Digital War
https://doi.org/10.1057/s42984-024-00094-z

ORIGINAL ARTICLE

Putting algorithmic bias on top of the agenda in the discussions on autonomous weapons systems

Ishmael Bhila[1]

Abstract
Biases in artificial intelligence have been flagged in academic and policy literature for years. Autonomous weapons systems—defined as weapons that use sensors and algorithms to select, track, target, and engage targets without human intervention—have the potential to mirror systems of societal inequality which reproduce algorithmic bias. This article argues that the problem of engrained algorithmic bias poses a greater challenge to autonomous weapons systems developers than most other risks discussed in the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE on LAWS), which should be reflected in the outcome documents of these discussions. This is mainly because it takes longer to rectify a discriminatory algorithm than it does to issue an apology for a mistake that occurs occasionally. Highly militarised states have controlled both the discussions and their outcomes, which have focused on issues that are pertinent to them while ignoring what is existential for the rest of the world. Various calls from civil society, researchers, and smaller states for a legally binding instrument to regulate the development and use of autonomous weapons systems have always included the call for recognising algorithmic bias in autonomous weapons, which has not been reflected in discussion outcomes. This paper argues that any ethical framework developed for the regulation of autonomous weapons systems should, in detail, ensure that the development and use of autonomous weapons systems do not prejudice against vulnerable sections of (global) society.

Keywords Autonomous weapons systems · Algorithmic bias · Automation bias · Inequality · Roboethics

Introduction

On the 17 May 2023, the Chair of the Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS), Ambassador Flavio Soares Damico opened the morning session of the GGE meetings. On the agenda was the discussion of paragraph 23–30 of the 2023 Report of the GGE, dealing broadly with issues of human–machine interaction in relation to autonomous weapons systems and how these could be regulated. Paragraph 27 of the draft report made passing reference to automation bias and "unintended bias". Noting that algorithmic bias would affect people of colour, minorities, and other vulnerable populations, the Philippine delegate pointed out that the report needed to make "a clearer reference to the need to spell out the risks arising from possible racial and gender bias".[1] In the same manner, the Canadian delegation noted that the language used in the making of the Chair's report would have to "expand on the concept of unintended biases… to include the language such as ethnicity, gender, age, and disability".[2] Costa Rica, Panama, and Mexico buttressed the same point, with Mexico going further to suggest that the outcome document should include measures to *prevent*- not *mitigate*- algorithmic biases that come with AI.[3]

Despite these calls for clear language on the prevention of algorithmic bias in autonomous weapons systems, the draft report that was produced on the next day omitted issues of race, and the final report did not include any of the suggested strong language, instead encouraging measures to "reduce automation bias in system operators" and "reduce unintended bias in artificial intelligence capabilities related to the use of the weapon system".[4] The report ignored all the calls for the recognition of such a central problem in the use of AI, particularly autonomous weapons systems which may disproportionately impact vulnerable populations. The following sections of this paper will show how AI systems have disproportionately affected vulnerable populations,

✉ Ishmael Bhila
ishmael.bhila@port.ac.uk

[1] School of Law, University of Portsmouth, Portsmouth, UK

Published online: 31 May 2024

making a case for a closer consideration of such problems when discussing autonomy in weapons.

This paper, based on postcolonial critique of the sociotechnologies of war, contributes to the emerging discussion on inequality and bias in autonomous weapons systems. From the onset, it should be noted that the paper does not address challenges with autonomous weapons systems that target military objects; the paper is concerned with the development of autonomous weapons systems that identify, track, select, target, and engage human targets. While there is extensive literature on bias in AI, the same level of scrutiny is yet to be applied to autonomous weapons systems. A few scholars have addressed the problem of the potential risk of bias posed by autonomy in weapons. Figueroa and others address algorithmic bias against persons with disabilities and the silence of that discussion in international discussions on autonomous weapons systems (Figueroa et al. 2023). Shama Ams' paper deals with the convergence of military and civilian uses of AI and addresses algorithmic bias in passing (Ams 2023), and Catherine Jones' paper focuses on Western-centric research methods, with automation bias in lethal autonomous weapons used only as an example (Jones 2021). While there have been increased scholarly debates on military applications of AI, the analysis of how the most significant forum to discuss the potential regulation of autonomous weapons systems has accounted for algorithmic bias has not been examined so far. This article therefore makes a key empirical contribution to discourse relating to the global governance of military applications of AI.

The problem of engrained algorithmic bias poses a greater challenge to the justifications for the use of autonomous weapons systems by their developers than the risks of proliferation, incidental loss of life, access by terrorists, and other identified risks. This is mainly because it takes longer to rectify a discriminatory algorithm as seen in the many examples given in the following sections of this paper than it is to issue an apology for a mistake that occurs occasionally. Powerful states have controlled both the discussions and the outcome, which has focused on issues that are pertinent to them while ignoring what is existential for the rest of the world. This paper unpacks these dynamics, making a case for centring the issue of algorithmic bias in outcome documents to reflect the discussions that take place within the GGE on LAWS discussions.

Based on multidisciplinary literature from science and technology studies (STS), engineering, computer science, social science, and other fields, Section 1 draws attention to the emergence of biases in deep learning processes, natural language processing (NLP), machine learning, and system training and how these can have a profound impact on the development and use of autonomous weapons systems. In understanding these biases, the paper shows how these shortcomings can be transferred to autonomous weapons systems, and how the risk of bias escalates in new contexts from the system's environment of development, particularly in different geographies and communities in the Global South.

Having shown the biases in AI and autonomous weapons systems development, Section 2 goes on to argue that international discussions on autonomous weapons systems should give centrality to the problem of algorithmic bias as it would affect most of the global population if not properly addressed. Section 3 argues that both procedural and substantive international law should contain strong language that can achieve, to use the Mexican delegation's terms, the prevention rather than the reduction of algorithmic bias in autonomous weapons. While procedural law deals with the rules, processes, and procedures of how international lawmaking practices are conducted, substantive law seeks to address inequalities, enhance the voice of the marginalised, eradicate prejudices, acknowledge differences, and accomplish structural change (Fredman 2016). The paper concludes by showing how language in the CCW process has avoided adequately addressing a clear problem and makes a case for a more sensitive approach that does not perpetuate existing inequalities.

Methodology

This paper is a result of ongoing PhD research on the participation of small states in the making of international law relating to autonomous weapons systems. Based on the postcolonial technoscientific framework, the paper critiques the situated knowledge that has marginalised issues that are pertinent to the discourse and practice of algorithmic warfare and to those who are most likely to be affected by them. The paper adopts a qualitative research methodology, analysing state submissions/proposals in the United Nations Convention on Certain Conventional Weapons (CCW) since the year 2017 with the start of formal discussions through the establishment of the Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS). State submissions are proposals made to the group to guide discussion and to suggest what a normative and practical outcome on the subject should look like. For this paper, I analysed states' treatment of the problem of algorithmic bias in their submissions and the differences in state interests towards mitigating the issue. A total of 73 working papers, submissions and other proposals were analysed covering the period 2017–2023.

The methodology also relied heavily on statements made by states in the CCW GGE on LAWS meetings that have been taking place on average twice each year since 2017. These statements are in the form of legal debates guided by the Chair of the meeting who sets the questions and agenda, usually culminating in a chair's report at the end

of each session. I sought to highlight when and how states voiced concerns about algorithmic bias, the reaction to those concerns, and an analysis of the outcomes of those interventions.

Finally, the research analysed whether concerns and suggestions about algorithmic bias were incorporated in the Chair's reports, and if they did, how these suggestions were included. In this sense, I looked not only at the idea of the inclusion of the concept but also the quality and importance that it was awarded in relation to other issues. The statements, submissions, Chairs' agendas and reports, and related documents are publicly available on the United Nations Office at Geneva databases.[5]

The nature and forms of automation and algorithmic bias

The sociotechnologies of security (or insecurity) are open to failure, owing mainly to their "inherent contradictions and irremediable fault lines" (Suchman et al. 2017). Autonomous weapons systems are based on sensors, AI, and other emerging technologies for profiling, biometrics, thermal imaging, data mining, satellite observation, and population metrics; the use of which is based on hierarchies of knowledges, assumptions, vocabularies, and modes of attention (Wilke 2017). As autonomous weapon systems are expected to identify, monitor, and engage targets, their ability to tell significant facts about human life, particularly in contexts foreign to their conditions of design and development, is highly overestimated (Adelman 2018).

This section considers several areas which characterise autonomous weapons systems and how these are liable to racialisation, discrimination, and bias. The paper acknowledges the positive aspects of AI. However, the purpose of this study is to analyse algorithmic bias and its potential impact in the development and use of autonomous weapons systems. The positive aspects of AI both in civilian and military spaces are well documented and are still being realised. This paper also focuses on the algorithms that animate autonomous and AI technologies despite autonomous weapons systems being not always based on AI technologies. The paper considers autonomy in weapons as a spectrum with the potential for having challenges at any level, not as a fixed system based solely on one type of technology. This elusive nature of autonomy in weapons makes it essential to have robust regulatory frameworks before they are deployed. This contribution seeks to add to the conversation on a comprehensive regulatory regime for AI in the military domain, focusing only on the pressing issue of algorithmic bias as it pertains to autonomous weapons systems. We ought to learn from what is already known about the problems of AI. The argument is not about banning the development of AI, and the paper does not engage on the debate on whether autonomous weapon systems are legally or ethically permissible in international law and in practice, it simply aims to attract more attention to the problem of algorithmic bias when discussions on autonomous weapons systems are done.

Autonomous weapons, like most AI-based security systems, engage in data collection, storage, and management to enable the conduction of intelligence, surveillance, and reconnaissance (ISR). For example, uncrewed aerial vehicles (UAVs) can collect information about the profiles and nature of targets, improve their functionality without human oversight, and can be programmed to include numerous responses to respective challenges (Konert and Balcerzak 2021). This data collection, storage, and management has the potential to lead to racialisation through the biased creation and utilisation of data (M'charek et al. 2014). In 2021, the USA government published that multiple civilians had been killed through "targeted killing" using drones, and Peter Lee gave the example of Afghan civilians who were killed having been misidentified as terrorists, noting that these new weapons are used deliberately to coerce populations (Lee 2021), echoing Judith Butler's argument that those targeted are viewed as people whose lives are injurable and lose-able (Butler 2009). Algorithms are only as good as the data they are fed, which means that who creates them and where they are used matter the most. The culture, beliefs, and value system of the developer are influential in how the algorithms will perform in settings that are different from where they were programmed. Algorithmic bias is classified into three categories: preexisting bias- which is influenced by unequal social structures and culture, technical bias—that emanates from technical shortcomings, and emergent bias—which is a result of a change in environment or context within which the algorithm is used (Friedman and Nissenbaum 1996). I propose that efforts to regulate autonomous weapons systems should consider bias at all these levels to avoid unintended harms against marginalised and vulnerable populations.

It is essential to consider the risk posed by targeting humans using sensors in war as this has serious ethical and bias implications. The USA, for example, uses electronic and visual data collected through sensors to gather intelligence in its "global war on terror" (US Office of the Secretary of Defense 2007). Roboticists like Ronald Arkin proposed the use of robots (autonomous weapons systems) in war that could be emotionless and that utilise electro-optics, robotic sensors, and synthetic aperture to observe and target humans (Arkin 2010). However, several scholars have criticised this uncritical trust in the use of automated sensors and AI in targeting who to kill. Critics of those who are pro-autonomous weapons have argued that targeting using algorithms is murky when it comes to distinguishing between civilians and combatants, especially

in unfamiliar cultural contexts (Sharkey 2010). Others questioned the ability of autonomous weapons to identify legitimate targets and to make strategic and tactical decisions (Roff 2014; Johnson 2022; Hunter and Bowen 2023).

This takes us to the question of the "target of colour." AI systems can learn from conversations, observation, and identification of patterns, all of which are liable to systemic bias (Klugman 2021). With the proliferation of the use of large language models (LLMs) in many domains, companies like Palantir developed an artificial intelligence platform (AIP) for defence to "unleash the power of LLMs and cutting-edge AI for defence".[6] Palantir has worked with the USA and UK governments for the provision of military and security data and surveillance services, and it has expanded its market to European security services, accompanied by controversies of data privacy concerns (Johnston and Pitel 2023). Data used to train machine learning models can reproduce inequalities and be incomplete, leading to biased outcomes (Ferrara 2023). Military applications of facial recognition are already in implementation with the Ukrainian battlefield having already integrated Clearview AI's facial recognition software to identify enemies (Dave and Dastin 2022). The gathering of biometric data using AI is highly faulty among people of colour, with one large-scale study showing that AI largely misidentified people from East and West Africa and East Asian people migrating to the US, while algorithms made in China were effective at identifying East Asian people while misidentifying American Indians, African Americans, and other Asian populations (Grother et al. 2019). Buolamwini and Gebru discovered that machine learning algorithms are more likely to discriminate and misclassify darker-skinned females (at a rate of 34.7%) as compared to light-skinned males (0.8%) (Buolamwini and Gebru 2018). This is mainly a result of sampling bias whereby an algorithm is trained to recognise a certain section of society (Ferrara 2023) and autonomous weapons systems are largely developed and trained in the USA, China, states in the European Union, and a few other leaders in AI development, which leaves populations in parts of the world where majorities are not white at a very high risk. Google's Google Photo algorithm was recorded to have misidentified a black couple as gorillas and still could not find a viable solution to their biased algorithm after years (Vincent 2018; Grant and Hill 2023). If facial recognition software is used by states to identify security threats (Israel HLS & CYBER 2022), the risk of killing the wrong people grows extremely high if autonomous weapons systems are used among racially different populations with the high probability in AI of misidentifying people of colour. Software like Faception (FACEPTION|Facial Personality Analytics, 2023) that claim to be able to identify a terrorist or paedophile through facial recognition are highly controversial but have been used by governments (Buolamwini and Gebru 2018).

In addition to the biased collection and usage of biometric data, AI also has the capability to learn directly through voice recognition and interaction with humans (Kim et al. 2019; Klugman 2021). Speech recognition AI utilises, interprets, and employs language in ways that are not anticipated by humans (Bylieva 2022). Robots like Ameca use generative AI to speak several languages and to interact directly with people (Chan 2023), and in July 2023 at the United Nations' AI for good conference in Geneva a group of nine robots held a press conference where they addressed questions from humans (Ferguson 2023). However, the humans addressing the robots at the AI for good conference were told to speak slowly and there were obvious inconsistencies and poses between responses. Speech recognition AI systems struggle when interacting with unfamiliar speech patterns. A study on the use of Apple's iOS Siri system showed that it had challenges understanding children's speech, owing to issues like pitch and patterns of voice and speech, and the types of questions children ask (Lovato and Piper 2015). A study of five automated speech recognition (ASR) systems in the USA discovered that the average word error rate for transcribing speech by black speakers was much higher than it was for white speakers (Koenecke et al. 2020). Another study in Britain noted that automatic speech recognition reproduced and perpetuated existing linguistic discrimination against marginalised groups (Markl 2023). If this type of AI is used in autonomous weapons systems, such margins for error can have catastrophic consequences. For example, a robot may be tasked to "select and engage" a target based on its own understanding of who is or is not a belligerent in a community with a foreign language. If such a system finds a group of young men in an African community, for example, wresting and insulting each other and shouting at the top of their voices while enjoying themselves, the chances of it profiling them as combatants is extremely high, simply because it lacks an understanding of the culture and language patterns.

Closely related to the problem of voice recognition is the problem of bias in translation. With autonomous weapons systems largely developed in the West and by a few more countries in the world, the chances of some of these systems having to rely on translation are very high. A good example of bias in machine translation is the gender bias in Google translation. Many languages are gender neutral or have gender-based words, which makes translation to the English language highly inaccurate even with the most modern AI systems. The cultural differences between the source language and target language can lead to gender bias in translation of several languages. For example, the translation of *dia seorang dokter,* which is gender neutral, from Indonesian by Google Translate translates to *he is a doctor* while *dia*

seroang perawat which is also gender neutral translates to *she is a nurse* (Fitria 2021). This is the challenge of stereotyping in machine translation (Savoldi et al. 2021). The translator automatically assumes that a doctor should be male while a nurse should be female. The study that showed this bias was done in 2021, and at the time of this study in 2023, the same bias was unchanged in Google Translate. In the event that such a bias is contained in an autonomous weapon system, there could be biased identification of targets for a long time before a technical error is corrected, which would be catastrophic, unethical, and tragic. In addition, it would be tragic to discover a problem of algorithmic bias through experience as opposed to discovery by research especially in the domain of military technology which has implications over life and death. With the existence of multiple languages in vulnerable societies, machine translation systems also carry the bias of under-representation where certain groups are not even visible (Savoldi et al. 2021). For example, an AI system that relies on natural language learning and translation would easily misidentify people in multiple ethnic settings in Zimbabwe. A simple search on Google Translate of the word *mukororo,* a traditional Ndau word that means *son* automatically mistranslates to Shona, the dominant language in Zimbabwe, the result of which is *thief*. The correct Shona word for thief is not mukororo but *gororo*. In such a case, an AI system can easily misidentify someone as a thief who is simply being endearingly being referred to as a son. In 2017, Facebook translated the phrase "good morning" from Arabic into "attack them" in Hebrew which led to the arrest of a Palestinian man by Israeli police (Hern 2017). A study of hate speech detection tools showed that members of minority groups, especially Black people, in America were likely to be labelled as offensive by hate speech detection identification tools because of their dialect, also exposing them to real-life violence (Sap et al. 2019). These cases show that translation AI has already been proven as faulty in many cases, and any military AI that would be based on such systems is likely to be ethically questionable.

The USA has partnered with Scale AI to develop "Scale Donovan", an AI platform that uses LLMs based on the same faulty philosophy that led to the killing of civilians at a wedding in Mali- relying on AI for the identification of who is "friendly" or an enemy through live data and depending on AI's ISR information. Such a catastrophic mistake was made with a "human in the loop" which makes it plausible to assume that worse can happen if autonomy in weapons does not account for bias. Everyday language used in social settings is complex, which makes it risky to deploy harmful technologies that cannot reason beyond colloquialisms (for example, the statement "an all-Muslim movie was a 'box office bomb'" would easily be interpretated as stereotypical by most people, assuming that all Muslims are terrorists- a bias that cannot be easily explained and understood by an AI system) (Sap et al. 2020). Large language models reveal a spectrum of behaviours that are harmful, especially through the reinforcement of social biases (Ganguli et al. 2022). Algorithmic bias in AI systems can lead to the reinforcement and escalation of social inequalities and biased decisions (Kordzadeh and Ghasemaghaei 2022), which would lead to the application of force on the wrong targets by emerging technologies in the area of autonomous weapons systems.

The identification of what is perceived as hostile by AI can also be very problematic. If autonomous weapons systems and emerging technology-based systems, select and target threats. The global war on terror led by the USA and its allies depends largely on ISR done by semi-autonomous or autonomous systems, a practice that is controversial and has led to the killing of multiple civilians in environments foreign to those of the deployers. In Mali, the French army killed multiple people at a wedding after a Reaper Drone provided wrong ISR information, mistaking wedding attendees for insurgents (Stoke White Investigations 2021). These challenges should be addressed, and regulations should be put in place before autonomous weapons systems are deployed.

Recognition of algorithmic bias in global policy and national legal contexts

The problem of exclusion, termed as representational harm by Kate Crawford, is a widely recognised challenge in AI debates (Ruttkamp-Bloem 2023). Various studies have acknowledged that there is a crisis with regards to diversity in AI (West et al. 2019). The challenges of bias in AI have been flagged in recent years in soft law (recommendations, guidelines, standards, codes of conduct, and other non-binding laws), and the development of hard law on AI is still in its infancy (Gutierrez 2023). The European Union's 2021 proposal for an Artificial Intelligence Act (Article 33 and 37) proposed the regulation of AI systems that use "'real-time' and 'post' remote biometric identification systems" that have the risk of bias and discrimination according to sex, ethnicity, age, or disability based on historical societal patterns (European Commission 2021). The EU AI Act proposes a risk-based approach, and issues of discrimination and bias in AI are classified as "high risk."

In the same manner, UNESCO's recommendations on AI ethics encouraged its member states to be wary of the cultural impacts of AI, noting that natural language processing should be cognisant of the "nuances of human language and expression" (UNESCO 2022). The Council of Europe's Committee on Artificial Intelligence (CAI) went even further in its Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law to

suggest that states should manage risk through ensuring that those who may be affected by AI should have their perspectives heard when risk and impact assessments are done.[7] However, this provision in the draft convention would have been useful in addressing the effects of algorithmic bias, particularly in more risky technologies like autonomous weapons systems, but it falls agonisingly short in tackling the problems of discrimination and bias in AI. Such an approach of including the potential victims in the development of normative frameworks is what this paper advocates, especially in negotiations for the regulation of the development and use of autonomous weapons systems at the UN.

The Organisation for Economic Cooperation and Development (OECD), which has 38 member states, developed their own recommendations on AI that were endorsed by a large global state population, recognising in the first instance that "a well-informed whole-of-society public debate is necessary for … limiting the risks associated with" AI (OECD 2022). The recommendations by the OECD are based on "human-centred values and fairness" that include equality, non-discrimination, the inclusion of underrepresented populations, diversity, fairness, and social justice, with the target goal of reducing inequalities and addressing bias.[8] Some comprehensive studies on AI policies, for example Maas and Villalobos' work, have identified some seven "institutional models" for AI governance (scientific consensus building, political consensus building and norm-setting, coordination of policy and regulation, enforcement of standards or restrictions, stabilisation and emergency response, international joint research, and distribution of benefits and access) but have barely analysed the issue of algorithmic in those models (Maas and Villalobos 2023). Scholars working on global regulation of AI have acknowledged algorithmic bias but have chosen not to focus "on the relative urgency of existing algorithmic threats (such as e.g., facial recognition or algorithmic bias)" but to find ways in which those looking to regulate could find convergencies for ethical AI (Stix and Maas 2021, p. 261). This underlines the urgency for the recognition of algorithmic bias in AI not only in practice but also in global regulation efforts.

The number of global and domestic legislations that aim to mitigate the risks of AI has increased by more than six times since 2016 (Maslej et al. 2023). In the USA, the National Institute of Standards and Technology noted in its Artificial Intelligence Risk Management Framework that "AI systems are inherently socio-technical in nature, meaning that they are influenced by societal dynamics and human behaviour" (NIST AI 100-1, 2021). The framework goes on to identify potential harms, including harm to groups or communities through discrimination. An interesting section in the framework addresses "risk prioritisation". In this section, the policy argues that sometimes, there are risks that are not worth prioritising, especially if they cannot be fully eliminated. This, however, calls to question how risks are defined by different people. It is highly questionable whether a developer who is not likely to be affected by racial bias in weapons systems would regard it as a priority. It is therefore essential to have a framework that decides what is essential and what is not based on equal consultation and participation rather than business-informed decisions by developers.

Despite the mentions and references to bias in AI in soft law in global context, the issue is yet to be fully addressed, with no policy fully devoting space to the risks of algorithmic bias and how it should be prevented. With governments not committed to developing binding regulations for AI to maximise its benefits (Marchant et al. 2020), it is hard to see challenges that are socially embedded like algorithmic bias being given the attention and urgency they deserve.

The CCW negotiations and algorithmic bias

The Heyns Report (A/HRC/23/47) of 2013[9] introduced the issue of "lethal autonomous robotics" (LARs). In that report, there was mention of the respect for human life, the Martens Clause, and other challenges posed by autonomous weapons systems, but the issue of race and bias was yet to be introduced in the discussion. However, as the depth in discussion developed over the years, states began to recognise the importance of taking biases in AI seriously when thinking about autonomous weapons systems. At the first formal Group of Governmental Experts (GGE) meeting in 2017, the USA submitted a proposal arguing that prohibitions should be directed towards "intentional wrongdoing," with unintended consequences referred to as "mere accidents or equipment malfunctions" that do not violate the law of war.[10] This logic would mean that there would be no responsibility for systems that are "unintentionally" racist that would disproportionally affect vulnerable groups. In a 2018 CCW submission, the International Committee of the Red Cross (ICRC) made a passing note that "unpredictable and unreliable operations may result from a variety of factors, including … in-built algorithmic bias".[11] In 2019, the Chair's report noted that there was need for further clarification on aspects like "possible bias in the datasets used in algorithm-based programming relevant to emerging technologies in the area of autonomous weapons systems".[12] Thompson Chengeta observed the same challenge and explained that.

> "an earlier version of the 2019 GGE Report included a paragraph that noted that the use of AADs may compound or worsen social injustices such as racial and gender discrimination. During the discussions, no state representative contested that paragraph. Later in the evening of the same day when another version of the report was provided, the paragraph had been removed.

> The delegations from South African and Canada questioned why this had occurred, but no remedy was provided and the text addressing discrimination risks remained excluded" (Chengeta 2020, p. 177).

Regardless of these calls, in 2020 only passing references were made to the risks of "unintended engagements" posed by autonomous weapons systems.[13] The GGE took one step forward and backtracked twice on the issue despite the hope of the equalisation represented by UN organs. With small states being the ones facing potential effects of biases in autonomous weapons systems, the debates developed in a binary manner that continued to either ignore or silence calls for the recognition and discussion of algorithmic bias.

Several states have raised the issue of algorithmic bias since 2020. In 2021, the Holy See argued that "autonomous weapons systems, equipped with self-learning or self-programmable capabilities, necessarily give way to a certain level of unpredictability, which could, [lead to] such systems [making] mistakes in identifying the intended targets due to some unidentified "bias" induced by their 'self-learning capabilities'".[14] A joint working paper in the same year by Argentina, Costa Rica, Ecuador, El Salvador, Panama, Palestine, Peru, the Philippines, Sierra Leone, and Uruguay noted that "weapon systems are not neutral. Algorithm-based programming relies on datasets that can perpetuate or amplify social biases, including gender and racial bias, and thus have implications for compliance with international law".[15] Additionally, Argentina, Ecuador, Costa Rica, Nigeria, Panama, the Philippines, Sierra Leone, and Uruguay submitted a Draft which they labelled as 'Protocol VI' whose Article 3 Section 3 suggested that "each High Contracting Party shall ensure that weapon systems do not rely on datasets that can perpetuate or amplify social biases, including gender and racial bias." In 2023, Pakistan submitted that "there are already known problems of data bias and unpredictability that are compounded by growing autonomy of these weapons, based on machine learning algorithms".[16] A 2023 paper by a group of nine European and Latin American states noted that a normative framework should be developed that considers "the avoidance of data bias and programming shortfalls in complex systems".[17] A March 2023 proposal by Palestine also argued that the process of using encoded data to target, select, and engage humans with force would "likely entrench bias and discrimination through flawed profiling of human characteristics, particularly if seeking to target some people rather than others".[18] All these concerns by various states are testament of the centrality of the problem of algorithmic bias in negotiations for a normative framework for autonomous weapons systems.

Interestingly, however, the risks posed by autonomous weapons systems that have largely been considered by highly militarised states like the USA, UK, Russia, Australia, and others include unintended engagements, civilian casualties, incidental loss of life, the risk of proliferation, loss of control of the system, and the risk of acquisition by terrorist groups.[19] Issues of racial, ethnic, and gender bias in autonomous weapons systems are omitted in almost all their submissions, whether deliberately or unconsciously. The absence of racial and other biases in the discourse used by these powerful states in the CCW has also relegated the issue of algorithmic bias to the periphery of the outcomes of the discussions in the GGE. Western philosophy of science, which informs such discourse, has marginalised such pertinent concerns like algorithmic bias to the periphery. The problem of engrained algorithmic bias poses a greater challenge to the justifications for the use of autonomous weapons systems by their developers than the risks of proliferation, incidental loss of life, access by terrorists, and other identified risks. This is mainly because it takes longer to rectify a discriminatory algorithm- as seen in Google's failure to fix its racist and sexist translations for years even until now- than it is to issue an apology for a mistake that occurs occasionally. These powerful states have controlled both the discussions and the outcome, which has focused on issues that are pertinent to them while ignoring what is existential for the rest of the world.

The Convention on Certain Conventional Weapons (CCW), on paper, represents actors from across the global divide, with Civil Society, Think Tanks, States from all regions, and regional and international organisations represented. However, in practice, this representationalism is neither existent nor desired by some of the actors. At the time of writing, the CCW had 126 state parties, four of which were signatories. Adopted in 1980, the convention seeks to ban or restrict the development and/or use of certain types of weapons that may cause unnecessary harm in war or that may have an indiscriminate impact on civilians. The CCW is uniquely positioned to address issues of emerging weapons, with Article 8 (2)(a) stating that high contracting parties can suggest new protocols not already covered to be added (Convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects 1980). However, for military superpowers who want to maintain military superiority through weapons based on emerging technologies, the introduction of new prohibitions is not an attractive prospect, which has hampered the effectiveness of the CCW (Carvin 2017). To this end, the USA in its 2018 working paper argued that states should not seek "to codify best practices or set new international standards for human–machine interaction in this area" as it was impractical, favouring instead voluntary measures by states to comply to IHL.[20] For the majority of the world, however, whose security is not guaranteed and whose vulnerabilities are many, international law provides the best option for security.

The biases that are synonymous with emerging technologies and weapons are bound to affect smaller states, fragile communities, minorities, vulnerable populations, and people of colour more than they do the dominant states.

Within this context, the CCW has failed to be inclusive and equal. Practice in international law has shown that the existence of diverse perspectives in disarmament discussions is of utmost importance for the success of multilateral decision-making (Borrie and Thornton 2008). Scholars like Thompson Chengeta have argued that the CCW is not the correct forum for discussing autonomous weapons systems (Chengeta 2022). This research shows that the CCW falls short in several ways in encouraging inclusion and equality.

One explanation that can be offered for the CCW's failure to address issues concerning algorithmic bias in the debates on autonomous weapons systems is that most vulnerable states are almost always excluded, particularly due to structural constraints. African and Caribbean states are scarcely represented in the discussions on autonomous weapons systems within the CCW. These are the states that are predominantly black in their racial composition. Of the 24 states in the Caribbean, only Antigua & Barbuda, Cuba, Dominican Republic, Grenada, and Jamaica are parties to the CCW. Less than half of African states (26) are either High Contracting Parties or Signatories to the CCW. A staggering 65 UN member states are not parties to the CCW, with only Andorra being European. This means that if all the 126 states parties to the CCW were to attend and contribute to the discussions on autonomous weapons systems, a disproportionately high number of vulnerable states are left out from the onset.

In addition, many small states who are part of the CCW do not have the capacity to be in the discussions on a yearly basis. During fieldwork for this study, I realised that during the 15–19 May 2023 session of the GGE, very few African and Caribbean states were represented. On the 15th, the first meeting did not have a single Black-African state represented, and throughout the whole session, only South Africa, Nigeria, Algeria, Sierra Leone, and Cameroon were represented among African states. Among those present, only Algeria (Monday 15 May) and South Africa (Friday 19 May) made very brief statements. For Caribbean states, Cuba- which has always been present at GGE meetings on autonomous weapons systems made several contributions to the discussions. The research showed that most smaller states simply cannot afford to provide and fund personnel to attend these meetings, even if they are part of the CCW. This reflects on the structural inadequacies of the CCW as a forum for international law-making. Caribbean and Latin American states who have not been actively involved in the CCW attended the February 2023 Latin American and Caribbean Conference on the Social and Humanitarian Impact of Autonomous Weapons organised by Costa Rica,[21] showing their willingness to discuss and address challenges posed by autonomous weapons systems. These states came up with the Belén Communique which reiterated their commitment to actively engage in the debates to push for a legally binding instrument on autonomous weapons systems.[22] In addition, Caribbean states also convened a conference on autonomous weapons systems in September 2023, coming up with the CARICOM declaration which emphasised the need for regulating autonomous weapons systems so that they "should not be leveraged to undermine human rights, exacerbate prevailing inequalities, nor deepen discrimination on the basis of race, ethnicity, nationality, class, religion, gender, age, or other status."[23] Similarly, in December 2023, the Philippines organised a conference on autonomous weapons systems, bringing an Asian perspective to the debate. It is essential therefore to question why the CCW continues to be an unattractive forum for the discussions on autonomous weapons systems.

Combining the structural and procedural inadequacies of the CCW with the disproportionate dominance of highly militarised states in the CCW, it can be gleaned that substantive issues that smaller states grapple with in international security sometimes do not find expression in discussion outcomes in the forum. This is a worrying trend in international law-making which is likely to perpetuate international security problems that these forums seek to address. The neglect and relegation of issues of ethnic, racial, religious, gender, disability, and other biases in algorithms that (will) control autonomous weapons systems is likely to lead to the proliferation not the mitigation of conflicts and global polarisation when the effects begin to be fully felt among vulnerable populations. If these challenges are to be addressed, the voices that call for caution on the dehumanising potential of autonomous weapons systems must be heeded to.

Conclusions

The discussions at the CCW have gone on for years. For most states, the end goal is for a legally binding instrument that will regulate the development and use of autonomous weapons systems. For the highly militarised few, the debates are an opportune moment to reaffirm the applicability of existing international humanitarian law, which regrettably does not address issues like algorithmic bias. Both efforts, however, would be exercises in futility when it comes to the protection of the most vulnerable in global society if they are not meaningfully consulted in the process and if calls for the mitigation of bias in autonomous weapons systems are ignored or given a peripheral position in the discussions.

The problem of algorithmic bias has been extensively researched in academic and policy literature, but this has not translated to policy results at the UN level when it comes to

attempts to regulate autonomous weapons systems. This gap is even more surprising because side events at the CCW have been dedicated to such issues, Civil Society and academic advocacy have raised the same issues, and some states have voiced the concerns to do with racial, gender, ethnic, and other forms of bias in the discussions and in their submissions. Given these continued efforts, it is worrisome that the reports from the discussions have continuously relegated the issue of algorithmic bias and have not treated it with the detail that would be expected.

To prevent the risks in AI, the perspectives and concerns of those who are likely to be affected should be considered with full attention. It is true that states from the Global South, or small and vulnerable states have participated in the discussions within the confines of the GGE on LAWS in the Convention on Certain Conventional Weapons (CCW). However, representational equality does not automatically mean substantive equality. The relegation of the issue of algorithmic bias, particularly when it comes to race, ethnicity, religion, gender, and disability as raised by many states, in the CCW shows how the substantive outcomes of discussions may not reflect pertinent issues for the vulnerable members of global community. With the discussions still ongoing, we can only hope that such critical issues will gain traction and be given full attention for the protection of vulnerable states and peoples.

Notes

1. The Philippines statement, GGE on LAWS, 17 May 2023, accessible at https://conf.unog.ch/digitalrecordings/index.html?guid=public/61.0500/D90790E4-53C2-4A71-A4D3-9AFAE8A80A26_10h07&position=2498&channel=ENGLISH
2. Canada statement, GGE on LAWS, 17 May 2023, accessible at https://conf.unog.ch/digitalrecordings/index.html?guid=public/61.0500/D90790E4-53C2-4A71-A4D3-9AFAE8A80A26_10h07&position=3585&channel=ENGLISH
3. Mexico statement, GGE on LAWS, 17 May 2023, accessible at https://conf.unog.ch/digitalrecordings/index.html?guid=public/61.0500/D90790E4-53C2-4A71-A4D3-9AFAE8A80A26_10h07&position=8624&channel=ENGLISH
4. Report of the 2023 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Para. 27.
5. The recordings of the state debates/statements can be found at https://conf.unog.ch/digitalrecordings/index.html?guid=public/ and the state submissions and other conference documents are available at https://library.unoda.org/ and https://meetings.unoda.org/ccw-/convention-on-certain-conventional-weapons-group-of-governmental-experts-on-lethal-autonomous-weapons-systems-2023.
6. See https://www.palantir.com/aip/defense/ for more details.
7. Article 24(2b) Risk and Impact Management Framework- Committee on Artificial Intelligence (CAI), Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law.
8. Section 1.2 and 1.4(c).
9. Report of the Special Rapporteur on extrajudicial, summary, or arbitrary executions, Christof Heyns, 9 April 2013.
10. CCW/GGE.1/2017/WP.6 Working Paper entitled Autonomy in Weapons Systems submitted by the United States of America, 10 November 2017, Para. 30.
11. CCW/GGE.1/2018/WP5 Working Paper entitled Ethics and autonomous weapon systems: An ethical basis for human control submitted by the ICRC, 29 March 2018, Para. 45.
12. CCW/GGE.1/2019/3 Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/285/69/PDF/G1928569.pdf?OpenElement Para. 20(a).
13. CCW/GGE.1/2020/WP.7 Chairperson's Summary Para. 37(a).
14. CCW/CONF.VI/WP.3 Working Paper entitled Translating Ethical Concerns into a Normative and Operational Framework for Lethal Autonomous Weapons Systems submitted by Holy See, 20 December 2021, Para. 9.
15. CCW/GGE.1/2021/WP.7 Joint Working Paper, 27 September 2021, Para. 9(e).
16. CCW.GGE.1/2023/WP.3 Working Paper entitled Proposal for an international legal instrument on Lethal Autonomous Weapons Systems (LAWS) submitted by Pakistan, Para 13.
17. Joint Commentary of Guiding Principles A, B, C, and D by Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico, and New Zealand.
18. State of Palestine's Proposal for the Normative and Operational Framework on Autonomous Weapons Systems, March 2023.
19. CCW/GGE.1/2022/WP2 Working Paper entitled Principles and Good Practices on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States, Para. 32.

20. Working paper by the USA on "Human–Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems CCW/GGE.2/2018/WP.4 Para 45.
21. Details about this conference are accessible here https://conferenciaawscostarica2023.com/?lang=en
22. Belén Communique, 'Further action 2' https://conferenciaawscostarica2023.com/wp-content/uploads/2023/02/EN-Communique-of-La-Ribera-de-Belen-Costa-Rica-February-23-24-2023..pdf
23. Section II, CARICOM Declaration, 2023, found at https://www.caricom-aws2023.com/_files/ugd/b69acc_c1ffb97ed9024930a3205ae4e34c1b45.pdf

Declarations

Conflict of interest The author states that there is no conflict of interest.

References

Adelman, R.A. 2018. Security glitches: The failure of the universal camouflage pattern and the fantasy of "identity intelligence." *Science, Technology, & Human Values* 43 (3): 431–463. https://doi.org/10.1177/0162243917724515.

Ams, S. 2023. Blurred lines: The convergence of military and civilian uses of AI & data use and its impact on liberal democracy. *International Politics* 60 (4): 879–896. https://doi.org/10.1057/s41311-021-00351-y.

Arkin, R.C. 2010. The case for ethical autonomy in unmanned systems. *Journal of Military Ethics* 9 (4): 332–341. https://doi.org/10.1080/15027570.2010.536402.

Borrie, J., and A. Thornton, eds. 2008. *The value of diversity in multilateral disarmament work*. New York: United Nations.

Buolamwini, J. and Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification, in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. *Conference on Fairness, Accountability and Transparency*, PMLR, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html. Accessed: 14 August 2023.

Butler, J. 2009. *Frames of war: When is life grievable?* London: Verso.

Bylieva, D. (2022) *Language of AI*. https://doi.org/10.48417/TECHNOLANG.2022.01.11.

Carvin, S. 2017. Conventional thinking? The 1980 convention on certain conventional weapons and the politics of legal restraints on weapons during the cold war. *Journal of Cold War Studies* 19 (1): 38–69.

Chan, K. (2023) *What's new in robots? An AI-powered humanoid machine that writes poems*, *AP News*. https://apnews.com/article/robot-show-artificial-intelligence-chatgpt-0d0b4e0bfeec1860f16298bc70322e99. Accessed 15 August 2023.

Chengeta, T. 2020. Autonomous armed drones and the challenges to multilateral consensus on value-based regulation. In *Ethics of drone strikes: Restraining remote-control killing*, ed. C. Enemark, 170–189. Edinburgh: Edinburgh University Press.

Chengeta, T. (2022) Is the convention on conventional weapons the appropriate framework to produce a new law on autonomous weapon systems, in Viljoen, F. et al. (eds) *A life interrupted: Essays in honour of the lives and legacies of Christof Heyns*. Pretoria University Law Press, pp. 379–397. https://www.pulp.up.ac.za/edited-collections/a-life-interrupted-essays-in-honour-of-the-lives-and-legacies-of-christof-heyns. Accessed 5 April 2023.

Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (1980). https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/. Accessed 7 August 2023.

Dave, P. and Dastin, J. 2022. Exclusive: Ukraine has started using clearview AI's facial recognition during war, *Reuters*, https://www.reuters.com/technology/exclusive-ukraine-has-started-using-clearview-ais-facial-recognition-during-war-2022-03-13/. Accessed 15 Aug 2023.

European Commission (2021) *Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and Amending Certain Union Legislative Acts*, *2021/0106 (COD)*. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence. Accessed 9 Aug 2023.

FACEPTION\Facial Personality Analytics faception. https://www.faception.com. Accessed 14 Aug 2023.

Ferguson, D. 2023. Robots say they have no plans to steal jobs or rebel against humans, *The Guardian*, https://www.theguardian.com/technology/2023/jul/08/robots-say-no-plans-steal-jobs-rebel-against-humans. Accessed 15 Aug 2023.

Ferrara, E. 2023. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies.

Figueroa, M.D., et al. 2023. The risks of autonomous weapons: An analysis centred on the rights of persons with disabilities. *International Review of the Red Cross* 105 (922): 278–305. https://doi.org/10.1017/S1816383122000881.

Fitria, T.N. 2021. Gender bias in translation using google translate: Problems and solution. Rochester, NY. https://papers.ssrn.com/abstract=3847487. Accessed 15 Aug 2023.

Fredman, S. 2016. Substantive equality revisited. *International Journal of Constitutional Law* 14 (3): 712–738. https://doi.org/10.1093/icon/mow043.

Friedman, B., and H. Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on Information Systems* 14 (3): 330–347. https://doi.org/10.1145/230538.230561.

Ganguli, D. et al. 2022. Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. arXiv. https://doi.org/10.48550/arXiv.2209.07858.

Grant, N. and Hill, K. 2023. Google's photo app still can't find gorillas and neither can apple's, *The New York Times*, 22 May. https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html.Accessed 25 Jan 2024.

Grother, P.J., Ngan, M.L. and Hanaoka, K.K. (2019) *Face recognition vendor test part 3: Demographic effects*. National institute of standards and technology. https://www.nist.gov/publicatio

ns/face-recognition-vendor-test-part-3-demographic-effects. Accessed 14 Aug 2023.
Gutierrez, C.I. 2023. Uncovering incentives for implementing AI governance programs: Evidence from the field. *IEEE Transactions on Artificial Intelligence* 4 (4): 792–798. https://doi.org/10.1109/TAI.2022.3171748.
Hern, A. 2017. Facebook translates "good morning" into "attack them", leading to arrest, *The Guardian*, https://www.theguardian.com/technology/2017/oct/24/facebook-palestine-israel-translates-good-morning-attack-them-arrest. Accessed 15 Aug 2023.
Hunter, C., and B.E. Bowen. 2023. We'll never have a model of an AI major-general: Artificial intelligence, command decisions, and kitsch visions of war. *Journal of Strategic Studies*. https://doi.org/10.1080/01402390.2023.2241648.
Israel HLS&CYBER 2022—The international conference & exhibition Expo-Wizard. https://hls-cyber-2022.israel-expo.co.il/expo. Accessed 14 Aug 2023.
Johnson, J. 2022. The AI commander problem: Ethical, political, and psychological dilemmas of human-machine interactions in AI-enabled warfare. *Journal of Military Ethics* 21 (3–4): 246–271. https://doi.org/10.1080/15027570.2023.2175887.
Johnston, I. and Pitel, L. (2023) *German states rethink reliance on Palantir technology*. https://www.ft.com/content/790ee3ae-f0d6-4378-9093-fac553c33576. Accessed 25 Jan 2024.
Jones, C.M. 2021. Western centric research methods? Exposing international practices. *Journal of ASEAN Studies* 9 (1): 87–100. https://doi.org/10.21512/JAS.V9I1.7380.
Kim, Na-Young., Yoonjung Cha, and Hea-Suk. Kim. 2019. Future english learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning* 22 (3): 32–53. https://doi.org/10.15702/mall.2019.22.3.32.
Klugman, C.M. 2021. Black boxes and bias in AI challenge autonomy. *The American Journal of Bioethics* 21 (7): 33–35. https://doi.org/10.1080/15265161.2021.1926587.
Koenecke, A. et al. 2020. Racial disparities in automated speech recognition, *Proceedings of the national academy of sciences*, 117(14), pp. 7684–7689. https://doi.org/10.1073/pnas.1915768117.
Konert, A., and T. Balcerzak. 2021. Military autonomous drones (UAVs)—from fantasy to reality. Legal and ethical implications. *Transportation Research Procedia* 59: 292–299. https://doi.org/10.1016/j.trpro.2021.11.121.
Kordzadeh, N., and M. Ghasemaghaei. 2022. Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems* 31 (3): 388–409. https://doi.org/10.1080/0960085X.2021.1927212.
Lee, P. 2021. *Modern warfare: 'precision' missiles will not stop civilian deaths—here's why*, *The Conversation*. http://theconversation.com/modern-warfare-precision-missiles-will-not-stop-civilian-deaths-heres-why-171905. Accessed 25 Jan 2024.
Lovato, S. and Piper, A.M. 2015. Siri, is this you?: Understanding young children's interactions with voice input systems', in *Proceedings of the 14th international conference on interaction design and children*. New York, NY, USA: Association for computing machinery (IDC '15), pp. 335–338. https://doi.org/10.1145/2771839.2771910.
M'charek, A., K. Schramm, and D. Skinner. 2014. Topologies of race: Doing territory, population and identity in Europe. *Science, Technology, & Human Values* 39 (4): 468–487. https://doi.org/10.1177/0162243913509493.
Maas, M. and Villalobos, J.J. 2023. *International AI institutions: A literature review of models, examples, and proposals*. Legal Priorities Project. Available at: https://www.legalpriorities.org/research/international-ai-institutions.html.Accessed 3 Oct 2023.
Marchant, G.E., Tournas, L. and Gutierrez, C.I. 2020. Governing emerging technologies through soft law: Lessons for artificial intelligence. Rochester, NY. https://papers.ssrn.com/abstract=3761871. Accessed 22 Aug 2023.
Markl, N. 2023. *Language variation, automatic speech recognition and algorithmic bias*. PhD Thesis. The University of Edinburgh. https://era.ed.ac.uk/handle/1842/41277. Accessed 25 Jan 2024.
Maslej, N. et al. 2023. *AI Index Report 2023—Artificial intelligence index*. Stanford-California: Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/report/. Accessed 22 Aug 2023.
NIST AI 100-1. 2021. AI risk management framework', *NIST*. https://www.nist.gov/itl/ai-risk-management-framework. Accessed 22 Aug 2023.
OECD. 2022. Recommendation of the council on artificial intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed 22 Aug 2023.
Roff, H.M. 2014. The strategic robot problem: Lethal autonomous weapons in war. *Journal of Military Ethics* 13 (3): 211–227. https://doi.org/10.1080/15027570.2014.975010.
Ruttkamp-Bloem, E. 2023. Epistemic just and dynamic in Africa AI ethics. In *Responsible AI in Africa: Challenges and opportunities*, ed. D.O. Eke, K. Wakunuma, and S. Akintoye, 13–34. Cham: Springer.
Sap, M. et al. 2019. The risk of racial bias in hate speech detection, in *Proceedings of the 57th annual meeting of the association for computational linguistics. ACL 2019*, Florence, Italy: Association for Computational Linguistics, pp. 1668–1678. https://doi.org/10.18653/v1/P19-1163.
Sap, M. et al. 2020. Social bias frames: Reasoning about social and power implications of language, in *Proceedings of the 58th annual meeting of the association for computational linguistics. ACL 2020*, Online: Association for Computational Linguistics, pp. 5477–5490. https://doi.org/10.18653/v1/2020.acl-main.486.
Savoldi, B., et al. 2021. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics* 9: 845–874. https://doi.org/10.1162/tacl_a_00401.
Sharkey, N. 2010. Saying "no!" to Lethal autonomous targeting. *Journal of Military Ethics* 9 (4): 369–383. https://doi.org/10.1080/15027570.2010.537903.
Stix, C., and M.M. Maas. 2021. Bridging the gap: The case for an "Incompletely Theorized Agreement" on AI policy. *AI and Ethics* 1 (3): 261–271. https://doi.org/10.1007/s43681-020-00037-w.
Stoke White Investigations (2021) *France's shadow war in Mali: Airstrikes at the bounti wedding*. London, UK: Stoke White Investigations Unit/ Stoke White Ltd. https://www.swiunit.com/post/france-s-shadow-war-in-mali-airstrikes-at-the-bounti-wedding. Accessed 27 July 2023.
Suchman, L., K. Follis, and J. Weber. 2017. Tracking and targeting: Sociotechnologies of (In)security. *Science, Technology, & Human Values* 42 (6): 983–1002. https://doi.org/10.1177/0162243917731524.
UNESCO (2022) *Recommendation on the ethics of artificial intelligence*. Paris: United Nations Educational, Scientific and Cultural Organisation. https://unesdoc.unesco.org/ark:/48223/pf0000381137. Accessed 22 Aug 2023.
US Office of the Secretary of Defense (2007) *Unmanned Systems Roadmap: 2007–2032*. US Department of Defense. https://www.globalsecurity.org/intell/library/reports/2007/dod-unmanned-systems-roadmap_2007-2032.pdf.
Vincent, J. (2018) *Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech*, *The Verge*. https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai. Accessed 15 Aug 2023.
West, S.M., Whittaker, M. and Crawford, K. (2019) Discriminating systems: Gender, race, and power in AI, *AI Now Institute*.

Wilke, C. 2017. Seeing and unmaking civilians in Afghanistan. *Science, Technology, & Human Values*. https://doi.org/10.1177/0162243917703463.

Ishmael Bhila Doctoral researcher in international law and international security studying the participation of small states in the making of international law on autonomous weapons systems.