Section I: Contextualization

In this section the author will provide some contextualization of the AI and AWS debate by offering some conceptual guidance on the terms and technologies subject matter to this article. Please bear in mind that this is a *legal* article and does not intend to be authoritative on the scientific or purely technical aspects of these terms, for most of them are not even conclusively defined by the leading experts in these fields, and because the exponential rate at which they evolve renders them to be of the most fluid nature.

Let us begin with a semantic examination of the relevant notions for this study:

The English word *artificial,* derived from the Latin *artificialis,* is equivalent to the terms factitious, synthetic, fake, unnatural - a thing that is artificial is man-made or constructed by humans, usually to appear like a thing that is natural.[3]

On the other hand, the word *intelligence* is more difficult to define. Intelligence is broadly explained as 'the ability to learn, understand, and make judgments or have opinions that are based on reason' or simply as 'thinking ability'.[4] However, this notion is still contest-

[3] 'Artificial, adj', *Cambridge Dictionary,* available at: *<https://dictionary.cambridge.org/us/dictionary/english/artificial>* accessed 2 July 2020.

[4] 'Intelligence, n', *Cambridge Dictionary,* available at: *<https://dictionary.cambridge.org/us/dictionary/english/intelligence>* accessed 2 July 2020.

ed amongst psychologists as some of them relate it to the human intellect and thus limited to the cognitive brain. Therefore, this notion would traditionally be linked to the *human* condition.

For Stephen Hawking, "intelligence is the ability to adapt to change".[5] This quote serves us as a good bridge for the next -*compound*- concept, Artificial intelligence.

There is no consensus on a universal definition of this concept as it is continuously challenged and refined by amongst the leading experts in the field.[6] The only definitive and agreed upon aspect about AI is that it is a *very* disruptive technology.

As a research field, the name Artificial Intelligence was decided upon during a workshop at Dartmouth College in 1956, where a group of remarkable scientists gathered for an 8-week brainstorming session on the conception of "Machines that think".[7]

---

5 Stephen Hawking stated this at his graduation from Oxford University, as cited in: The Telegraph, *Professor Stephen Hawking: 13 of his most inspirational quotes*, London, (8 January 2016) available at: <*www.telegraph.co.uk/news/science/stephen-hawking/12088816/Professor-Stephen-Hawking-13-of-his-most-inspirational-quotes.html*> accessed 2 July 2020.

6 Matthew U. Scherer, *Regulating Artificial Intelligence Systems, Risks, Challenges, Competencies and Strategies*, 29 Harv. J. L. & Tech, at 353, 359-62 (2016) [*hereinafter:* Scherer, *Regulating AI*].

7 Giovanni Sileno, *History of AI, Current Trends, Prospective Trajectories,* Winter Academy on Artificial Intelligence and International Law, Asser Institute (2021); Sileno mentions the group of ~20 remarkable scientist and engineers, who were at the Dartmouth workshop in 1956 including: John McCarty (LISP language, situation calculus, non-monotonic logics) Marvin Minsky (frames, perceptron, society of minds), Herbert Simon (logic theorist, general problem solver, bounded rationality), Allen Newell (logic theorist, general problem solver, the knowledge level), Ray Solomonoff (father of algorithmic probability, algorithmic information theory), Arthur Lee Samuel (first machine learning algorithm for checkers), W. Ross Ashby (pioneer in cybernetics, law of requisite variety), Claude Shannon (father of information theory) and John Nash (father of game theory) [*hereinafter:* Sileno, *History of AI*].

Broadly speaking, it can be understood to be the use of computer systems to carry out tasks previously requiring human intelligence, cognition or reasoning.[8] It is a category of research meant to develop systems that are able to solve problems or achieve goals in different degrees of difficulty by reasoning, *i.e.*, by imitating human problem-solving abilities, in some cases including the ability to learn from experience and therefore improve the machine's abilities without any human intervention,[9] and that is designed to act as a rational agent.[10]

In this respect, the Organization for Economic Cooperation for Development (OECD) has dictated five basic principles for regulating AI —*lato sensu*— in a general agreement document signed by 42 Member States:[11]

- AI should benefit people and the planet, driving inclusive growth, sustainable development and well-being.
- AI systems should be designed with respect for the law, human rights, democratic values and diversity, as well as including safeguards that allow human intervention.
- AI systems should be transparent, and there must be a clear understanding of how they work.
- AI must operate in a stable and secure manner throughout their existence and that the potential risks can be assessed continuously.

---

[8] 'Artificial Intelligence, n', *Cambridge Dictionary*, available at: *<https://dictionary.cambridge.org/us/dictionary/english/artificial-intelligence>* accessed 2 July 2020.

[9] Igor Kononenko & Matjaz Kukar, *Machine Learning and Data Mining: Introduction to Principles and Algorithms,* at 38 (2007); *see* also Scherer, *Regulating AI*, *supra* note 6, at 361.

[10] Stuart Russell & Peter Norvig, *Artificial Intelligence - A Modern Approach*, at 4-5 (3rd ed. 2010) [*hereinafter:* Russell & Norvig, *AI-A Modern Approach]*.

[11] Organization for Economic Cooperation and Development (OECD), *OECD principles on AI*, available at: *<https://www.oecd.org/going-digital/ai/principles/>* accessed 5 July 2020.

– It is required that organizations and individuals that develop, distribute or operate AI systems are responsible for the proper functioning in line with the above-mentioned principles.

Nonetheless, the potential of this technology is very wide in scope, therefore it is necessary to make a distinction between the different types of AI:[12]

– Artificial narrow intelligence (ANI) has a narrow range of abilities and is the AI that is most prevalent in our world today. It is programmed to perform a specific task extremely well.
– Artificial general intelligence (AGI) is on par with human capabilities. This technology hasn't been achieved – yet. The purpose of AGI is to think, understand, and act in a way that is indistinguishable from that of a human in any given situation. Noticeably, the aim of many projects related to AI is that these systems can adapt to different situations and operate without human control. What scientists are missing at this point is a way to make machines conscious by programming a full set of cognitive abilities that until now are known only to humans.
– Artificial superintelligence (ASI) is even more capable than a human. It is intended to outperform our abilities and thus surpass the limitations of our species. It is super-powerful and self-aware beyond the human sense. This embodies the evolution of this field, on which the theory of "singularity" is based, meaning that one day machines will be smart enough to program and improve themselves until they become independent from their human creators.

On the one hand, some technoskeptics believe that this scenario is far-fetched, on the other, Google's inventor, Ray Kurzweil's predicts that the "singularity" will occur in the year

[12] Brodie O'Carroll, *What are the 3 types of AI? A Guide to Narrow, General, and Super Artificial Intelligence,* Codebots (2017) available at: *<https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible>* accessed 5 July 2020.

2045 with the creation of a self-conscious AI that will be one billion times more powerful than all human brains.[13]

The avenue to achieve this is through a technology called machine learning, a concept that was already introduced in the 1940s by mathematician Alan Turing and developed through his *Imitation Game* which defined an operational standard for intelligence, known as the "Turing Test",[14] which became the basis of the AI we currently know.[15]

In essence, machine learning is a process that enables artificial systems to improve from experience,[16] enables machines to adapt to new environments, and to act in a manner that will result in achieving the assigned goal regardless of unforeseen obstacles and with no explicit direction from their programmer.[17]

Ideally, machine learning would become a solution in order to more efficiently, effectively, and accurately tackle unpredictability, without receiving orders from the programmer.[18]

---

[13] AI Business, *Ray Kurzweil Predicts that the singularity will take place in 2045,* (2017) available at: <*https://aibusiness.com/document.asp?doc_id=760200*> accessed 5 July 2020.

[14] Russell & Norvig, *AI-A Modern Approach, supra note* 10, at 16-17, 1021.

[15] *Id.,* at 16-17.

[16] Sileno, *History of AI, supra* note 7.

[17] Russell & Norvig, *AI-A Modern Approach, supra* note 10, at 693; Carry Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, SSRN, Georgetown Law Journal, at 1156 (2017) [*hereinafter:* Coglianese & Lehr, *Regulating by Robot]*; Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. Rev. at 871, 875 (2015-2016) [*hereinafter:* Rich, *Machine Learning]*.

[18] Russell & Norvig, *AI-A Modern Approach, supra* note 10*,* at 693; Coglianese & Lehr, *Regulating by Robot, supra* note 17, at 1156; Rich, *Machine Learning, supra* note 17, at 875.

On the other hand, deep learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.[19] It relies on a hierarchy of representation learning, producing different levels of abstraction.[20] Basically, it is an expansion of machine learning onto multiplied layers, thereby assimilating an exponential amount of data.

These technologies have a remarkable asymmetrical advantage over humans given that they can accrue knowledge from potentially infinite databases, and in turn leave us with a very limited understanding of their capabilities.

Accordingly, there is a *huge* caveat derived from machine learning and deep learning that must be properly taken into consideration, one that could be referred to as a "known unknown". This is no other than the *decision-making process* a.k.a., "the black box". The reason for this is that, unlike human linear decision-making processes, artificial decision-making processes are too perplexing for us to understand because the machine itself creates its own algorithms, and as such, these processes constitute the black box.[21]

This means that regardless of its original "set up" it is the program itself that decides upon the proper weight to ascribe to each element it perceives.[22]

---

[19] Jason Brownlee, *What is deep learning?,* (2019) available at: *<https://machinelearningmastery.com/what-is-deep-learning/>* accessed 4 February 2021.

[20] Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning,* MIT Press (2016).

[21] W. Nicholson II Price, *Black-Box Medicine*, 28 Harvard Journal of Law & Technology at 419, 432-34 (2014-2015); Rich, *Machine Learning, supra* note 17, at 886 [*hereinafter:* Price, *Black-Box Medicine]*.

[22] Coglianese & Lehr, *Regulating by Robot, supra* note 17, at 1156. They say, "we cannot really know what precise characteristics any machine-learning algorithm is keying in on".

In addition, the programmer does not know what rule, or even which specific characteristics, were utilized by the machine in yielding a certain conclusion, nor can the programmer deconstruct the inferences or trackback the decision processes that were applied.[23]

To illustrate, a utility function that is programmed to mitigate or avoid human suffering can decide to kill instead of injuring a person – since people do not suffer when they are dead.[24]

In other words, the programmer controls the input introduced to the program in its learning phase, she provides optimization guidelines for the interpretation of these inputs (what is known as the utility function) and is privy to the output the program extrapolated - but for all other purposes, the artificial entity is considered to be a black box that yields no intuitive nor causal explanation for its actions.[25]

In sum, the use of machine learning and deep learning programs must always conform with the full awareness that it entails the inherent risk that there is no way to predict, understand or audit a specific decision concluded by the AI in terms that are understandable to humans.[26]

---

[23] *Id.*

[24] For more examples *see* Ryan Calo, *Robotics and the Lessons of Cyberlaw, 103 California Law Review* at 542-43 (2015) [*hereinafter:* Calo, *Robotics and the Lessons of Cyberlaw*].

[25] Liron Shilo, *When Turing Met Grotius AI, Indeterminism, and Responsibility,* SSRN, at 14 (2018) [*hereinafter:* Shilo, *When Turing Met Grotius]*.

[26] *Id.,* at 11-12, 18-19. For more on the Black-Box *see* Price, *Black-Box Medicine, supra* note 21, at 432-437; 442-467. He explains the pros and cons of black-boxes in the context of medicine; Rich, *Machine Learning, supra* note 17, at 886, 923-24. He describes the trade-off between conceding to the use of algorithms that have black boxes but are highly accurate in their predictions, and concluding that despite the favorable accuracy in prediction, the black boxes should be more transparent for the sake of analyzing the 4th amendment implication of the use of such algorithms.

In order to digest these concepts, the following quote comes to mind:

> As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns— the ones we don't know we don't know.[27]

Autonomous Weapon Systems (AWS)

It is notable that the underlying technology of Artificial Intelligence has the potential to accommodate both civilian and military uses. This section focuses on the second domain which applications could include intelligence, surveillance and reconnaissance (ISR), navigation, multi-domain command and control, missile defense, cyber defense, information manipulation, target recognition and weapons development.[28]

The latter weapon systems are completely *sui generis* and give birth to a *de jure* and *de facto* category of their own.

The incumbent are highly sophisticated *entities* that are able to mimic human decision-making abilities in order to execute a variety of tasks *without* any human intervention.[29]

---

[27] Response from Donald Rumsfeld to a question from the US Department of Defense during a press conference in February 2002, available at: <*https://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636*> accessed 7 July 2020.

[28] Hitoshi Nasu, *Artificial Intelligence and the Obligation to Respect and to Ensure Respect for International Humanitarian Law,* Exeter Centre for International Law, at 5 (2019).

[29] *See* Michael N. Schmitt, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, Harvard National Security Journal Features, at 4 (2013) [*hereinafter:* Schmitt, *AWS and IHL*]; Kenneth Anderson et al., *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, 90 International Legal Studies,

Also known as lethal autonomous weapons systems (LAWS), these entities are designed to *actively initiate* and make *lethal* decisions rather than merely acting as defensive and/or reactive systems.[30]

As such, this category of entities is not currently defined in our legal order given the predominant and novel feature of their autonomous nature on the cognitive and decision-making levels. That is why, it is paramount to remember that although they may be weaponized, they should not be examined or defined merely as weapons, as they are much more than that.

Upon careful observation, one realizes that these entities are neither weapons, conventional platforms, nor moral agents tantamount to humans for legal purposes. That being said, they are often referred to as the first, on occasion as the second, and frequently treated as the third.[31]

Naturally this is another concept that lacks definitional consensus within the field, and which discretional margin ranges enormously. On one end of the spectrum, an AWS is considered as an automated component of an existing weapon, and on the other, as a platform that is itself capable of sensing, learning, and launching attacks.[32]

---

at 386 (2014); *see* Philip Alston, *Interim report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, 18, U.N. Human R. Comm., U.N. Doc. A/65/321 (23 August 2010); Ian Kerr and Katie Szilagyi, *Asleep at the switch? How Killer Robots Become a Force Multiplier of Military Necessity, 333* (2016); Orna Ben Naftali and Zvi Triger, *The Human Conditioning: International Law and Science Fiction*, 14 Law, Culture And The Humanities, at 38 (2016).

30 Shilo, *When Turing Met Grotius, supra* note 25, at 2.

31 *Id.*, at 15.

32 Dustin A. Lewis, Gabriella Blum & Naz K. Modirzadeh, *War-Algorithm Accountability,* HLS PILAC, at 5 (2016) [*hereinafter:* Lewis et al., *War-Algorithm]*.

Relatedly, and in an effort to simplify —yet broaden— the concept, Harvard University scholars have introduced the term "war algorithms", defined as any algorithm that is expressed in computer code, that is effectuated through a constructed system, and that is capable of operating in the context of armed conflict. These systems include self-learning architectures that are at the center of the most heated debates about the perceived replacement of human judgment with algorithmically-derived choices.[33]

As explained above, the fact that these choices might be difficult for humans to anticipate or unpack *vis-á-vis* the prospect of them being autonomous as well as able to physically act upon the world, definitely confronts the concepts of law as we know them.[34]

Consequently, they challenge fundamental and interrelated notions of public international law, international humanitarian law, international criminal law and related accountability frameworks. Those concepts include attribution, control, foreseeability, and reconstructability.[35]

Understandably, the American scientist Max Tegmark calls the shift into autonomy "the third revolution of weapons", after the invention of gunpowder in the thirteenth century and that of nuclear arms in the twentieth century.[36]

---

[33] *Id.*, at 10.

[34] Calo, *Robotics and the Lessons of Cyberlaw, supra* note 24, at 542.

[35] Lewis et al., *War-Algorithm*, *supra* note 32, at 77.

[36] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence,* New York (2017).

Autonomy

At this point it is important to zoom in and focus on two key aspects of this new category – autonomy and decision-making.

Given the fact that our societies are increasingly more interconnected every day, it is relatively simple to carry out not only traditional cyberattacks, but cyber kinetic attacks. These are virtual assaults that have tangible consequences on the physical world resulting in the causation of damage, injury or death solely through the exploitation of vulnerable information systems and processes. Basically, this means using the cyberspace to inflict physical damage on nuclear power plants, water facilities, oil pipelines, factories, hospitals, banks, transit systems and apartment structures.[37] The attacker can be located in a safe, far removed location while remotely taking a toll on human lives and destabilizing national or foreign governments by targeting critical infrastructure - all that is required is an internet connection.

Albeit how dangerous this notion is on its face alone, it is noteworthy that this is known to be a human controlled, operated and directed kind of warfare - *in principle*.

Yet, what we currently conceive as battlespace[38] is becoming gradually, but progressively, human-free, both in practice and in theory. It is easy to understand the reason for this since machines have an advantage over humans given the fact that they can execute tasks quicker, more precisely and cheaper than if they were performed by us.

Theoretically, us humans are still in charge of *when* to start a war, against *whom* it will be fought, (*jus ad bellum*) which *weapons, means,* and *methods* would be used, and what

[37] Naveen Goud, *What is a Cyber Kinetic Attack?* available at: <*https://www.cybersecurity-insiders.com/what-is-a-cyber-kinetic-attack/*> accessed 20 July 2020.

[38] Formerly known as 'battlefield', as will be explained in Section II: International Humanitarian Law Considerations.

objectives are to be achieved (*jus in bello*). However, on a tactical level, machines are becoming the "micro-managers" and the executors of *how* to achieve these goals. They will increasingly decide who, what and when to attack. In a nutshell, this is what's known as battlespace automation.[39]

The question we must ask ourselves at this point is – can we allow for these decisions to be legally (*and ethically*) delegated to machines?

Moreover, it is only a matter of time before this battlespace automation makes its way into the next level – becoming autonomous.

The International Committee of the Red Cross (ICRC) is the humanitarian organization mandated to safeguard the Geneva Conventions, and it has stated that it is not opposed to new technologies of warfare *per se*.[40]

In all cases, any new technology of warfare must be used, and must be capable of being used, in compliance with existing rules of international humanitarian law - this is a minimum requirement. Nonetheless, the unique characteristics of new technologies of warfare such as AWS, the intended and expected circumstances of their use, and their foreseeable humanitarian consequences raise questions of whether existing rules are sufficient or need to be clarified or supplemented, in light of their foreseeable impact.[41]

---

[39] Shilo, *When Turing Met Grotius, supra* note 25, at 1.

[40] International Committee of the Red Cross (ICRC), *Artificial intelligence and machine learning in armed conflict: A human-centered approach,* Geneva, at 1 (6 June 2019) [*hereinafter:* ICRC, *AI and machine learning in armed conflict].*

[41] ICRC, *International Humanitarian Law and the Challenges of Contemporary Armed Conflicts,* report for the 32nd International Conference of the Red Cross and Red Crescent, Geneva, at 38-47 (October 2015).

Granted, certain military technologies – such as those enabling greater precision in attacks – may assist conflict parties in minimizing the humanitarian consequences of war. However, as with any new technology of warfare their lawfulness depends on the way they are used in practice.

For the ICRC, a significant application is the use of AI and machine learning tools to control physical military hardware, in particular, the increasing number of unmanned robotic systems – in the air, on land and at sea – with a wide-range of sizes and functions. AI and machine learning may enable increasing autonomy in these robotic platforms, whether armed or unarmed, and controlling the whole system or in specific functions – such as flight, navigation, surveillance or even targeting.[42]

The organization is also interested in the application of AI and machine learning to the development of cyber weapons as "digital autonomous weapons", for they are expected to change the nature of both, the cyberdefense and cyberattack capabilities, increasing the scale, and changing the nature and severity of attacks.[43]

Naturally, a recurrent concern amongst scholars is that their autonomous nature creates a responsibility gap[44] – mainly derived from the black box – which would negate any moral culpability of human actors who took part in their creation, utilization or deployment. It is clear that the questions regarding attribution of responsibility are paramount because the fact that

---

[42] ICRC, *AI and machine learning in armed conflict, supra* note 40, at 3.

[43] Brundage, M. et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, (2018); United Nations Institute for Disarmament Research (UNIDIR), *The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapon Systems and Cyber Operations*, UNIDIR, 2017.

[44] Davison, N., *Autonomous weapon systems under international humanitarian law*, in *Perspectives on Lethal Autonomous Weapon Systems,* United Nations Office for Disarmament Affairs (UNODA) Occasional Papers No. 30, at 16 (2016) [*hereinafter:* Davison, *AWS under IHL*].

someone may be held responsible for deviating from the agreed upon rules, is a *sine qua non* condition of fighting a just war.[45]

Decision Makers?

Perhaps the broadest and most far-reaching application is the use of AI and machine learning for decision-making purposes, by enabling widespread collection and analysis of data sources to identify people or objects, assessment of patterns of life or behavior, making recommendations for military strategy and operations, or making predictions about future actions or situations.[46]

These automated decision-making systems are effectively an expansion of intelligence, surveillance and reconnaissance tools, using AI and machine learning to automate the analysis of large data sets to provide "advice" to humans in the making of particular decisions, and increasingly, to automate both the analysis and the subsequent initiation of a decision and/or action by the system.[47] Relevant AI and machine-learning applications include pattern recognition, natural language processing and image, facial and behavior recognition. The ICRC's concern revolves around the fact that the possible use of these systems is extremely broad and may include decisions about who or what to attack and when,[48] about who to detain and for

---

[45] *See* Robert Sparrow, *Killer Robots*, 24 Journal of Applied Philosophy, at 62, 67 (2007) [*hereinafter:* Sparrow, *Killer Robots*].

[46] ICRC, *AI and machine learning in armed conflict, supra* note 40, at 4.

[47] *Id.*

[48] USA, *Implementing International Humanitarian Law in the Use of Autonomy in Weapon Systems*, Working Paper, *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* (CCW) Group of Governmental Experts, March 2019.

how long,[49] about military strategy – even on the use of nuclear weapons[50] – as well as those regarding specific operations, such as attempts to predict or pre-empt adversaries.[51]

It is precisely this concern that has led to copious heated arguments about the role of decision-making in war, and who is better situated to make life-and-death decisions—humans or machines. However, this is a nearly theological question because in the case of a machine, it is not technically clear that it can always comply with IHL or the rules it was programmed to follow, whereas a human can deliberately decide not to respect IHL.[52]

In any case, there is also significant disagreement over the cost/benefit analysis that might result from distancing human combatants from the battlefield and whether the potential life-saving benefits of AWS are outweighed by the risks inherent to the fact that war also becomes, in a practical sense, easier to conduct[53] and thus leading to an increased vulnerability of civilian populations.

Proponents of this technology argue that AI and machine learning-based decision-support systems may enable better decisions by humans in the manner in which they conduct hostilities and that these will comply with international humanitarian law. They say this

---

[49] Ashley Deeks, *Predicting Enemies*, Virginia Public Law and Legal Theory Research Paper No. 2018-21, at 1549-1554 (2018).

[50] Boulanin, V., (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*. Vol. 1, Euro-Atlantic Perspectives, Stockholm International Peace Research Institute (SIPRI) (2019).

[51] Hill, S., and Marsan, N., *Artificial Intelligence and Accountability: A Multinational Legal Perspective,* Big Data and Artificial Intelligence for Military Decision Making, Meeting Proceedings STO-MP-IST-160, NATO, (2018).

[52] Marco Sassóli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified,* 90 International Law Studies, U.S. Naval War College, at 310 (2014) *[hereinafter:* Sassóli, *Autonomous Weapons and IHL]*.

[53] Lewis et al., *War-Algorithm*, *supra* note 32, at 8.

will consequently minimize risks for civilians by facilitating faster and more widespread collection and analysis of information.

However, the ICRC is rightly reminded of the black box problem explained earlier - the fact that the same algorithmically generated analyses might likely also facilitate worse decisions, violations of international humanitarian law and exacerbate risks for civilians, especially given the current limitations of the technology, such as unpredictability, lack of explainability and biases. From a humanitarian perspective, this is a chief concern since they pose risks of injury or death to persons or destruction of objects, and because these *decisions* are governed by the *lex specialis* of international humanitarian law.[54]

[54] ICRC, *AI and machine learning in armed conflict, supra* note 40, at 7.