Human control and discrimination: responding to the challenge of drone-bombing by organized criminal groups in Mexico

Control humano y discriminación: respondiendo al desafío de los drones explosivos utilizados por grupos de crimen organizado en México

Ana Martín Beringola
Universidad Internacional de Valencia
Orcid: 0000-0002-4102-5410

Fecha de recepción: 01 marzo 2024
Fecha de aceptación: 07 junio 2024

"[A]ll predictions agree that if man does not master technology, but allows it to master him, he will be destroyed by technology" (Icrc, Commentary to Art. 37 Api, para. 1476).

Summary: I. Introduction. II. Delimiting the scope of violence. 1. Drone bombing by organized criminal groups. 2. The humanitarian nature of violations. III. Challenges posed by artificial intelligence in armed conflict. 1. Human control. 2. Machine bias. 3. Responsibility. IV. Suggestions for a domestic legal framework. 1. Integrating meaningful human control. 2. Addressing bias. V. Conclusion.

Abstract: This article addresses the challenge of drone bombing by organized criminal groups in Mexico as a surging but unregulated phenomenon. It is argued that, while the Government is using a law enforcement approach, drone bombing, which intensifies the nature of confrontations, falls under certain circumstances under the regime of international humanitarian law. This results in the need to consider the difficulties that autonomous weapons systems pose in situations of armed conflict. Accordingly, the article examines *human control* and *discrimination* as key

issues in the international debate around autonomous weapons systems. Based on this discussion, the article brings forward some proposals to inform a domestic legal framework applicable to drone bombing in Mexico –and autonomous weapons at large– in consistency with the State´s international obligations and the domestic policy framework.

Resumen: Este artículo aborda el desafío que plantean los drones explosivos utilizados por los grupos de crimen organizado en México como un fenómeno de creciente importancia pero no regulado. Se argumenta que, si bien el Gobierno utiliza un enfoque policial frente a estos ataques, los drones explosivos, al intensificar la naturaleza de la violencia, recaen bajo ciertas circunstancias en el ámbito del derecho internacional humanitario, lo cual hace necesario considerar las dificultades que los sistemas de armas autónomas plantean en situaciones de conflicto armado. En consecuencia, el artículo examina el *control humano* y la *discriminación* como cuestiones clave que plantea el debate internacional sobre sistemas de armas autónomas. Teniendo en cuenta este debate, el artículo realiza algunas propuestas que deberían informar el marco legal aplicable a los drones explosivos en México –y a las armas autónomas en general– coherente con las obligaciones internacionales del Estado y con su política interna.

Keywords: *international humanitarian law, human rights, discrimination, human control, autonomous weapons systems, drones.*

Palabras clave: *derecho internacional humanitario, derechos humanos, discriminación, control humano, sistemas* de *armas autónomas, drones.*

I. Introduction

The use of Artificial Intelligence (AI) has rapidly expanded as evidenced by its multiple applications, not only in daily life but, also, in exceptional situations like armed confrontations. This is taking place in Mexico where organized criminal groups are increasingly using unmanned armed drones to combat the State or other criminal organizations. The phenomenon, which commenced in 2017, is surging and even becoming a daily reality in some States (Sullivan 2021: 867; Bunker and Sullivan 2021; Plaw

et al. 2020: 101). Drones –aircrafts remotely piloted– are autonomous weapons systems (Aws). This implies that, after initial activation by a person, Aws are able to "select and apply force to targets without human intervention […] in response to information from the environment received through sensors and on the basis of a generalized 'target profile'"(Icrc 2021)[1].

Artificial intelligence (Ai), which involves the use of computer systems [algorithms] to carry out tasks that replicate human cognition, planning or reasoning (Icrc 2023) is being integrated in the military worldwide for its many benefits –intelligence, autonomous weapons and vehicles, resource management, logistics, surveillance, recognition– (Chandler 2021). Ai is featured by its capacity to act independently and efficiency (Mainzer 2020: 3), hence, the generalized interest to incorporate it in the field of State security interests. When Ai is embedded in Aws, these weapons allow obtaining a more complete analysis of conflict environment with capacity to advance critical functions in armed conflict such as targeting and engagement in the use of force. Their benefits include precision selecting targets –face, voice recognition, etc.–, reduction in the number of casualties, greater protection of civilian infrastructure (Salgado Espinoza 2020: 65) and, even indication of the time and location of crimes helping the process of accountability (Chandler 2021: 21)[2]. However, dealing with *targeting*, Ai-powered weapons pose a huge challenge. The limited human control over Aws "the user does not choose, or even know, the specific target and the precise timing and/or location of the resulting applica-

[1] International Committee of the Red Cross (Icrc).

[2] This is not to speak of the many benefits that artificial intelligence has in armed conflict, for instance, the identification of patterns of crime through satellite image, establishing evidence of incitement to commit crime through the use of algorithms in the social media, locating mass graves helping investigations, crime prevention and reparations. See, for instance Håkansson Schmidt (2022), see references; Porup, J.M. (2017), "Hunting for Mexico´s mass graves with machine learning" in *BIZ & IT*. Available at: «*https://arstechnica.com/information-technology/2017/04/hunting-for-mexicos-mass-graves-with-machine-learning/*» [Consulted on 28 February, 2024].

tion of force" (Icrc 2021) triggers ethical and legal debates around the issues of *control* and *responsibility*[3]. On the one hand, handing out *human control* to Aws for critical target functions would imply leaving live or death decisions in the hands of machines. On the other hand, the use of Aws raises the question of who is criminally responsible for machine failures if a human agent has no meaningful control –no agency *ergo* no intent– over the self-initiation of the machine in targeting and engagement (Asaro 2016: 190-192).

This paper addresses the legal gap surrounding the use of explosive drones by organized criminal groups in Mexico with the aim of suggesting a way forward to inform the domestic legal framework. In this way, the article paves the way for the future use of Aws by non-State and State actors. The paper highlights that Aws raise the intensity of violence resulting in the need to rethink the nature of armed violence in Mexico and the potential application of international humanitarian law (Ihl), currently not applied in the State´s confrontation with cartels. As it is argued, under *certain* conditions, the legal requirements triggering the application of Ihl exist in Mexico. Accordingly, the article delves on the circumstances that make the application of Ihl a complicated task at the international and national levels in that State. One reason hampering the application of Ihl is the lack of international legal standards on Aws. Another reason is official reluctance to recognize the potential application of Ihl to national security issues. Discrimination is identified as another issue requiring consideration in a future domestic legal framework in Mexico. Studies reveal that Ai tends to *generalize* patterns or features of conflict environment resulting in discrimination in the coding of algorithms. Consequently, machine bias is often behind failures in targeting and engagement leading to the violation of Ihl principles, including distinction, proportionality, non-discrimination that result

[3] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): "Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems", CCW/GGE.1/2022/WP.5.

in unnecessary harm (Chandler 2021; Houghton 2022; Håkansson Schmidt 2022). Discrimination in the use of AI is rampant in Mexico (Spitx *et al.* 2023) and it is especially detrimental for the most vulnerable groups such as women, girls and indigenous peoples (Casados *et al.* 2020). Moreover, as structural discrimination is exacerbated in armed conflict, machine bias is expected to aggravate these inequalities (Red de Investigación Feminista en Inteligencia Artificial *et al.* 2023). This fact weakens the official strategy under the National Plan of Peace and Security to end violence by tackling its root causes of discrimination and promoting human rights.

From this backdrop, the article´s methodology first focuses on the international legal framework applicable to drone attacks in Mexico and, more broadly, to Aws. Then, with this insight, it engages on a discussion of *de lege ferenda* at the domestic level in Mexico to address the phenomenon of drone bombing. Accordingly, the article is divided in three sections. The first section discusses the phenomenon of drone bombing in the context of confrontations between criminal organizations and the State in Mexico providing arguments for the application of IHL in certain situations. Then, the article addresses the ethical challenges posed by Aws making their regulation complicated in IHL, paying special attention to the issues of *human control*, *machine bias*, and responsibility. Following this discussion, the article finally suggests the tenets of a responsible domestic legal framework in line with Mexico´s international obligations and national policies. It is hoped that, in doing so, Mexico will set an example of good practice which might help moving the global discussion ahead in that direction.

II. Delimiting the scope of violence

1. Drone bombing by organized criminal groups

Over the past years the intensity of violence between Mexican authorities and criminal groups, and between criminal groups

in Mexico, has drastically surged[4]. This fact questions the effectiveness of the official strategy which shifted from waging war against drug cartels towards addressing the structural causes of violence underpinned by gross inequalities and poverty (Plan Nacional 2018-2024: 2,7). Complex reasons explain the difficulty for halting organized criminal groups in Mexico, including the deterioration of living standards, the growth of the illegal drug market and a failed public strategy against organized crime (Padin 2023: 94). Growing militarization in Mexico, whose military budget increased by 52% during the term of President Andrés Manuel López Obrador (2018-2024), has resulted in the country advancing 11 positions within the most powerful armies from 2022 to 2023. However, militarization has not reversed trends in violence. Instead, both militarization and violence have gone in tandem questioning the official goal of caring for human rights to defeat drug-related violence (González 2023)[5]. Interestingly, what has been almost unnoticed is that the socialization of technology among cartels –especially the use of AI– has been a major driver of the rising levels of organized crime and violence in Mexico (Plaw *et al.* 2020: 95).

Aircrafts without a pilot and remotely controlled, commonly known as drones, are responsible for a large amount of the new forms of violence involving Mexican cartels. While the use of drones by criminal groups is not new, the constant technological improvements of AI are now reverberating into more intense confrontations. Cartels have resorted to drones strategically to gain an advantage over rivals in four main ways (Bunker and Sullivan 2021). From 2010, drones were employed to smuggle narcotics due to their considerable capacity and the difficultly of radars

[4] Over the past 15 years, more than 35.000 intentional killings have been registered per year and the number of persons disappeared exceeds 100.000. See, WOLA (2022): "México profundiza la militarización. Los hechos muestran que es una estrategia fallida" in *WOLA*. Available at: «*https://www.wola.org/es/analisis/mexico-profundiza-militarizacion-hechos-muestran-estrategia-fallida/*» [Consulted on 16 February, 2024].

[5] Popularly known in Spanish as "política de abrazos, no balazos".

to detect them (Plaw *et al.* 2020: 101). A few years later drones served for intelligence, surveillance and reconnaissance purposes, including night vision combat. From 2017, they started to be used as weapons for targeting purposes. In 2020, from the beginning of the pandemic, they have served propaganda purposes competing against the State for political support. Whilst all of these functionalities make up the cartels criminal enterprise, the *weaponization* of drones is directly linked with an increase in violence and harm in armed confrontations in Mexico.

Over 600 attacks with explosive-equipped drones have been attributed to criminal groups in Mexico, the majority in the states of Guerrero, Tamaulipas and Michoacán (Mexico Now 2023). In the latter, drone-bombing is a nearly daily occurrence, with 260 of such attacks recorded in 2023 (Cbs 2023). Drones are used by non-state actors as weapons and as weapon platforms and are becoming increasingly sophisticated –they can be single-use– and, often, multi-use platforms (Bunker and Sullivan 2021). They often incorporate homemade weapons such as explosives, nails and pellets due to the online training of technological skills by cartel members. Further, drones are operated by specialist *droneros* –drone operators–. Certain criminal groups, as the Cartel Jalisco Nueva Generación (Cjng), are training groups of drone operators to produce explosives and operate drones (Mondragón Toledo 2023). Critically, a consequence of using Aws like drones is the increasing asymmetry of warfare between the parties and the lower threshold of escalation, which is resulting in more severe harms[6]. For instance, to respond to drone firepower, the Mexican Government is acquiring helicopters equipped with rotating electric machine guns as, allegedly, the only way to overpower the increasing drone bombing capacity of the Jalisco cartel (Stevenson 2021).

[6] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): "Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems", CCW/GGE.1/2022/WP.5.

The escalation of violence due to the use of explosive-equipped drones can be observed in the evolution of incidents. In Tecate (Baja California, July 2018), in a single-use incident, a police commander found a weaponized drone with two grenades and a note threatening not to mess with the local cartels. In Aguililla (Michoacán), the Jalisco and Viagras cartels regularly fight each other with bomb-dropping drones resulting in serious damage to civilian infrastructure and violence. This has triggered the use of dangerous counter-weapons by the Government. For example, the public authorities have resorted to Blackhawk helicopters for targeting drones, also called *indiscriminate blanket fire* (firing 6000 rounds per minute), which are prohibited in most civil conflicts (Stevenson 2021). Another example of escalation is violence in Nuevo Caracol (Guerrero, May 2023). While drone attacks were usually carried out between cartels, these attacks are now spreading to the rural community where they are taking their toll: civilian killings, damage of civilian infrastructure and the spread of terror have led to the forced displacement of an estimated number of 600 people in 2023, which has resulted in 67 communities demanding protection to the local and federal authorities (Mondragón Toledo 2023).

2. The humanitarian nature of violations

There is an alarming gap in Mexico´s law and policy to address the use of AI in armed confrontations. Neither the National Plan on Peace and Security (National Plan 2018-2024) nor the National Agenda on Artificial Intelligence consider the implications of using AI in the military and security sectors in spite of the seriousness of the extant evidence and its trend to spread and become more sophisticated (Bunker and Sullivan: 2021).Yet, the high levels of armed violence in Mexico involving drones and the loophole around the use of Aws bear the question: What is the nature of violations resulting from the use of Aws by Mexican cartels? Answering to this question is of the utmost importance for minimizing legal uncertainty, arbitrary responses and, instead, clarifying the standards that would provide suitable protection against

violations as the use of Aws by non-State actors continues to increase. However, answering to this query is challenging for two reasons. One is the lack of binding international standards to guide a domestic legal framework in Mexico. Another is the difficulty of qualifying the nature of the conflict between the cartels and the State, and between cartels –i.e. armed or civilian in nature– whose discussion is viewed as opening Pandora´s box (Padin 2023).

Whether armed hostilities involving Mexican cartels constitute an armed conflict or internal disorders has been considerably debated. At stake is the application of either the law of armed conflict (LOAC), IHL or, alternatively, a law enforcement regime. The former would imply applying binding international standards on Mexico, namely, the Geneva Conventions (1949), Additional Protocol I[7] and customary international law. Conversely, the later would entail political freedom to regulate the conflict as an internal issue whilst taking a positive note of relevant soft law standards[8] and ensuring compliance with International Human Rights Law (IHRL). It is worth noting that, although Mexico has not ratified Additional Protocol II[9] concerning non-international armed conflicts –an indicator that it does not consent to international interference in domestic confrontations– Article 3 Common to the Geneva Conventions applicable to internal armed conflicts is legally binding. Similarly, Mexico must abide by customary international humanitarian law, applicable at all times (ICRC 2005), and international jurisprudence. It is also the case that an eventual qualification of certain situations

[7] (ICRC), Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 UNTS 3 (8 June 1977).

[8] Refer to "The Code of Conduct for Law Enforcement Officials", adopted by General Assembly resolution 34/169 (1979), and to the "Basic Principles on the Use of Force and Firearms by Law Enforcement Officials of 1990", adopted by the Eighth United Nations Congress on the Prevention of Crime and the Treatment of Offenders, Cuba, 1990.

[9] ICRC, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non- International Armed Conflicts (Protocol II), 1125 UNTS 609, 8 June 1977.

in Mexico as armed conflict would not exclude the application of IHRL, whose protection does not cease in armed conflict. IHRL would apply in a complementary way allowing greater protection in situations not covered by the *lex specialis* or IHL (ICJ 2004: para. 106), for instance, acting as a safeguard of fair treatment and judicial guarantees against members of organized criminal groups.

While law enforcement, characterized by police arrest powers[10] is the official approach to address organized crime in Mexico, the reality is a *de facto* militarization. The Government has consistently recalled the application of IHRL when its armed forces act in security operations. This official position, complemented by application of law enforcement standards and the lack of ratification of Additional Protocol II to the Geneva Conventions, are good indicators of the official view that internal hostilities do not attain the threshold of Common Article 3 which would trigger the application of IHL in Mexico (Padin 2023: 792). In this regard, the National Peace and Security Plan (2018-2024) establishes the aim to pacify the country through the explicit goal of national peacebuilding and reconciliation, which is to be achieved by securing respect for human rights. In this regard, the problem of national insecurity and violence is officially recognized as a one of structural inequalities which must be addressed from a human rights-based approach (Chanona 2019: 6). To accomplish this aim, the National Peace and Security Plan articulates a proposal for transitioning from a public security model reliant on the military forces to one of enhanced police forces, as shown by the creation of the National Guard (Guardia Nacional), a police body with larger powers. However, this official position is contradicted by the means of implementation. The National Guard is largely made up of military staff and dependent upon the National Defense Secretary (SEDENA),

[10] The term "police" refers to officers who exert the powers of arrest or detention to implement the law and order, which may imply the need to resort to the use of force. See General Assembly of the United Nations (2014): "Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns", para. 22, A/HRC/26/36.

which has resulted in an increasing militarization of public security issues. The official approach is further complicated by a lack of clear protocols on the use of force under the National Peace and Security Plan, which is critically silent on the standards applicable to the key substantive issue of the use of force against organized criminal groups (Chanona 2019: 7-8).

Regarding the nature of the conflict involving State and non-state actors, international standards consistently establish the existence of an internal armed conflict when two concurrent conditions are present: a degree of organization and protracted violence[11]. On the one hand, international jurisprudence considers the *organization* criterion fulfilled in the presence of indicators such as the existence of a command structure, the ability to gain access to weapons, military equipment and training, the ability to plan, coordinate and to carry out military operations[12]. On the other hand, international precedents declare that violence is protracted based, not as much on duration, as on the intensity of hostilities appreciated by non-exhaustive factors, including the existence of armed clashes, the mobilization and distribution of weapons, the types of weapons and military equipment, the number of casualties due to shelling and fighting, and the quantity of troops deployed[13].

Whether the criteria of organized and protracted violence determining the existence of an armed conflict and triggering the application of IHL exist in Mexico is highly debatable. Critics argue that the main goal of cartels is to obtain economic gain and, consequently, they adapt their organizational structure not to confront

[11] International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Duško Tadić*, 02 October 1995: para. 70.

[12] International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Ramush Haradinaj, Idriz Balaj and Lahi Brahimaj,* 3 April 2008: para. 60. International Criminal Court, *The Prosecutor v. Thomas Lubanga Dyilo,* 14 March 2012: para. 537.

[13] International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Ljube Boškoski and Johan Tarčulovski,* 10 July 2008: para 177. *The Prosecutor v. Thomas Lubanga Dyilo:* para. 538.

the Government *per se* but to protect their *plazas*. Accordingly, as cartels diversify their profitable activities, their structure becomes more decentralized (Padin 2023: 780-781; Muggah 2023: 570). Furthermore, despite the surge in violence in Mexico from 2007, studies show that only a small number of homicides is related to violence between State officials and illegal armed groups. This pattern is mainly attributed to situations combining drug consumption and poverty as pulling factors that affect the most vulnerable populations, therefore, confirming that economic (not political) factors are the key drivers of most murders directly related to organized crime. Accordingly, what extant research reveals is that *most* of the violence involving organized criminal groups in Mexico does not reach the threshold of armed conflict and, therefore, it is not suitable for the IHL paradigm (Padin 2023: 786).

These observations are, however, not conclusive. Despite the increasing fragmentation of Mexican cartels and their fighting for lucrative aims (Bellal 2017: 86), the latest ICRC´s interpretation of armed conflict supports the notion of *aggregated intensity*. According to this view, when a coalition of armed actors has sufficient level of coordination, the intensity of hostilities "between each of them and an opposing party may be aggregated when considering whether the threshold of intensity has been reached" (D´-Cunha *et al.* 2024:3). According to this view, the lack of *traditional* hierarchical structure of Mexican cartels would not prevent from considering the aggregation of violence of each of these criminal *coalitions* for the purpose of establishing the intensity requirement of an armed conflict. Additionally, even if cartels´ criminal activities mainly rely on economic objectives, international jurisprudence considers that this fact does not prevent them from becoming parties to an armed conflict. In this regard, the International Criminal Tribunal for the Former Yugoslavia (ICTY) has noted that the existence of an armed conflict "is based solely on two criteria: the intensity of the conflict and organization of the parties,

the purpose of the armed forces to engage in acts of violence or also achieve some further objective is, therefore, irrelevant"[14].

According to experts, the violence involving *certain* criminal groups in Mexico amounts to a non-international armed conflict in terms of organization and intensity. It is considered that, there are at least two internal armed conflicts taking place between the Government and the CJNG and the Sinaloa Cartel (CDS) (Bellal 2017: 87). Additionally, it has been noted that violence between the CDS and the CJNG reaches the threshold of an internal armed conflict as well (RULAC 2022). The CDS consists of a coalition of leaders of criminal organizations that acts as a federation relying on corruption and internal alliances. For more than ten years, the CDS has been involved in violent turf wars against the Mexican armed forces and, more recently, against other cartels for the control of drug trafficking routes (Bellal 2017: 87; RULAC 2022). In 2022, violence was particularly intense, although only one incident –a prison riot with 11 people killed, mostly civilians– has been attributable with certainty to the CDS (RULAC 2022). As for the CJNG, which emerged in 2010 as a splinter of the CDS, it has expanded its operations fast to many federal states. Considered one of the most aggressive groups, it uses highly sophisticated weapons in high-intensity attacks both against the State security forces and against other cartels, especially the CDS, to control illicit drug routes (Bellal 2017: 87).

The increasing use of drone bombing by powerful Mexican cartels reinforces the argument that criminal groups, such as the CDS and CJNG, are parties to an internal armed conflict that would trigger the application of IHL. On the one hand, the training of specialized drone operators, as is the case of the CJNG (Mondragón Toledo 2023), reflects enhanced organizational capacity in terms of access to weapons, military training and equipment, ability to plan, coordinate and carry out armed attacks, even if its orga-

[14] International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Limaj et al., 30* November 2005.

nizational structure has become more loosely connected. On the other hand, incidents, like those mentioned above, where drone bombing is repelled by battle-grade force such as Blackhawk helicopters and rocket launchers by the State security forces (Shuldiner 2022), prove that escalation is a direct consequence of the cartels´ increasing reliance on explosive-equipped drones. While violence between the Government and cartels, and between cartels, remains high, it becomes increasingly difficult to attribute violent incidents to specific armed groups, which has led some observers to declassify these armed conflicts due to a lack of certainty (Rulac 2022). A declassification –coming to an end– of an internal armed conflict takes place when one of its two conditions ceases to exist (D´-Cunha *et al.* 2024). However, in the case of drone bombing, there is abundant evidence –as described by this section– that cartels, including the Cds and Cjng, are increasingly using this technique, and that this fact is aggravating the scale and consequences of armed confrontations, even if complaints of armed attacks by the affected communities are often not followed-up by the authorities (Janetsky 2024; Investigator.Mx 2024). As it becomes plausible that Ihl applies in certain contexts involving drone bombing attacks in Mexico –and this possibility increases in the future as drones and Aws become more commonplace– a dilemma arises: There are no international legal standards on the use of Aws in armed conflict to guide policy-making in Mexico. Accordingly, it is necessary, as will be examined in the next section, to review the main challenges that are preventing from reaching an agreement on the international standards applicable to Aws and, consequently, that are making it difficult address the phenomenon of drone bombing in Mexico.

III. Challenges posed by artificial intelligence in armed conflict

Aws pose two major legal issues which are unregulated: human control regarding critical combat functions and discrimination or bias in the programming of algorithms embedded in Aws. This

section addresses these legal gaps in primary norms of international law and the resulting challenge they give rise to in secondary norms in terms of establishing international responsibility for these loopholes.

1. Human control

The autonomy of Aws constitutes their curse. AI allows the autonomous implementation of highly complex tasks resulting in a reduced ability of human agents to understand, predict, or mitigate their impacts (Asaro 2016: 190). One thing is decisive: the greater the autonomy of weapons, the lesser the ability of humans to control the machine. Actually, the degree of human-machine interaction can be programmed to enable humans exercise greater or lesser agency. This translates into different degrees of control. Human operators may be either *in the loop* –the machine identifies a target but a human must validate it before the machine initiates engagement– or *on the loop* –the machine identifies the target and commences engagement unless the human agent overrides it– (Coco 2023:3-4). Accordingly, Aws give rise to a critical ethical issue: a *control* problem regarding the critical functions of selecting the target and engagement in the use of force, hence, the difficulty of attributing responsibility for the harms caused by Aws when human control cannot be adequately established. The more autonomous weapons are, the more limited human control –and agency– over them, even in situations of negligence, resulting in the dilemma of who is responsible which risks turning upside down the paradigm of accountability for gross violations in armed conflict. Likewise, the human control problem may compromise other responsibility regimes. This may apply to the international responsibility of States for gross violations committed by its agents, and it may also interfere with domestic regulations sanctioning Aws, including the use weapons and explosives from the airspace. Observers have suggested that a taxonomy of the use of increasingly autonomous Aws in critical targeting and engagement functions may include the following levels of autonomy:

> "1) A human engages with targets, selects targets and initiates attack; 2) A program suggests alternative targets and a human decides which to attack; 3) A machine selects targets which a human must validate before engagement; 4) A machine selects and engages targets under supervision of a human who may override the machine´s choice and halt the attack; 5) a machine selects targets and initiates attack according to the mission´s goals defined at the planning/activation stage without additional human intervention" (Amoroso and Tamburrini 2020: 191).

Even though all these levels, except level 1, involve a degree of unpredictability in critical functions, the degree of autonomy and corresponding human control over Aws clearly varies (Acquaviva 2021: 7). Only levels 1 to 3 allow human agents to be *on the loop* and, therefore, the possibility of retaining a *significant* degree of human control. Levels 4 to 6 allow no significant human control, and control is completely absent in stage 5. Furthermore, the scenario complicates if Aws failures are combined with situations where human operators do not exercise control properly such as situations of automation bias –excessive human trust in the machine´s determinations despite contradicting information from other sources– and complacency –excessive trust in the machine´s determination resulting in diminished attention from human operators– (Coco 2023: 1). Who is under control in these situations: humans, robots, or both?[15] Who violates the laws of armed conflict? Can we conceive IHL without significant human intervention resulting in warfare being waged by robots instead of by humans?[16] Situations where human control is not significant or absent are the object of diplomatic debate and, still, are far from reaching the necessary consensus for their regulation necessary for a consistent and principled application of IHL.

[15] For an excellent discussion on the topic see, Acquaviva (2021).

[16] In the same sense, consider the ethical and legal concerns on lethal Aws raised by Special Rapporteur: i) compliance with IHL, ii) attribution of responsibility, iii) violation of human dignity, and iv) risk for peace and stability. See General Assembly of the United Nations (2013): "Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns", A/HRC/23/47.

As a warfare weapon, the legality of Aws must be considered within IHL´s general framework, where it is a highly divisive matter. A first consideration is that States have the duty to assess whether the employment of new weapons would be prohibited by IHL[17]. Thus, it impinges on each Government a preliminary duty to create internal procedures to elucidate the legality of a new Aws before putting it to use (ICRC 1987). The academic debate on the legality of Aws moves between the waters of a full ban and regulation[18]. Those in favor of a complete prohibition of Aws posit their incapacity to ensure respect for IHL because they lack the moral qualities of awareness and judgment which only humans have (Amoroso and Tamburrini 2020: 189). In this sense, it is observed that interactions among Aws in unstructured warfare contexts are highly unpredictable to the extent of rendering combat *beyond* human control, which would amount to accepting the possibility of removing ethics from warfare. Furthermore, it is argued that Aws make wars easier to wage due to the possibility of saving more military lives in one´s own camp –as more militaries operate remotely– which would pose increasing risks for the maintenance of international peace and security (Sharkey 2012)[19]. Conversely, arguments in favor of a controlled development of Aws and their regulation claim numerous benefits. These include regulating transparency, increasing control over State´s use of these weapons, and potential

[17] While Additional Protocol I refers to international armed conflicts only, consider that other IHL norms potentially applicable to Aws do apply to internal armed conflict. For instance, consider the prohibition to use indiscriminate weapons, which includes those whose effects "cannot be limited" which may be the case of Aws without human control. Additional Protocol I 1977: Article 36.

[18] Consider also international responses. For instance, the International Committee of the Red Cross (ICRC) supports banning unpredictable Aws (as does the UN Secretary-General) and autonomous target recognition, but it endorses the regulation and use of Aws. The Human Rights Commissioner for Human Rights has called for a moratorium on the sale and use of Aws until safeguards ensure compliance with IHRL standards (see Chandler 2021: 25).

[19] General Assembly of the United Nations (2013): "Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns", A/HRC/23/47.

to save human lives due to the accuracy of their data. It has even been argued that *ethically restrained* Aws would abide by the laws of war better than humans, who are subject to emotional reactions in the battlefield that do not affect robots (Arkin 2009). Moreover, prohibiting Aws would stifle innovation, economic growth and development, and it would involve renouncing to the possibility of devising alternatives for more proper management of AI in fields where it can provide considerable advantages (Asaro 2016).

Diplomatic responses have failed to provide an international legal framework on Aws so far, a core reason being the objection amongst powerful States to limit their technological lead (Amoroso and Tamburrini 2020: 192)[20]. Nonetheless, the international community is engaged, although limitedly, in the process of regulating lethal Aws[21]. The most relevant forum, and most concrete proposal, are the Guiding Principles on Lethal Autonomous Weapons Systems (LAWS), adopted by consensus by the Group of Governmental Experts (GGE) in the context of the Convention on Certain Conventional Weapons (CCW)[22]. These recommendations –where Mexico has played a leading role– aim to establish the basis for a "Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapons Systems"[23]. They have merit for adopting a middle way position between the opposite approaches, i.e.,

[20] Critical detractors of regulation include the Us and the Russian Federation.
[21] The debate on Aws began in the Human Rights Council. Then, the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions called for a high-level panel to articulate an international policy in its 2013 report. From 2013, the Group of Governmental Experts within the CCW has taken up the issue and made substantive recommendations after eight years of discussions.
[22] The aim of the CCW is to restrict and ban the use of weapons deemed to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately.
[23] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): "Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems", CCW/GGE.1/2022/WP.5.

a complete ban and the regulation of AWS, in the spirit of garnering enough consensus to agree on legally binding standards.

The Guiding Principles´ middle way position between a full ban *versus* the regulation of Aws is rooted in the claim that human agency must prevail and underpin all prohibitions and regulations of Aws. Accordingly, the Guiding Principles suggest a straightforward ban on those Aws that remove human control, do not respect IHL principles, and whose effects cannot be sufficiently understood. From this position, which ensures respect for existing international law, the Guiding Principles leave space for negotiating a regulatory framework. In this regard, the Guiding Principles suggest two key regulations: ensuring *sufficient* human control –that will require a definition– in the critical functions of targeting and the use of force –while recognizing that the scope of human control may vary depending on the context environment– and establishing individual criminal responsibility for violations committed using Aws. Accordingly, this international initiative makes room for lethal Aws in IHL but always within the framework of sufficient human control, respect for IHL and international responsibility.

The notion of *sufficient* human control undergirding the GGE Principles emerges as a key concept in international negotiations on Aws. This term matches what observers consider the *meaningful* human control of Aws which, accordingly, may provide helpful guidance establishing the boundaries of what this degree of human control entails. Meaningful human control has been defined by a threefold threshold: i) humans act as a fail-safe actor –the machine design prevents/mitigates the endangering of life or property in situations concerning direct attacks against civilians or excessive damage–; ii) human accountability –individual criminal responsibility for Aws violations–, and iii) respect for human dignity –moral agency– so that decisions relative to harm caused to people and property are not left to artificial agents (Amoroso and Tam-

burrini 2020: 192)[24]. Internationally agreed standards on the scope of meaningful/sufficient human control seems, therefore, of the utmost importance as a bulwark that mitigates the ethical dilemma posed by the autonomy of Aws. Another advantage of the notion of sufficient/meaningful human control under the Guiding Principles –which may be also of great help in future negotiations– is that it makes room for a "jointly differentiated and prudential" approach (Amoroso and Tamburrini 2020: 192), thereby, avoiding a one-size-fits- all approach to Aws and allowing States a margin of framed autonomy for decision-making in critical targeting and engagement functions.

2. Machine bias

Machines, like the humans who create them, are not infallible and make mistakes. Differently from humans though, machine failures are not underpinned by agency and moral judgment but by codes of algorithms which, in fact, make robots as good as their data (Farrés 2021: 3). The problem with algorithms, as those informing Aws, is that they are frequently informed by bias or discrimination. Gender-based discrimination is a good example. According to studies, gender discrimination is pervasive across systems of AI in warfare. Recognition tools are 70% more likely to recognize the voice of a man than of a woman, and relief systems tend to overlook the specific needs of women and girls (Houghton 2022: 2). Women are frequently stereotyped as submissive and sexualized whereas men tend to be, by default, associated with combatants, resulting in the filtering of content that exacerbates preexisting inequalities (Un Håkansson Schmidt 2022: 1-3). Likewise, gender intersecting with race discrimination is embedded in military codes. While facial recognition software makes 1% of error for light-skin women, this figure elevates to 35% for dark skin women (Håkansson Schmidt 2022: 1). Images from the most populated countries

[24] Note the links between this conceptualization of human control and the conclusions of the Special Rapporteur on extrajudicial summary or arbitrary executions (Heyns 2014).

–as China and India– only represent 3% of the total images of AI systems. Accordingly, while a white woman wearing a white dress is recognized as a bride, a North Indian woman wearing a sari is labelled as an artist (Chandler 2021: 15). Overall, there is a tremendous problem of representation informing the algorithms of Aws resulting in that bias underpins the machine´s critical functions in armed conflict.

Aws failures in selecting targets and engagement recurrently violate core IHL norms such as the principles of distinction –between civilians and combatants– and proportionality – refraining from causing excessive civilian loss in comparison with concrete and direct military advantage–[25]. Interestingly, observations from the field reveal that drone bombing is the autonomous weapon *par excellence* responsible for causing these violations. This might be due to algorithm bias combined with the features of Aws, which are deprived of sensory systems for separating civilians from combatants and lack situational awareness for making proportionality decisions (Sharkey 2012: 789-790). In Afghanistan, drones for signature strikes –taking place outside traditional battlefields– were programmed to consider any military-age male a military target. This gross generalization of the target profile led to a weaponization of civilians in Afghanistan by turning civilian men into military targets. This discriminatory coding resulted in a gross breach of the principle of distinction and it had a disproportionate impact on the number of civilian casualties that combatants must spare above all (Farrés 2021: 2). In Libya, *kamikaze* Aws –selecting and attacking human targets based on proximity to an object– have been labelled *battlefield experiments* and are considered key contributors of the serious humanitarian crisis in that country (Chandler 2020: 17-18). Similarly, research on the Syrian conflict suggests that selecting aerial bombardment as a method of warfare has a drastic gender discriminatory impact. According to estimates, these methods have caused a disproportionate number of women and girls being killed in the conflict –37% women v 23% men– when com-

[25] Additional Protocol I, Arts. 48 and 51.

pared to other methods of warfare such as shooting –13% women v 23% men– (Farrés 2021: 6).

Avoiding bias in Aws requires the adoption of anti-approaches which are now absent in the coding industry, including a conflict analysis of preexisting discrimination and the mainstreaming of non-discrimination through the coding cycle. A conflict analysis of discrimination should consider factors such as the conflict dynamics, culture, and existing inequalities that help understand the nature and effects of violence. Then, these should be integrated in the coding of algorithms so that Aws can factor this sensitive information in. In this regard, Chandler notes, "[i]f AI [artificial intelligence] 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 intelligence" (Chandler 2020: 9). If, as studies have shown, algorithm bias is structural, its elimination requires the mainstreaming of non-discriminatory approaches throughout the coding cycle, including the manufacturing, developing, use and control of Aws. Some key non-exhaustive measures include: inclusive gender programming –women represent 24% of computer scientists–; hiring experts to sensibly inform algorithm data (Chandler 2020: 9); a context-specific approach taking note of preexisting inequalities affecting the most vulnerable groups in peacetime which are likely to exacerbate with technologies (Spitx *et al.* 2023: 6-7), and reflecting the views of underrepresented groups, mainly women, children and other marginalized groups at a higher risk of having experienced the harms of machine failures (Håkansson Schmidt 2022: 2). Above all, informing Aws with an inclusive non-discriminatory approach to conflict analysis demands what feminist scholars have called a *human and differentiated impact* which takes note of the gender and other intersecting inequalities affecting marginalised groups who are the most affected by armed conflict (Red de Investigación Feminista en Inteligencia Artificial *et al.* 2023: 3). Overall, if Aws bias is to be avoided, it makes sense that algorithms are specifically programmed to avoid discrimination.

3. Responsibility

IHL is built upon the paradigm that only humans can be accountable for the agency and choices resulting in violations in armed conflict. Accordingly, "no one shall be convicted of an offence except on the basis of individual criminal responsibility"[26]. This paradigm is challenged by Aws where human agency –hence intent– typically preceding targeting and the use of force is blurred by the autonomy of these weapons. The Rome Statute (Rs) establishing the International Criminal Court (ICC), and requiring knowledge and intent to establish liability[27], illustrates well that Aws challenge the paradigm of individual criminal responsibility. As Coco convincingly notes, when Aws make targeting choices – regardless of whether humans are *in the loop* or *on the loop*), international criminal law (ICL) is ill-equipped to establish the *mens rea* or subjective elements (Coco 2023: 11-12). Considering the war crime of intentionally directing attacks against the civilian population[28], even if a human must validate or supervise the machine´s targeting choices, human failure to recognize a wrong choice by the machine would likely deny their knowledge of the civilian status of victims and, consequently, also their intent. The outcome would be similar for the war crime of intentionally directing an attack causing excessive incidental death, injury or damage[29]. The perpetrator would not be blamed for not being aware of the machine´s failure to evaluate that the attack would cause excessive damage. Likewise, the responsibility of commanders and superiors[30] would not be engaged if they did not know or should have known that the machine would fail in critical functions of targeting and engagement. Even if such knowledge existed, robots –lacking human

[26] Additional Protocol I, Art. 75(4)(b); Additional Protocol II, Art. 6(2)(b)
[27] UN General Assembly, Rome Statute of the International Criminal Court (last amended 2010), ISBN No. 92-9227-227-6, UN General Assembly (17 July 1998), Article 30.
[28] Rs, Art. 8.2, section b, fraction i.
[29] *Ibidem*, fraction iv.
[30] Rs, Article 28.

agency– could not be considered subordinates who committed crimes for which superiors are responsible in the current state of international law (Coco 2023: 15-16).

The *control* problem brings about the immediate dilemma of who is internationally responsible for IHL violations committed as a result of Aws failures during targeting and the use of force. If humans cannot be held criminally responsible, who then? Scholars are putting their legal imagination to work to answer this question due to the increasing use of Aws –among which critically drone-bombing– in conflict settings[31]. Acquaviva (2021) suggests various alternatives of how a responsibility framework might work. A first option would consider that only humans are responsible for Aws mishaps because IHL and its sanctioning through ICL are built on that premise. This solution may require some adaptations though. For instance, the Rs could be amended to introduce individual criminal responsibility for the mere use of Aws and, therefore, exceptionally require a *dolus eventualis* –a lower threshold than intent– for accepting the risk of using Aws in critical functions. Also, superiors could be held responsible based on the possibility that they infer knowledge and intent from situations where they "became aware of the crimes, accepted them and [thus] came to intend them"[32]. A second option for the attribution of responsibility would consider Aws legal persons subject to liability. Even if this scenario seems unplausible today because robots have no agency and, therefore, cannot be accountable for their actions, this might change in the future. In this line, scholars have suggested establishing different types of agencies that distinguish between the purposes of algorithms and those of individuals who use them,

[31] For instance, see Paton Walsh, Nick (2023): "Inside the unit deploying small explosive drones in Ukraine´s cash-strapped war" in *CNN*. Available at: «*https://edition.cnn.com/2023/12/12/europe/ukraine-kherson-explosive-drones-intl/index.html*» [Consulted on 28 February, 2024].

[32] International Criminal Tribunal for the Former Yugoslavia, *Prosecutor v. Momcilo Krajisnik (Trial Judgment),* 27 September 2006: paras. 1096 ff; International Criminal Tribunal for the Former Yugoslavia, *Prosecutor v. Momcilo Krajisnik (Appeals Chamber),* 17 March 2009.

and attribute responsibility accordingly (Asaro 2016: 193-194). A third alternative which is receiving attention would envisage a hybrid system of human and machine responsibility (Acquaviva 2021: 37). When humans retain meaningful control, there would be no questioning that ICL applies to them. Yet, in the absence of meaningful human control, criminal punishment may be foregone and there might be room to consider other forms of responsibility. These could include: State responsibility for allowing the deployment of unlawful/unpredictable weapons; civil remedies against the companies who created and procured Aws, and against the individuals who programmed these weapons wrongly. Overall, the view is that the harms caused by Aws cannot go unpunished and responsibility should be assumed by the different societal actors involved.

IV. Suggestions for a domestic legal framework

Mexico´s regulatory gap dealing with Aws is inconsistent with the State´s leadership on this matter, especially in view of the escalation of violence involving explosive-equipped drones. On the one hand, Mexico´s proactiveness within the CCW submitting jointly with Brazil and Chile) a draft proposal for a legally binding treaty on Aws attests to strong views on the IHL standards that should apply to these weapons. Essentially, Mexico´s claim, also underpinning the CCW Guiding Principles, is that an ethical approach to regulate Aws is necessary to ensure respect for the principles of humanity and dignity underpinning IHL and IHRL. Accordingly, respecting these principles requires, as core tenets, retaining the meaningful human control of Aws and ensuring individual accountability for the failures of Aws resulting in IHL and IHRL violations. On the other hand, Mexico´s domestic legal system does not address the increasing use of Aws, especially in confrontations with cartels, from the ethical lens of human control and responsibility that it sponsors internationally. Even though the National Agenda on Artificial Intelligence is underpinned by an

ethical pillar that enshrines human rights and non-discrimination –especially on gender and indigenous grounds–, it is silent on the role of the military and the police as key security issues affected by AI (Casados *et al.* 2020: 47-75). Notwithstanding these inconsistent approaches, the existing policy framework contains the necessary elements to articulate a sensitive regulation of Aws and give an adequate solution to the phenomenon of armed drones. Based on the discussion of the previous section, it is argued that a domestic legal framework of Aws in Mexico should consider two relevant issues –human control and discrimination– to which the article now turns.

1. Integrating meaningful human control

In November 2023, the Congress backed a presidential proposal to regulate the use of drones for criminal purposes which, at the time of writing, is being debated by the authorities and it is likely to make its way into the Federal Penal Code and the Federal Law of Fireweapons and Explosives. The proposal criminalizes various scenarios: i) *whoever* uses drones *to throw* any explosive object or weapons as well as chemical substances over persons or goods –10 to 20 years of imprisonment, increased to 30 years if the military or security forces are affected–; ii) whoever imports, manufactures, arms or adapts drones for the transport of explosives –5 to 10 years of imprisonment–, and iii) using drones for the surveillance of public authorities to evade or attack them –3 to 10 years of prison–. For the purpose of this discussion, it is the use of drones for critical target functions that matters.

The initiative to criminalize the use of explosive drones is progressive in various regards. It sets a leading example regulating an issue which is absent in many jurisdictions around the world and for which there are no international or regional standards. Also, the proposal encompasses responsibility for a broad range of situations such as negligence, automated bias or complacency. These situations are uncovered by current ICL standards under

the Rs. This requires intent of first or second degree[33], despite scholars have considered the inclusion of lower thresholds than intent desirable in domestic regulations, given the nature of Aws (Coco 2023: 14). In addition, the presidential proposal establishes the responsibility of manufacturers and others who contribute to the process of making the use of robots for criminal purposes possible. This promotes a policy of realistic accountability and inclusiveness of the different societal actors responsible for the harms caused by Aws (Acquaviva 2020: 37; Asaro 2016: 192).

The Congress' proposal to criminalize explosive drones is, however, affected by certain weaknesses in the light of Mexico´s international obligations and policies. One shortcoming is that the initiative does *not address* the issue of human control. Hence, it is contradictory with Mexico´s position under the Gge-Ccw sponsored framework where the human control of Aws is considered the key principle of the proposed regulation[34]. The presidential initiative criminalizes anyone *who throws* –explosives from a drone–. Therefore, it merely covers situations that involve an active human subject making decisions remotely and, therefore, which involve significant human control. However, it does not address the kern of the ethical debate around Aws which focuses on the lack of human control. In view of the fact that explosive drones, Ai-powered weapons affected by the control problem, are increasingly used in confrontations with cartels, it would be reasonable to expect that this point would not be missed out in domestic regulations.

Another limitation of the presidential initiative is that it establishes a *blanket* prohibition of drone bombing which may contradict Mexico´s Ihl obligations as well as the Ccw Guiding

[33] Rs, Articles 25 and 28.

[34] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): "Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems", CCW/GGE.1/2022/WP.5.

Principles on Aws. Indeed, IHL allows the use of legal weapons between the warring parties provided that its principles are respected. Moreover, States have so far not fully banned Aws in armed conflict but, as the GGE Guiding Principles show, they are paving the way towards the regulation of Aws[35]. The proposal being discussed in Congress thus seems suitable for law enforcement situations where domestic regulations apply and attacks should be measured by *more stringent* IHRL requirements on the lethal use of force[36]. Yet, as of today, IHL obligations do not ban drone bombing if human control is retained and humanitarian principles are respected. A future domestic framework could therefore consider the possibility of distinguishing between situations where law enforcement and IHL apply as different scenarios and with different legal implications.

The above considerations allow suggesting in this article that a future regulation of explosive drones aligned with Mexico´s IHL obligations would consider prohibiting: Whoever uses or contributes to the use of drones, *with or without* meaningful human control, to throw any explosive artifacts or weapons harming persons or objects in violation of international obligations under IHRL and IHL. This formulation seems appropriate for two reasons. Firstly, it would allow attributing individual responsibility when drone bombing was carried out without meaningful human intervention, even against those highly responsible in the chain of command, not clearly covered by the present initiative. Secondly, the suggested provision would allow considering both situations of law enforcement and armed conflict, potentially applicable to armed confrontations in Mexico where the State has different legal obligations under IHL and IHRL. Accordingly, faced with a situation of drone bombing, public officials may ask: Is it a law enforcement

[35] *Ibid*.

[36] These include sufficient legal basis, necessity and proportionality. See General Assembly of the United Nations (2014): "Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns", paras. 86-92 and 136, A/HRC/26/36.

or a IHL situation? If it falls under law enforcement, a blanket prohibition on the use of drone bombing may apply –in line with the presidential initiative–. If it falls under IHL, this regime is silent about the use of Aws and the only limitations on the use of weapons require respect for the law of armed conflict. However, in line with Mexico´s foreign policy under the GGE-CCW framework, drone bombing attacks falling under IHL should be prohibited whenever IHL is violated *and* whenever the absence of meaningful human control of Aws results in IHL violations that demand accountability.

2. Addressing bias

The presidential initiative criminalizes the activities involved in the process leading to the use of explosive drones for critical functions, including the manufacturing, arming, acquisition or adaptation of drones to transport explosives. Yet, if the aim is to hold social actors responsible for the harm caused by Aws, there is still a margin of appreciation for more inclusive regulation. As explained above, an important reason why Aws make mistakes during targeting and engagement is discrimination imbedded in the process of coding. If, as explained above, bias is structural, there is considerable potential in establishing a duty of non-discrimination throughout the coding cycle which only adequate regulation can bring into effect.

Discrimination in AI is a matter of great concern and which severely affects Mexico. Not only because inequalities pre-existing confrontations are likely to exacerbate during armed hostilities (Red de Investigación Feminista en Inteligencia Artificial *et al.* 2023: 4). Also, because algorithm bias further aggravates these inequalities (Spitx et al 2023: 6-7). Discrimination in Mexico is intimately connected with organized crime. Studies reveal that AI in Mexico, and technology in general, is designed for men resulting in the aggravation of gender inequalities and human rights violations in realms such as violence against women, the labor market, education and health, among others. To halt these discriminatory processes, studies are recommending Mexican policymakers

to adopt an inclusive anti-discriminatory approach considering gender and other intersecting grounds in the coding of AI so that robots can capture the complexity of societal needs (Spitx et al 2023: 4-7; Casados *et al.* 2020: 62-63).

Mexico´s national policies recognize the *rationale* linking discrimination to violent crime. Accordingly, there is a good reason to consider this connection in a domestic regulation of AWS as well. The National Agenda on Artificial Intelligence enshrines respect for human rights and non-discrimination as an ethical axis across the agenda. It highlights the need to identify the most vulnerable groups –noting women and indigenous peoples– and discriminatory practices in order to obtain representative data for algorithm coding. Actually, the aim of the National Strategy is to use the potential of AI to revert discrimination in Mexico by means of inclusive data that improves decision-making and promotes human rights (Casado et al 2020: 60-62). This goal is also linked with the essence of the National Plan on Peace and Security. The latter recognizes that social inequalities are the root cause of violent crime, hence, it declares that the way out of violence is to promote human rights and non-discrimination that will bring peace (Chanona 2019: 6).

To benefit from the advantages of AI improving data accuracy and promoting respect for human rights in conflict, scholars have suggested the regulation of a non-discriminatory approach dealing with AWS. A duty of non-discrimination should cover those involved in the chain preceding the use of AWS and allow the attribution of criminal and civil responsibility for the damages caused by bias embedded in AWS coding (Acquaviva 2021: 37-38; Asaro 2016: 192). This approach suits well Mexico´s domestic law and policy as it would strengthen the aims of the National Agenda on Artificial Intelligence and the National Plan on Peace and Security underpinned by human rights and non-discrimination. Thus, establishing a duty of non-discrimination in a future regulation of AWS would bring in alignment domestic policies dealing with

AI and security around the promotion of equality. Furthermore, a duty of non-discrimination squarely fits the approach of Federal Law to Prevent and Eliminate Discrimination (2003) which allows liability for omissions and actions *without* the intention of discrimination but with a discriminatory effect[37]. This is the case of manufacturers, developers, traders, and others involved in the process of AI, who may act based on cultural prejudices but without intending to harm. Accordingly, a non-discriminatory approach to AWS in Mexico may consider extending liability to those contributing –even if unwillingly– to biased algorithms that result in machine failures and, thus, IHL and IHRL violations that could have been prevented with due diligence.

As outlined above, a future regulation of AWS in Mexico may provide for: A duty of non-discrimination for those involved in the process of manufacturing, developing, deploying and controlling AWS, establishing their criminal or/and civil responsibility, as corresponds. In this sense, faced with bias dealing with military applications of AWS, public officials may ask questions such as (Chandler 2021: 23; Houghton 2022: 3-4): Was the attack committed due to discrimination or did it result in discrimination on one or intersecting grounds due to bias in the coding of AWS? Was the participation and representation of vulnerable groups –women and children, indigenous, etc.– considered in the processes of algorithm coding and decision-making? Was conflict prevention considered in the system design? Did the model used consider the protection of human rights, including reparations? Did these protections apply equally to women, girls, men and boys without discrimination? If ending armed confrontations in Mexico largely depends on promoting human rights and equality, ensuring that AWS data is not discriminatory becomes a critical ethical and legal component of the national strategy.

[37] Ley Federal para Prevenir y Eliminar la Discriminación, publicada en el Diario Oficial de la Federación el 11 de junio de 2003. Article 1.

V. Conclusion

This article addressed the legal gap on the use of Aws in armed confrontations in Mexico, especially drone bombing, dealing with organized criminal groups. It argued the need for an IHL approach in certain instances of violence involving Aws which is sensitive to the issues of human control, discrimination and individual criminal responsibility. Two issues which are hampering the application of this perspective were identified. Firstly, Mexico´s armed hostilities are addressed from a law enforcement perspective due to the absence of political will to apply IHL obligations. Secondly, there is a lack of international standards on Aws due to ethical concerns around the issues of human control and international criminal responsibility for the violations caused by Aws. These issues raised a major concern: If Aws have autonomy to carry out the critical functions of targeting and engagement, is there room for human accountability? It was argued that, as a key supporter crafting the Guiding Principles in the framework of the CCW, Mexico holds a strong international position on the IHL standards that should apply to Aws. These are underpinned by the principles of significant human control and individual criminal responsibility. Accordingly, the same standards should underpin Mexico´s domestic legal framework to address the increasing escalation of violence involving drone bombing by powerful criminal groups.

Based on a discussion of the key legal issues posed by Aws, the article suggested the tenets of a domestic legal framework aligned with Mexico´s IHL obligations and national policies underpinned by non-discrimination and human rights. First, it was suggested that a domestic framework should expressly address the control problem establishing the illegality of Aws without meaningful human control as well as those Aws that violate IHL. A more blanket criminalization –i.e. the 2023 presidential initiative on drone bombing– would be suitable for law enforcement situations which fail to reach the threshold of armed conflict. Second, it was recommended that a domestic legal framework of Aws

should establish a duty of non-discrimination for those involved in putting Aws to work –from developers to the military– holding them responsible for machine bias resulting in IHL and IHRL violations. This broad form of social accountability is necessary for prevention purposes and to ensure reparations. Moreover, a duty of non-discrimination regarding Aws is consistent with and enhances the implementation of core Mexican policies in relation with AI and security, namely, the National Agenda on Artificial Intelligence and the National Plan on Peace and Security, both underpinned by human rights and non-discrimination standards.

Protecting human control and non-discrimination are the core ethical reasons for which establishing a regulatory framework of Aws is critical for Mexico and for any State. Meaningful human control ensures that human agency –judgment, not automation– is behind targeting and use of force. Therefore, it involves a progressive choice to promote the rights to dignity –decisions affecting human life should be taken by humans, not by robots– and reparations in armed confrontations. As for the duty of non-discrimination in the coding of Aws, this standard shows commitment to put technology to work to improve human rights in armed conflict, promoting a culture of peace. Mexico´s support for these standards is commendable. Establishing a domestic regulation along these lines can set a compelling example to advance this urgent topic in armed confrontations where the autonomy of weapons is becoming increasingly harmful.

Bibliography

Acquaviva, Guido (2021): "Autonomous weapons systems controlled by Artificial Intelligence: a conceptual roadmap for international criminal responsibility", in *SSRN*. Available at «*https://ssrn.com/abstract=4070447*» [Consulted 31 January, 2024].

Amoroso, Daniele and Tamburini, Guglielmo (2020): "Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues", in *Current Robotics Report,* Vol. 1, 187-194.

Arkin Ronald C. (2009): *"Governing Lethal Behavior in Autonomous Robots",* Chapman and Hall/CRC, New York.

Asaro, Peter M. (2016): "The Liability Problem for Autonomous Artificial Agents", in *Association for the Advancement of Artificial Intelligence,* Spring Symposium, Num. 4, 190-194.

Bellal, Annyssa (2017): "The War Report: Armed Conflicts in 2017" in *Geneva Academy.* Available at: «*https://www.geneva-academy.ch/joomlatools-files/docman files/The%20War%20Report%202017.pdf*» [Consulted on 23 February, 2024].

Bunker, Robert J. and Sullivan, John P. (2021): "Mexican Cartels are Embracing Aerial Drones and They´re Spreading" in *War on the Rocks. Available* at: «*https://warontherocks.com/2021/11/mexican-cartels-are-embracing-aerial-drones-and-theyre-spreading/*» [Consulted on 31 January, 2024].

Casados, Daniel *et al.* (2020): "Agenda Nacional Mexicana de Inteligencia Artificial", in *Coalición IA2030Mx.* Available at: «*https://www.ia2030.mx/agenda2020*» [Consulted on 27 February, 2024].

Cbs (2023): "Drug cartels are sharply increasing use of bomb-dropping drones, Mexican army says", in *Cbs News*. Available at: «*https://www.cbsnews.com/news/drug-cartels-more-bomb-dropping-drones-mexico-army/*» [Consulted on 01 February, 2024].

Chandler Katherine (2021): "Does Military AI Have Gender? Understanding Bias and Promoting Ethical Approaches in Military Applications of Ai", in *UNDIR Research*

Report. Available at: «*https://unidir.org/publication/does-military-ai-have-gender-understanding-bias-and-promoting-ethical-approaches-in-military-applications-of-ai/*» [Consulted on 27 February, 2024].

Chanona, Alejandro (2019): "México: la reforma del sector seguridad y la estrategia contra la delincuencia organizada de Andrés Manuel López Obrador", in *Real Instituto Elcano ARI/3/2018.* Available at: « *https://media.realinstitutoelcano.org/wp-content/uploads/2021/11/ari3-2018-chanona-mexico-reforma-seguridad-estrategia-contra-delincuencia-organizada-lopez-obrador.pdf*» *[*Consulted on 05 February, 2024].

Coco, Antonio (2023): "Exploring the Impact of Automation Bias and Complacency on Individual Criminal Responsibility for War Crimes", in *Journal of International Criminal Justice,* Vol. 21, Issue 05, 1077–1096.

D´Cunha, Samit *et al*. (2024): "Defining armed conflict: some clarity in the fog of war", in *Humanitarian Law & Policy.* Available on: «*https://blogs.icrc.org/law-and-policy/2024/05/02/defining-armed-conflict-some-clarity-in-the-fog-of-war/*» [Consulted on 23 May, 2024].

Farrés Jiménez, Andrea (2021): "Embedding Gender in International Humanitarian Law: Is Artificial Intelligence Up to the Task?", in *Just Security.* Available at: «*https://www.justsecurity.org/77970/embedding-gender-in-international-humanitarian-law-is-artificial-intelligence-up-to-the-task/*» [Consulted on 27 February, 2024].

González, Isabella (2023): "El Ejército mexicano incrementa su poderío militar: escala 11 posiciones en el ranking mundial de las Fuerzas Armadas", in *El País.* Available at: «*https://elpais.com/mexico/2023-03-15/el-ejercito-mexicano-incrementa-su-poderio-militar-escala-11-posi-*

ciones-en-el-ranking-mundial-de-fuerzas-armadas.html» [Consulted on 27 February, 2024].

Håkansson Schmidt, Alexandra (2022): "Action Brief-From Warfare to Peace building: Employing Artificial Intelligence for Women, Peace and Security", in *Un Women*. Available at: «*https://asiapacific.unwomen.org/en/digital-library/publications/2022/04/from-warfare-to-peacebuilding*» [Consulted 13 February, 2024].

Houghton, Isabelle (2022): "The Implications of Artificial Intelligence in Women, Peace and Security: Part 2", in *Our Secure Future*. Available at: «*https://oursecurefuture.org/our-secure-future/news/implications-artificial-intelligence-women-peace-and-security-part-2*» [Consulted on 27 February, 2024].

Icj (2004): *"Advisory opinion of the International Court of Justice on the Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory"*, in *United Nations*. Available at: «*https://www.un.org/unispal/document/auto-insert-178825/*» [Consulted on 28 February, 2024].

Icrc (2023): "What you need to know about artificial intelligence in armed conflict" in *International Committee of the Red Cross*, United Kingdom. Available at: «*https://www.icrc.org/en/document/what-you-need-know-about-artificial-intelligence-armed-conflict*» [Consulted on 20 May, 2024].

Icrc (2021): "Position on Autonomous Weapon Systems: Icrc Position and Background Paper" in *International Review of the Red Cross,* Geneva, 1-12.

Icrc (2005): "International Humanitarian Law Databases, Rules", in *Icrc*. Available at: «*https://ihl-databases.icrc.org/en/customary-ihl/v1*» [Consulted on 20 February, 2024].

Icrc (1987): "Commentary of 1987, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the

Protection of Victims of International Armed Conflict (Protocol I), 8 June 1977", in *Icrc*. Available at: «*https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-36/commentary/1987?activeTab=undefined*» [Consulted on 28 February, 2024].

Informador.Mx (2024): "Todo sobre la emboscada del CNG que dejó 3 lesionados al sur de Jalisco", in *Informador.Mx*. Available at: «*https://www.informador.mx/jalisco/CJNG-TODO-sobre-la-emboscada-que-dejo-3-lesionados-al-sur-de-Jalisco-20240426-0065.html*» [Consulted on 29 May, 2024].

Janetsky, Megan (2024): "A drug cartel has attacked a remot Mexican community with drones and gunmen, rights group says", in *Associated Press*. Available at: «*https://apnews.com/article/mexico-cartel-attack-guerrero-drones-helidoro-castillo-ed795770ce9a09b1c063d91913268b99*» [Consulted on 29 May, 2024].

López Obrador Andrés Manuel (2018): "Plan Nacional de Paz y Seguridad 2018-2024", in *AMLO* Available at: «*https://lopezobrador.org.mx/wp-content/uploads/2018/11/PLAN-DE-PAZ-Y-SEGURIDAD_ANEXO.pdf*» *[*Consulted on 15 February, 2024].

Mainzer, Klaus (2020): *Artificial Intelligence: When do Machines Take Over?,* Springer, Berlin.

Mexico Now (2023): "Congress approves penalties for drone misuse", in *Mexico Now*. Available at: «*https://mexico-now.com/congress-approves-penalties-for-drone-misuse/*» [Consulted on 28 February, 2024].

Mondragón Toledo, Gabriel (2023): "NarcoDrones Have Become a Growing Scare Tactic in Mexico´s Drug Wars", in *Inkstick*. Available at: «*https://inkstickmedia.com/narcodrones-*

have-become-a-growing-scare-tactic-in-mexicos-drug-wars/» [Consulted on 01 February, 2024].

Muggah, Robert (2023): "Organized crime in armed conflicts and other situations of violence", in *International Review of the Red Cross,* Vol. 105, Num. 923, 569-574.

Padin, Juan Francisco (2023): "Opening Pandora´s box: The case of Mexico and the threshold of non-international armed conflicts", in *International Review of the Red Cross*, Vol. 106, Num. 923, 772-794.

Paton Walsh, Nick (2023): "Inside the unit deploying small explosive drones in Ukraine´s cash-strapped war", in *Cnn*. Available at: «*https://edition.cnn.com/2023/12/12/europe/ukraine-kherson-explosive-drones-intl/index.html*» [Consulted on 28 February, 2024].

Plaw, Avery *et al.* (2020): "The process of technologization of the drug war in Mexico", in *The Politics of Technology in Latin America* (Volume 1), Plaw Avery *et al.* (eds.), Taylor & Francis, United Kingdom, 95-108.

Porup, J.M. (2017), "Hunting for Mexico´s mass graves with machine learning", in *BIZ & IT*. Available at: «*https://arstechnica.com/information-technology/2017/04/hunting-for-mexicos-mass-graves-with-machine-learning/*» [Consulted on 28 February, 2024].

Red de Investigación Feminista en Inteligencia *Artificial et al* (2023): "Aportes para el llamado de la Anudh sobre los desafíos de la tecnologías nuevas y emergentes en el ámbito militar", in Ohchr. Available at «*https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/advisorycommittee/techmilitarydomain/submissions/3-academia-red-investigacion-feminista-ia.pdf*» [Consulted on 27 February, 2024].

Rulac (2022): "Non-International Armed Conflicts in Mexico", Geneva Academy of International Humanitarian Law and Human Rights, Geneva, 30 May 2021. Available at «*https://www.rulac.org/browse/conflicts/non-international-armed-conflict-in-mexico*» [Consulted on 28 February, 2024].

Salgado Espinoza, Raúl (2020): "Ethical controversies about Lethal Autonomous Weapons Systems, Views of small South American states" in *The Politics of Technology in Latin America (Volume 1),* Plaw Avery *et al.* (eds.), Routledge, London, 65-79.

Sharkey Noel E (2012): "The evitability of autonomous robot warfare", in *International Review of the Red Cross*. Available at: «*https://international-review.icrc.org/articles/evitability-autonomous-robot-warfare*» [Consulted on 27 February, 2024].

Shuldiner, Henry, (2022): "Military-Grade Weapons in Michoacán Put Mexican Forces on War Footing", in *Insight Crime*. Available at: «*https://insightcrime.org/news/war-type-weapons-in-michoacan-put-mexican-forces-on-battle-footing/*» [Consulted on 05 February, 2024].

Spitx, Sarah *et al.* (2023): "A Gender Perspective on Artificial Intelligence (Ai) and the Tech-Sector. Position Paper for the Multi-Stakeholder Dialogue in Mexico", in *Alexander von Humboldt Institute for Internet and Society.* Available at: «*https://www.hiig.de/publication/gender-ai-techsector-mexico/*» [Consulted on 27 February, 2024].

Stevenson, Mark (2021): "Mexico´s drug war uses drones, human shields, gunships", in *AP*. Available at: «*https://apnews.com/article/business-mexico-army-gangs-f63d6aac1d8057c8024ae7688c465696*» [Consulted on 01 February, 2024].

Sullivan, John P (2021): "Non-International Armed Conflict: Mexico and Colombia", in *Revista do Ministério Público Militar,* Vol. XLVI, Num. 35. Available at: «*www.academia.edu/62379204/Non_International_Armed_Conflict_Mexico_and_Colombia*» [Consulted on 05 February, 2024].

Wola (2022): "México profundiza la militarización. Los hechos muestran que es una estrategia fallida" in *Wola*. Available at: «*https://www.wola.org/es/analisis/mexico-profundiza-militarizacion-hechos-muestran-estrategia-fallida/*» [Consulted on 16 February, 2024].