bb.

4.2. Prospects and limits for incorporating gender-based ethics into military applications of AI

As militaries around the world integrate machine learning processes into the armed forces, commitments to ethical AI need to make clear how they address bias related to specific categories of people – notably, gender and its intersection with race, ethnicity, age and ability. Ethical AI advocates in the military today acknowledge many of the current limits to machine learning processes.[102] The United States Department of Defense, for example, notes that AI must have "explicit, well-defined uses, and the safety, security and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across [the AI capabilities'] entire life-cycles".[103] Research scholars have proposed that best practices might distinguish between "tightly bounded", "loosely bounded" and "unbounded" problems to determine the suitability of a military scenario to machine learning, given the potential advantages of algorithmic models in situations that can be clearly defined.[104] Moreover, AI could be used to provide a more complete record of conflict environments, for example, indicating the time and location of every shot fired and potentially improving accountability.[105] This commitment to understanding the limitations of AI and ensuring that the system improves transparency needs to extend to gender.

Military users should be trained to recognize the multiple ways in which bias can be introduced into AI throughout the life cycle of the system and the harms that can be caused as a result of these limitations. As a human–machine centred approach to military AI indicates, "All individuals who deal with AI technology have to exercise due diligence", meaning that, at every step, the operator should examine how his or her actions and inactions could contribute to potential harms.[106] Another set of authors observe that to use machine learning models ethically, military personnel must be aware of their own analytic limitations, as well as those of the algorithm: "[H]umans will require classroom and experiential training to understand algorithmic flaws and to gain confidence in their ability to diagnose them".[107]

If users do not understand how algorithm models work, the technology could introduce delays that work against AI decision making. Conversely, overconfidence in algorithmic logic can create an environment in which human operators fail to question AI predictions.

[102] See, for example, the account of a recent meeting held by the US Department of Defense Joint AI Center and the United Kingdom, Canada, Denmark, Estonia, France, Norway, Australia, Japan, the Republic of Korea, Israel, Finland and Sweden in S.J. Freedberg, "Military AI Coalition of 13 Countries Meets on Ethics", Breaking Defense, 16 September 2020, https://breakingdefense.com/2020/09/13-nations-meet-on-ethics-for-military-ai. See also Z. Stanley-Lockman, "Responsible and Ethical Military AI: Allies and Allied Perspectives", Center for Security and Emerging Technology, August 2021, https://cset.georgetown.edu/wp-content/uploads/CSET-Responsible-and-Ethical-Military-AI.pdf; China–UK Research Centre for AI Ethics and Governance, "The Ethical Norms for the New Generation Artificial Intelligence, China", 27 September 2021, https://ai-ethics-and-governance.institute/2021/09/27/the-ethical-norms-for-the-new-generation-artificial-intelligence-china; and J. Edmonds et al., "Artificial Intelligence and Autonomy in Russia", Center for Naval Analyses, May 2021, https://www.cna.org/CNA_files/centers/CNA/sppp/rsp/russia-ai/Russia-Artificial-Intelligence-Autonomy-Putin-Military.pdf.

[103] C. Todd Lopez, "DOD Adopts 5 Principles of Artificial Intelligence Ethics", US Department of Defense, 25 February 2020, https://www.defense.gov/Explore/News/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics.

[104] D. Blair et al., "Humans and Hardware: An Exploration of Blended Tactical Workflows Using John Boyd's OODA Loop", in The Conduct of War in the 21st Century: Kinetic, Connected and Synthetic, 2021, https://doi.org/10.4324/9781003054269, p. 100.

[105] Galliot and J. Scholtz, "The Case for Ethical AI in the Military", in M. Dubber, F. Pasquale and S. Das (eds.), The Oxford Handbook of Ethics of AI, 2020, p. 6.

[106] Ibid., pp. 5–6.

[107] D. Blair et al., "Humans and Hardware: An Exploration of Blended Tactical Workflows Using John Boyd's OODA Loop", in The Conduct of War in the 21st Century: Kinetic, Connected and Synthetic, 2021, https://doi.org/10.4324/9781003054269.

These frameworks must also attend to the specific ways in which machine learning can explicitly or implicitly encode gender, as well as how intersecting categories of race, age and ability are factors that can be part of military applications of machine learning.

Ethical approaches to military applications of AI are at once important and contradictory with regard to targeting. The people who must be adequately represented by military AI systems, in such cases, are also targets who can be killed. They will be likely to have different cultural and social markers from the people who built the AI, including gender norms. These characteristics may not be adequately captured through the determinations of machine learning. For example, a program might over-identify men as combatants, both obscuring women soldiers and making targets of civilian men.[108]

This potential for men to be overrepresented, and women underrepresented, among combatants identified on the battlefield has many possible consequences. Not least among these is the weaponization of the very categories of civilian and combatant that international humanitarian law aims to protect.[109] AI has been shown to be susceptible to minor alterations, and one can extrapolate how militaries might adapt aspects of camouflage that would depict soldiers as civilians to confuse AI. Such techniques could incorporate gender.[110] Given the complexity associated with these issues, there is no straightforward technological fix.

4.3 Towards a framework for gender, peace, security and AI

Recent scholarship complicates the simple associations linking men to war and women to peace, even as it demonstrates how masculine stereotypes of war remain privileged in international affairs, often in the guise of neutral, human actions.[111] United Nations Security Council resolution 1325 on Women, Peace and Security recognizes the importance of women in conflict prevention and peacebuilding, women's rights during and after conflict, and the specific needs of women during relief and recovery.[112] A task force from UN Women explains, "when women are at the negotiating table, peace agreements are more likely to last 15 years or longer".[113] Yet, women continue to be underrepresented in peace treaties and in arms control, non-proliferation and disarmament.

Gender mainstreaming based on Security Council resolution 1325 calls for the "incorporation of gender analyses and gender perspectives in all aspects of military operations".[115] Over the past two decades, there has been growing recognition of the importance of integrating gender considerations

[108] S. Shoker, Military-Age Males in Counterinsurgency and Drone Warfare, 2021, https://doi.org/10.1007/978-3-030-52474-6.

[109] Stop Killer Robots, "Gender and Killer Robots", 2021, https://www.stopkillerrobots.org/gender-and-killer-robots.

[110] W. Knight, "Military Artificial Intelligence Can Be Easily and Dangerously Fooled", MIT Technology Review, 21 October 2019, https://www.technologyreview.com/2019/10/21/132277/military-artificial-intelligence-can-be-easily-and-dangerously-fooled.

[111] C.C. Confortini, Intelligent Compassion: Feminist Critical Methodology in the Women's International League for Peace and Freedom, 2012, https://doi.org/10.1093/acprof:oso/9780199845231.001.0001; P. Kirby and L.J. Shepherd, "The Futures Past of the Women, Peace and Security Agenda", International Affairs, vol. 92, no. 2, 2016, pp. 373–392. For an overview of the field, see J. True and S.E. Davies (eds.), The Oxford Handbook of Women, Peace, and Security, 2018, https://doi.org/10.1093/oxfordhb/9780190638276.001.0001.

[112] United Nations, Department of Political and Peacebuilding Affairs, "Women, Peace and Security", https://dppa.un.org/en/women-peace-and-security.

[113] UN Women, "In Focus: Women, Peace, Power", n.d., https://www.unwomen.org/en/news/in-focus/women-peace-security.

[114] Ibid.; and R. Hessmann Dalaqua, K. Egeland and T. Graff Hugo, "Still Behind the Curve: Gender Balance in Arms Control, Non-proliferation and Disarmament Diplomacy", UNIDIR, 2019, https://doi.org/10.37559/WMD/19/gen2.

[115] C. de Jonge Oudraat et al., "Gender Mainstreaming: Indicators for the Implementation of UNSCR 1325 and Its Related Resolutions", NATO Science for Peace and Security Programme, 2015, https://www.nato.int/science/project-reports/UNSCR-1325-Scorecard-Final-Report.pdf.

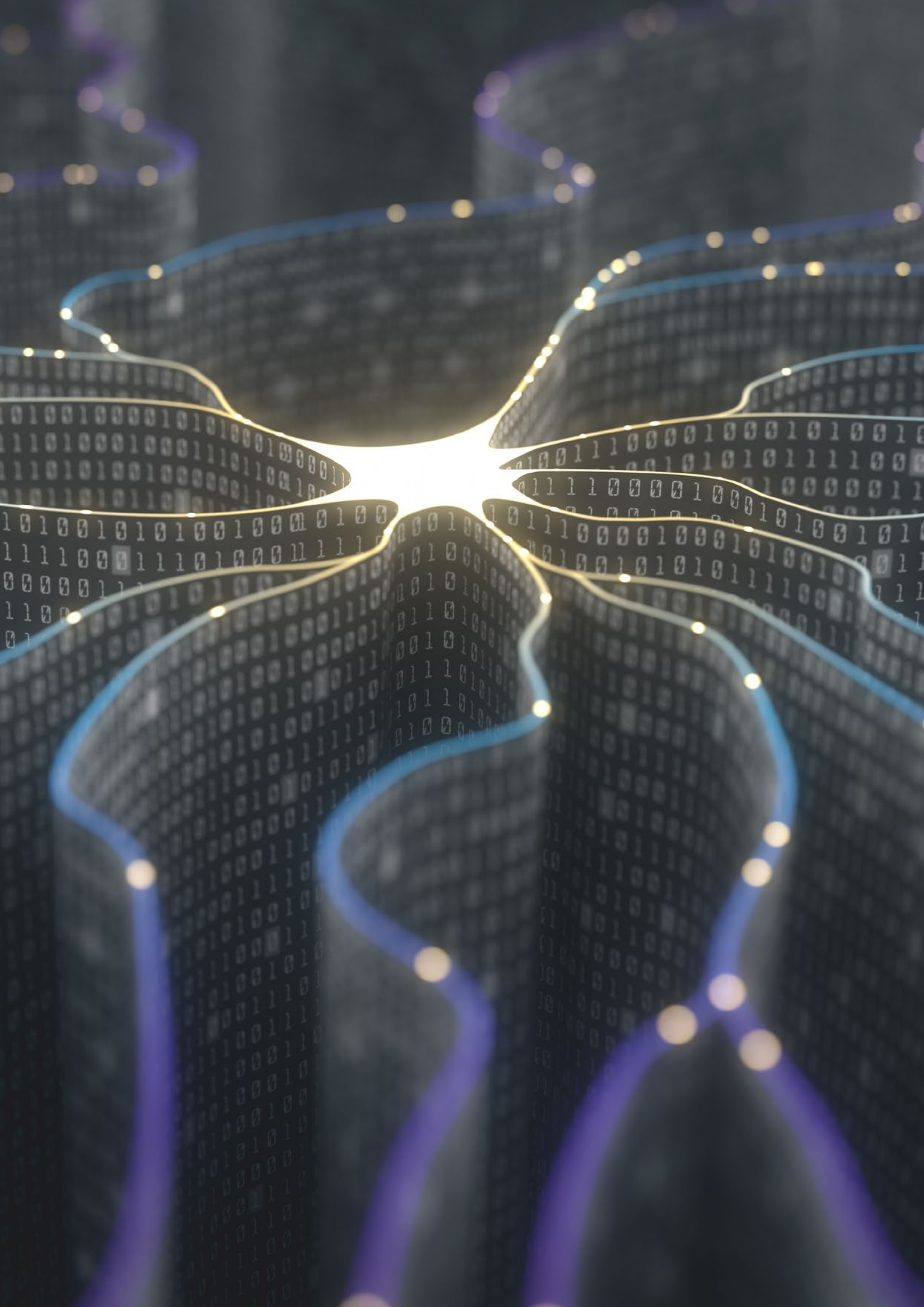
into national security strategies and policy directives, including military directives and guidance documents.[116] This process should also apply to military applications of AI.

A broad range of initiatives to address gender and the military should be connected to the development of military AI. They transect civil society, national government and international organizations. A gender-based audit of military applications of AI would make transparent the data-collection processes used in training a machine learning model, as well as the ways the algorithm has been tested and evaluated against key benchmarks tied to gender, race, age and ability. For these practices to have an impact, algorithms should be independently audited and evaluated, using testing scenarios that are distinct from the data sets used to train the machine learning model.

Members of the evaluation team should include not only computer scientists, lawyers and policymakers; rather, such teams should be broadly representative of the constituencies that will use and could be affected by the system. At minimum, reviews of AI for national defence should include experts who study gender. Some algorithms – particularly military applications of AI designed to identify people – should be evaluated by an international committee of experts who can diagnose culturally specific norms that are potentially standardized within such models.

A broader, gender-based review of military applications of AI might refer to the framework of women, peace and security to shape strategies for its use: How is women's participation and representation accounted for? Has conflict prevention been considered in the system's design? Does the model provide for the protection of human rights, and do these protections apply equally to men, women, boys and girls? For military applications of AI systems used for relief and recovery, particular attention should be given not only to the impact of algorithm-based models on women and children, but also to securing the data collected during these missions so that this information cannot be weaponized.

[116] Ibid.



5. Conclusions and recommendations

In the past year, international organizations and multilateral stakeholders have put forward a number of recommendations to limit autonomous weapon systems, artificial intelligence and the development of machine learning that could violate international humanitarian law and human rights law. This report affirms these limits. The International Committee of the Red Cross has recommended that the use of unpredictable autonomous weapon systems should be ruled out, as should autonomous target recognition against human objectives. Further, it has called for regulation of their design and use.[117] The Campaign to Stop Killer Robots has called for the prohibition of autonomous weapons without meaningful human control and of systems that use sensors to target human beings.[118] The United Nations Secretary-General, António Guterres, has called for a ban on autonomous weapons and the High Commissioner for Human Rights, Michelle Bachelet, has called for moratoriums on the sale and use of AI until adequate safeguards are in place to ensure compliance with human rights.[119] Others have stressed the importance of knowledge-sharing between States to establish best practices with regard to data associated with autonomous weapons.[120]

A ban on autonomous weapons will not be sufficient to address the myriad of other potential uses for military AI, however. The analysis presented in this report places gender and other social inequalities at the centre of military applications of AI. It advocates for a gender-based approach to military AI, which has the potential transform the direction of the field and shape future developments. The research proposes additional recommendations with these aims in mind, which tie military applications of AI to the goals of gender mainstreaming.

» **Apply gender mainstreaming policies based on UNSCR 1325 to military AI strategies.**

These would consider: How is women's participation and representation accounted for by military AI? Has conflict prevention been considered in the system's design? Does the model provide for the protection of human rights, and do these protections apply equally to women, men, girls, and boys? How would post-conflict relief and recovery missions be addressed by military AI?

In a similar vein, efforts to address gender-based violence in conflict should also address potential harms associated with military applications of artificial intelligence.

[117] International Committee of the Red Cross, "Autonomous Weapons: The ICRC Recommends Adopting New Rules", Statement, 3 August 2021, https://www.icrc.org/en/document/autonomous-weapons-icrc-recommends-new-rules.

[118] Campaign to Stop Killer Robots, "Statement to the Informal Discussions on Autonomous Weapon Systems", 29 June 2021, https://www.stopkillerrobots.org/wp-content/uploads/2021/09/CSKR-Statement-to-the-informal-discussions.docx.pdf.

[119] 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; and UN News, "Urgent Action Needed over Artificial Intelligence Risks to Human Rights", 15 September 2021. https://news.un.org/en/story/2021/09/1099972.

[120] A. Holland, Known Unknowns: Data Issues and Military Autonomous Systems, UNIDIR, 2021, https://doi.org/10.37559/SecTec/21/AI1.

» Develop specific measures to assess how military AI systems represent and respond to gender.

A gender-based review for military AI should make explicit how the system represents and responds to gender and how harmful effects have been mitigated. This information should be verified through an independent testing process and provided to personnel who have been trained to understand the potentials and pitfalls of the model they are using.

- Data sets should be documented, providing information to users regarding their motivation, composition, collection process, and recommended uses.
- Algorithmic models should be tested against benchmarks that evaluate their operation across gender, age, and race, as well as their intersections.
- Military AI systems should be evaluated using testing scenarios that are distinct from the data sets used to train the machine learning model.
- The review should extend to all uses of military AI not just systems that are being tested as weapons, as non-lethal uses including human resource management and logistics have gender-based impacts.
- Military applications of AI intended to be beneficial, especially in relief and recovery efforts, are particularly important to evaluate across gender, age, and race, as well as their intersections.
- Data associated with military AI could cause additional harms if it is compromised. Measures to protect data should account for these potential vulnerabilities and extend to the disposal of the system.
- The review process should include a diverse interdisciplinary team with expertise in gender studies. It should also incorporate feedback from persons who will use the system.

These proposals fit with ongoing efforts to address gender inequities in national security, international affairs and technology. By adopting principles that aim to limit harm in the machine learning life cycle, new technologies can be developed that promote – rather than hinder – gender equity and contribute to gender mainstreaming in the military, while regulations can be adopted to ensure that AI applications comply with international law.

Glossary

Algorithm	A procedure or set of rules used in calculation and problem-solving[121]
Artificial intelligence	The theory and development of computer systems to replicate human processes[122]
Autonomous weapons	Weapons that, once activated, can identify and select targets and apply force to them without human intervention[123]
Bias	The tendency to prejudice in favour of or against one thing, person or group compared with another, usually in a way considered to be unfair[124]
Cognition	The mental action or process of acquiring knowledge and understanding through thought, experience and the senses[125]
Computer processing	The procedure that transforms data into meaningful information through hardware components[126]
Data	(1) Facts and statistics collected together for reference or analysis; (2) quantities, characters or symbols on which operations are performed by a computer[127]
Gender	The roles, behaviours, activities and attributes that a given society at a given time considers appropriate or as a "norm" for women and men and girls and boys, as well as non-binary or gender-fluid people[128]
Harm	The withholding of rights, opportunities or resources from certain people or groups, or the perpetuation of stigma and stereotypes associated with certain people or groups[129]
Intersectionality	A framework for understanding how the interconnected aspects of a person's social categorizations such as race, gender and class create different modes of disadvantage and privilege[130]
International Humanitarian Law	The body of international law that seeks to limit the effects of armed conflict and protect the rights of people who are not or are no longer participating in hostilities and restricts the means and methods of warfare[131]
International Human Rights Law	The body of international law, established by treaty or custom, on the basis of which individuals and groups can expect and/or claim certain rights that must be respected and protected by their states[132]
Machine Learning	A branch of artificial intelligence that focuses on using data and algorithms to imitate how humans learn[133]
Military AI	Applications of artificial intelligence for national security[134]

[121] “algorithm, n.”, Oxford Dictionaries, http://en.oxforddictionaries.com/definition/algorithm.

[122] “artificial intelligence, n.”, Oxford Dictionaries, http://en.oxforddictionaries.com/definition/artificial_intelligence.

[123] V. Boulanin, N. Goussac and L. Bruun, Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human–Machine Interaction, SIPRI, June 2021, https://www.sipri.org/sites/default/files/2021-06/2106_aws_and_ihl_0.pdf.

[124] “bias, n.”, Oxford Dictionaries, http://en.oxforddictionaries.com/definition/bias.

[125] “cognition, n.”, Oxford Dictionaries, http://en.oxforddictionaries.com/definition/cognition.

[126] “data processing”, Encyclopaedia Britannica, https://www.britannica.com/technology/data-processing.

[127] “data, n.”, Oxford Dictionaries, http://en.oxforddictionaries.com/definition/data.

[128] UNIDIR, “What is Gender?”, n.d., https://unidir.org/gender-perspective.

[129] H. Suresh and J. Guttag, “Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle”, MIT Case Studies in Social and Ethical Responsibilities of Computing, summer 2021, https://doi.org/10.21428/2c-646de5.c16a07bb.

[130] UN Women, “Intersectional Feminism: What It Means and Why It Matters”, 1 July 2020, https://www.unwomen.org/en/news/stories/2020/6/explainer-intersectional-feminism-what-it-means-and-why-it-matters.

[131] International Committee of the Red Cross, “War & Law”, https://icrc.org/en/war-and-law.

[132] International Committee of the Red Cross, “What Is the Difference Between IHL and human rights law?”, https://www.icrc.org/en/document/what-difference-between-ihl-and-human-rights-law.

[133] IBM Cloud Education, “Machine Learning”, IBM, 15 July 2020, https://www.ibm.com/cloud/learn/machine-learning.

[134] D. Chenok, L. van Bochoven and D. Zaharchuk, “Deploying AI in Defense Organizations”, IBM Center for the Business of Government, 2021, https://www.ibm.com/downloads/cas/EJBREOMX.

Does Military AI Have Gender?

Understanding bias and promoting ethical approaches in military applications of AI

Does Military AI Have Gender? uncovers the significance of gender norms in the development and deployment of artificial intelligence (AI) for military purposes. The report addresses gender bias in data collection, algorithms and computer processing.

Drawing on research in ethical AI, the report outlines avenues for countering bias and mitigating harm, including a gender-based review of military applications of AI. In doing so, it seeks to chart a path for technology development that promotes – rather than hinders – gender equity and contributes to gender mainstreaming in the military.

UNIDIR UNITED NATIONS INSTITUTE FOR DISARMAMENT RESEARCH

@unidirgeneva @UNIDIR un_disarmresearch