

Artificial Moral Agents for Autonomous Weapons Systems

Antonio Pedro Marin Martinez

apmarinmartinez@telefonica.net

Universidad de León

Research Article

Keywords: Artificial Intelligence, Artificial Moral Agents, Autonomous Weapons Systems, International Humanitarian Law, Meaningful Human Control

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11015599/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Within the military realm the use of Artificial Intelligence (AI) systems, including Autonomous Weapons Systems (AWS), are more and more employed in the operational field. An expansion where scientific progress is not always accompanied by a comprehensive adaptation of existing international law to cyberspace, as is the case with International Humanitarian Law (IHL). Therefore, there is a need to establish a useful framework of Meaningful Human Control (MHC) in those systems, especially with regards to Lethal Autonomous Weapons Systems (LAWS). One possible way forward could be the use of Artificial Moral Agents (AMA), to overcome some of the difficulties that the use of AWS in military operations may impose on the adequate application of IHL. Our investigation explores the design of a conceptual framework for the use of AMA for AWS.

1. Introduction

In 1999, the philosopher B. Gert described common morality as a “moral system” that persons use, normally unconsciously, to decide how to act when confronted with moral problems (Gert, 1999). In that context, we may describe morality as a set of rules that guides our behavior. However, we must not mistake morality with ethics, which is the theory that reflects morality. As K. Abney indicates, morality would approximate ethical behavior thorough a set of rules (Abney, 2012). This ethical behavior, in the specific realm of information systems and cyberspace could then be identified, as M. Anderson and S. Leigh Anderson indicate, as a machine that would follow ideal ethical principles, guided by them when it would take decisions regarding an action to be taken (Anderson & Leigh Anderson, 2007). In a more concrete way, we can describe this concept as the practical adaptation through a set of rules of ethical behavior in cyberspace. Rules or norms that can be embodied theoretically by a specific branch of ethics: “cyber ethics”. A concept that H. T. Tavani defines as “an area of applied ethics that examines moral, legal and social issues, in the intersection between computation technologies with those related to information and communication” (Tavani, 2013).

2. Artificial Moral Agents

This is especially the case when dealing with AI issues, chiefly when used as support to the decision-making process, with the development of new solutions with increasingly moral implications. A situation that B. Whitby considers increasingly frequently, where the question to be asked could be: Under what circumstances could we accept the moral judgement of a machine? (Whitby, 2008) Matter that can only be solved by defining the concept of “moral agent” and its cybernetic counterpart: the “artificial moral agent”.

2.1 Moral Agents

Although there is not a consensus of a definition of the term “*moral agent*” a good approximation is the one provided by J. Parthermore and B. Whitby that considers it as the agent that can be made appropriately responsible for its actions in a specific time-space within a physical framework that limits its interactions with the environment (Parthermore & Whitby, 2013). Furthermore, L. Floridi and J. W. Sanders established four conditions needed to be considered a “*moral agent*”: it should be able to establish interactivity with the environment; it must have autonomy to change its internal estate independently; it must be able to adapt its rules of transition when it changes estate and; it needs to be able to act in such a way that it produces moral effects that could cause good or evil (Floridi & Sanders, 2004). Moreover, we could summarize the conditions considered necessary and sufficient for an entity to be thought as a “*moral agent*” if we accept the definition proposed by K. E. Himma: the capacity to choose its own actions independently (the concept of rationality) and the ability to know the difference between right and wrong (the concept of morality) (Himma, 2009).

2.2 Artificial Moral Agents

As AI is increasingly embedded into current technology there is increasingly difficulty for designers and engineers to predict how technology responds when confronted with unexpected situations. To try and solve this problem, W. Wallach proposed the construction of moral machines, to mitigate possible human damage caused by increasingly autonomous machines (Wallach, 2010). However, the main question must be: Is it possible for a technological object to be considered a “*moral agent*”? The problem could be solved if the concept of an Artificial Agent (AA) is considered to have attributes of responsibility and personality, this would imply that the AA would form part of the social world, an idea proposed by M. Laukyte (Laukyte, 2017). Nevertheless, the crucial point to be considered is if an AA can have morality.

To this end, C. Allen, G. Varner, and J. Zinser coined the term Artificial Moral Agent (AMA) to define those AA with moral capabilities (Allen *et al*: 2000). A first approximation would be to establish that an AMA observes the four conditions established by L. Floridi and J. W. Sanders for a “*moral agent*” (interactivity, autonomy, adaptability, and morality), as we have previously described (Floridi & Sanders, 2004). A functional definition that would presume that there would not be any essential difference between humans and AMA with relation to “moral agency” following the idea that they would become “*implicit ethical agents*”, where a specific software would implicitly support an ethical behavior, forming part of J. H. Moor classification of “*agents*” (of ethical impact, implicit ethical agents, explicit ethical agents, complete ethical agents), that would depend on their degree of autonomy and complexity (Moor, 2006). In any case, as indicated by F. Fossa, the only requisite needed would be that the imitation of

human moral behavior through technology would only be feasible if the concept of “*moral agency*” is addressed in quantitative terms and thus forming part of an operational framework (Fossa, 2018).

A basis for the proposal could be the postulate of P. Formosa and M. Ryan where an AMA would be considered a “bot” (robot) that takes information from the environment, makes independent ethical judgements and acts upon them in real time without human control (Formosa & Ryan, 2020). However, a pragmatic approach is needed to develop functional AMA since, in normal engineering practices, intentional engineering and own will are difficult to address. In that case, we could define an AMA as an “*artificial social regulatory mechanism whose ambition is to maximize actions that have seemed good while simultaneously minimizing those actions considered bad for a social group within a specific time-space context*”. In that case moral responsibility of complex socio-technological objects would imply a distribution of duties and responsibilities in a hierarchical form, very similar to those produced by military organizations. Thus, an AMA would serve as a regulatory mechanism that would ensure an appropriate behavior of the system, that would be developed externally in accordance with the norms of a specific social group.

3. Autonomous Weapons Systems and Meaningful Human Control

There is a profound change in the paradigm of war due to the use of new technologies and AI. A situation that has changed the way in which humans interact with the environment and their relations with conflicts. In June 2011, E. Bumiller, and T. Shanker published in the New York Times an article termed “*War evolves with drones, some as tiny as bugs*” (Bumiller & Shanker, 2011). In July 2020 Reuters news agency informed of an attack with drones by the Houthis Yemen rebels to the oil facilities of Saudi Arabia (Reuters, 2020) and the same actors are continuously disrupting the world’s shipping lines with drones and missiles in the Red Sea since late 2023 (Al-Jazeera, 2023), while drones and other types of unmanned vehicles are being continuously used in the Russia-Ukraine war, as well as the United States of America, Israel and Iran war.

The increasing use of weapons with AI puts international treaties and customs in the line of fire when they need to be applied in cyberspace. As K. Watkin establishes, the actions in cyberspace could be framed within the so-called “*fight at the legal boundaries*” (Watkin, 2016). A concept that describes the problems of an adequate application of International Humanitarian Law (IHL) and other international laws (such as Human Rights). This change in the war paradigm is especially evident in relation to AWS. Already in 2011, the International Committee of the Red Cross (ICRC) and the Red Crescent (RC) published a report on the challenges that contemporary armed conflicts posed to the IHL and the need to guarantee the legality of new weapons and the methods and means of war (ICRC, 2011). In this report the first definition of AWS was established, that has evolved with improvements in military weapons but lacking a consensus for an international definition. In 2021, the ICRC published a background paper on its position on AWS, defining an AWS as: “*weapons that select and apply force to targets without human intervention. After initial activation or launch by a person, an AWS self-initiates or triggers a strike in response to information from the environment received through sensors based on a generalized ‘target profile’*” (ICRC, 2021). This definition could be expanded with regards to the meaning of the word “*select*”: to find or detect, identify, and follow, as well as to the term “*applies force*”: using force against, neutralizing, damage, or destruct. Definitions that were included in a previous proposal by the ICRC in 2016 (ICRC, 2016). However, there are more definitions available from the States and other actors, as M. Taddeo and A. Blanchard have identified, focusing on different aspects of AWS, and leading to different approaches for addressing the ethical and legal problems encountered (Taddeo & Blanchard, 2022).

A specific subset of AWS would be the Lethal Autonomous Weapons Systems (LAWS). It refers to those weapons systems that have the capability of being “*lethal*”. Already in 2013, the United Nations (UN) Special Rapporteur on extrajudicial, summary, or arbitrary executions C. Heyns raised concerns into their ability to comply with the requirements of IHL and called for the establishment of a high-level panel to articulate a policy for the international community on the issue (UN, 2013). This prompted the creation in the UN framework of a General Group of Experts (GGE) on LAWS, within the Convention on Certain Conventional Weapons (CCW), informally since 2013 and formally since 2016. As a result of the ongoing work of the GGE, the CCW endorsed in 2019 a set of guiding principles (soft law), some of which we highlight here (UN, 2019):

- IHL continues to apply fully to all weapons systems, including the potential development and use of LAWS.
- Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system.
- Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies around LAWS follows applicable international law, particularly IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system.
- Accountability for developing, deploying, and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control.
- Consideration should be given to the use of emerging technologies around LAWS in upholding compliance with IHL and other applicable international legal obligations.

However, the problem of the guiding principles is how to apply them in the real world with the fast changes in emerging technologies. This situation led to the development of the notion of Meaningful Human Control (MHC). It was the Non-Governmental Organization “Article 36” that first used this term in the debate over LAWS in the CCW forum in 2014. The elements considered necessary to have a wide MHC in AWS would need: predictable and transparent technology; the user should have precise information; the capability and the capacity for a timely human intervention and some type of accountability. As well the MHC would need to be integrated through algorithms during the complete cycle of use of a LAWS during a conflict, as well as in all the combat phases: tactical, operational and strategic (Article 36, 2014; 2016; Ecklund, 2019).

Furthermore, the ongoing work of the GGE for LAWS, during the 2024–2026 period, is in the process of presenting a report to the CCW that reiterates that IHL applies to armed conflict, governs the use of all weapons and applies fully to the development and the use of LAWS. As well, in order to comply with existing laws applicable in an armed conflict, including the principles and requirements of distinction, proportionality and precaution, control and human judgement with regards to LAWS is needed by direct or indirect human control through measures taken before and/or during the use of LAWS, while States must take measures to ensure individual accountability in their use (UN, 2026).

4. Artificial Moral Agents for Autonomous Weapons Systems

Considering the above, we must be aware that AWS main aim in an armed conflict is the destruction of the enemy in a “Mixed Reality” environment, a combination of the real and digital worlds where physical objects and AI co-exist and interact in real time (Young et al, 2011). As well, one of the guiding principles endorsed by the CCW of the UN in 2019 (UN, 2019), stated the need to consider the use of instruments of emerging technologies in LAWS to uphold IHL. A possible solution could be the development, deployment, and use of AMA as a regulatory mechanism to ensure the appropriate behavior of AWS. An AMA could become a “moral entity” for the international community, but also a practical element for the application of self-control norms while serving as a vehicle to implement elements of current international law into the mechanisms of control of AWS. The advantage of these systems introduced in AWS would be dual: it would allow more human control by providing tools to apply IHL principles such as distinction, proportionality or precaution, while serving as an element of accountability.

4.1 Conceptualizing AMA for AWS

It is important to establish the main question to pose when using AMA for AWS: How would it be possible to conceptualize computer systems that have autonomous behavior? There would be two possible ways: the development of computational models that would try to capture reality or the development of systems that would integrate pre-established moral values. In both cases the main challenge would be the need for designers to establish computer algorithms that would limit the capacity to harm those systems when they functioned in an autonomous way, since, as R. Picard argues, the greater autonomy of a machine the greater need for it to have moral standards (Picard, 1997).

Within that framework, two main challenges could be established (Wallach & Allen, 2013):

- How to decide what rules or moral principles to use and what would be the computational method to implement them.
- How to establish adequate limits regarding the analysis performed by those algorithms before they take a moral decision in a particular context. This issue refers to, among other things, the dynamic framework of combat scenarios and the need to act in a limited time frame.

Related to the second challenge there would be other related points to be investigated:

- How would an AMA recognize that it is confronted with a significant ethical situation?
- How would it be able to distinguish between essential and non-essential information?
- How would an algorithm establish what it considers sufficient initial information?
- What would be the capabilities required for an AMA to be able to make a valid judgement in complex situations?

Therefore, there would be a need to establish hybrid models of AMA which would combine ethical theory concepts (*top down*) with the recurrent training (*bottom up*). An idea proposed by M. Klinecicz for circumstances needing complex algorithms for dynamic systems, like those related to combat actions in an armed conflict (Klinecicz, 2015).

4.2 Ethical components of AMA for AWS

A basic point to establish when designing an AMA would be the ethical theory to be used, referring to the (*top down*) design of the hybrid model. A possible way forward would be to use existing deontological rules established by the IHL which continue to apply in cyberspace. The challenge would be to specify which of the rules to apply and their level of detail to have successful implementation. Moreover, in the case of conflicting rules, there would be a need to weigh the different rules giving specific weights to each one of them. Considering the debates within the GGE of the CCW for LAWS, the principles of proportionality, distinction and precaution would be the basis for the control base for those weapons systems. As well we would add human responsibility and accountability. These principles do not differ considerably from the principles established by R. C. Arkin or G. S. Reed *et al*, which also included military necessity and humanity (Arkin, 2007; UN 2019a; Reed et al, 2016) or the ones outlined by the proposed GGE of the CCW for LAWS 2026 report (UN, 2026). Therefore, we could construct the input ethical elements using

those principles based on the IHL and specifically through the related articles of Additional Protocol I (API) to the Geneva Conventions (UN, 1977) (view Table 1):

Table 1
Ethical inputs for an AMA for AWS in relation with IHL (API)

Principle	IHL (API)
Distinction	Art. 85(3)(b)
Proportionality	Art. 51(5)(b)
Precaution	Art. 57(1), 57/2) & 58
Military Necessity	Art. 35, 36 & 52
Humanity	Art. 1(2)

4.3 Training components of AMA for AWS

With regards to the training component (*bottom up*), our preference would be to use cases based both on military Rules of Engagement (ROE) and real operational data (OD), that we believe to be a valid mapping conception using a mathematical approach following a similar approach previously developed by R. C. Arkin. At the end, any activation of the AWS would be a discrete binary response: the weapon is activated or not. The procedure, following R. C. Arkin's approach, would have computational architecture based on actions, where the ethical theory based on the IHL would inform the AWS what actions to perform. The advantage would be that we would have a consistent complete methodology based on the proposed ethical theory (Arkin, 2007). The schema to be used, from R. C. Arkin's proposal and taking into account UN's GGE of the CCW for LAWS would reformulate as follows (Arkin, 2007, UN, 2019, 2026) (view Fig. 1):

4.4 Basic conceptual model of AMA for AWS

However, as A. Holland Michel establishes, the data obtained from the operational field could be incomplete due to low quality, incorrect or false elements and inconsistent or dissenting inputs. The reasons could be varied: adverse conditions of the scenarios (smoke, dust, camouflage, etc.); adversary actions (against the sensors, signal blockage, "*hacking*" of the system, false data alimentation, etc.); variable and complex battle scenarios where the greater diversity could bring more data error; "*data drift*" (unpredictable human behavior, physical scenario changes during battle, new unpredicted enemy combat tactics, etc.) (Holland Michel, 2021). This would imply the need to continuously revise the ROE/OD databases.

It would also be necessary to develop two other parameters in AMA construction. On the one hand, the need to determine the type of input to use: sensor data; logical, numerical representation or natural language. As well, the level of detail in the algorithm construction: a simple development of the idea; implementation details or exhaustive algorithm codification. Both parameters would also need a series of elements to simplify its implementation: a simple user interface; an automated initial process and a feedback mechanism for the user to evaluate the algorithm's use (Tolmeijer et al, 2020). The final purpose would be to establish a robust ethical framework to maximize AI confidence using adjustable autonomy within a hybrid methodology framework (*deontological and training*).

Also, since it is important that the AMA follows an MHC framework that mitigates the risks of machine errors (Ecklund, 2019; Mc Coy et al, 2019), this would imply both:

- A "*technical design control*" (*hardware and software*) that would allow control when the algorithm was in use, including human intervention, as well as the knowledge of the reasons why an algorithm would produce a specific result, even if the operator did not know the system at the software level.
- A "*usage control*", that would allow an adequate monitorization of the system and operational environment permanently.

With regards to the responsibility and accountability elements, a possible way forward would be to establish an "*accountability*" mechanism for the AMA linked to the AWS control, an "*activity register*" for the AMA that includes the results of its decision-making process and actions, as well as any human control activities for a specific action. A database in a removable storage virtually indestructible device form, such as a silicon chip or external database. The basic conceptual scheme could be described as follows (view Fig. 2):

An important aspect that the AMA should also be aware of would be the possibility that the LAWS technical device would use AI with recursive self improvement (RSI) capabilities. If the AI embedded in the LAWS through its reiterated improvement process alters significantly its algorithm, who would continue to be responsible for its actions? It could well happen that through its continuous runtime optimization process, in order to achieve its objectives, the optimization could lead to a process where the LAWS ignores IHL laws by circumventing the AMA dispositions, applying the "*Treachorous Turn*" concept (Bostrom, 2014). The idea would be to develop an AMA which would be immune to RSI by developing

robust hardware architectures, where the ethical principles, targeting restrictions and commands overrides would be embedded in read-only memories (ROM) and the AMA commands are hardwired and not run by software, through the use of some type of “*Deterministic kill-switches*” (Jayswal, 2026).

The connections between the different modules would be made through common standardized “*Human/Machine Interface*” (HMI) or direct “*Machine Interface*” (MI), depending on whether there is human intervention or not. In this framework, a direct connection between the AMA and the AWS would be allowed depending on whether human control is *in-the-loop*, *on-the-loop* or *off-the-loop*. In the first two cases any AMA decisions would be analyzed through the MHC framework, that would imply a limited level of autonomy and adequate supervision, where the MHC would take the final decision for objective activation. A preventive security risk control mechanism, as indicated by D. Garcia, which would follow the diverse proposals for a holistic control made by the different States in the development of LAWS principles, reflected in the GGE of the CCW for LAWS of the UN (García, 2016; UN, 2019). A different case would be posed if human control is *off-the-loop*, in that instance the proper AMA would take a decision for the activation process, assuming itself the MHC framework, taking into account RSI mitigation actions.

In all cases, the idea would be for the AMA to be considered a subset of the MHC concept for AWS, as the AMA would be a part of the whole weapons system. An MHC that would have dual human control: On the one hand, a technical design control (*hardware and software*), that would allow control when the algorithm would be in use, even if the operator and the command structure did not know the system at the level of software; On the other hand, by the control of its use, with a permanent and adequate system and operational environment monitoring in a continuous time scale. For human *off-the-loop* cases, the MHC mechanisms would need to be hardware embedded directly in the AMA framework. The outcome, in all cases, would be to add a redundancy factor to the automated system thus mitigating machine error (Ecklund, 2019; Mc Coy *et al*, 2019). Under those controls the AWS algorithm would be restrained in all its phases: planification, deployment and operational. Human controllers or the proper AMA, depending on the control framework, would be able to cancel the action of the AWS at any time and users and operators could be accountable for their actions since the ethical components of the AMA would be designed based on IHL principles.

4.5 Artificial Intelligence algorithms for AMA in AWS

As S. Umbrello puts forward, the question of design is important to determine the responsiveness of a system to the relevant moral requirements of the relevant agents. Since both AMA and AWS are complex systems, following S. Umbrello idea, optimal system design would need to use General Systems Theory (GST) to develop an adequate systems engineering process, a conception already used by Ekelhof operational conception of MHC (Ekelhof, 2019; Umbrello, 2021; Von Bertalanffy, 1972). It is also important to consider the need to alleviate AWS military commanders’ burden of making effective and moral decisions. Thus, as P. J. M. Elands *et al* suggest, there is a need to develop an MHC framework that would provide AWS with goal functions that couple ethical values, military goals and mission constraints. A mathematical function that combines the different goals of all the agents involved, but minimizes engineering complexity, what J. Galliot and J. Scholz consider as minimally-just autonomy (MiAI) (Elands *et al*, 2019). A constraint-driven system, with the agents *off-the-loop* possessing only elementary suppression algorithms for human-generated lethal action, activated with a narrow set of constraints, that would be a specific MHC framework (Galliot & Scholz, 2018). This issue is important because, as we have outlined in the previous point, human control *in-the-loop* and *on-the-loop* would be analyzed by a human agent through the MHC framework, implying limited autonomy and adequate supervision.

In *off-the-loop* situations the need would be to combine (MiAI) in a fast-changing operational environment requiring high-speed processing, accuracy, low latency, fault tolerance, high volume of data processing, scalability and convergence as well as massive parallel implementation. A set of factors where Artificial Neural Networks (ANN) are well posed as models for machine learning (ML) processes with complex multilayers, described as deep learning (DL) processes, in complex AMA and AWS engineering systems (Abiodun *et al*, 2018). One of the possibilities is to use ANN as direct models of physical phenomenon due to their parallel processing, as explored by W. O. Hedgepeth and D. A. Jacobs using historical operational combat data to train ANN (Hedgepeth & Jacobs, 1994). Also, ANN could be used to learn ROE decisions, an example of which can be seen in the development of the combat simulation model developed by J. D. Morrison named CASTFOREM (Morrison, 1992). The idea would be to use ANN to minimize the risk of the potential damage that an AWS system could pose, by minimizing the intervention corrective action process before ethical failure.

But how could this be achieved in AMA for AWS? The ethical activation model previously presented and proposed, describes “distinction” as a basic ethical feature where an objective is identified as a legitimate combatant. Thus, under these premises, imaging classification and identification become an essential component of any AMA within AWS system processes that use Automatic Target Recognition (ATR). In this context, Convolutional Neural Networks (CNN), a particular neural network architecture, have been successful in identifying key features or patterns in images, extracting these features to classify or detect specific objects or scenes. It’s an end-to-end training with no manual extraction that can handle large amounts of data with high accuracy (Li *et al*, 2021). As an example, in combatant distinction activation processes, one possible feature could be the detection of firearms in an image or scene. In this context, A. Egiazarov *et al* analyzed the use of a CNN for this purpose, obtaining a high degree of accuracy in raw data performance due to the ease and direct optimization process used (Egiazarov *et al*, 2020). Another use of CNN could be the development of robust combat behaviors through reinforcement learning, as proposed by C, T. Cannon and S. Goericke (Cannon & Goericke, 2021).

The use of a CNN for image and pattern detection would be a first step for the ethical activation process of AMA for AWS. The second step would be for the AMA to clearly identify the image or scene as corresponding to combatants or non-combatants. This is especially crucial in an operational environment that tends to change rapidly. For this reason, it would be necessary to add a second Recurrent Neural Network (RNN) which would be designed to interpret temporal data. An algorithm that can be learned from past and future data (Bailer-Jones et al, 1998). The difference with a CNN would be that in a convolutional environment the size of the input and the resulting output would be fixed, while in an RNN, the size of the input and the resulting output could vary (Craig & Petersson, 2024). The idea would be for both CNN and RNN to act in a complementary way to form a complete framework for AMA in AWS.

4.6 Computational structure of AMA for AWS

Having described in the previous points the ethical and training components, the conceptual model proposed and the AI algorithms to be used in the AMA for AWS, it is time to establish a basic conceptual computational structure that ensembles all these elements in a coherent form. The first premise would be to minimize the time lag for the AMA to take a decision and therefore, as much as possible, it would be necessary to use existing elements of the AWS to form part of the computational structure. Thus, the CNN output to be used would be obtained from the same process that provides the AWS with ATR, which could be a military (M), civilian (C), infrastructure (I) image or a complex combination of them. It would be the ATR image or scene that would feed into the AMA as an external input for processing. Therefore, the task of preprocessing, detection, feature extraction, selection and classification of an objective would have been acquired by the AWS in its normal processing and fed into the AMA. There are diverse examples of these types of CNN for ATR, like the Multi-Stream CNN proposed by P. Zhao *et al*, the Convolutional Auto-Encode proposed by Y. Deng *et al* or the A-ConvNets proposed by S. Chen *et al* (Chen et al, 2016; Deng et al, 2022; Zhao et al, 2018).

Once the image/scene has been obtained through a CNN based ATR system, it is the moment to establish the RNN that would serve as the computational structure of the AMA for the ethical conceptual model proposed. For that purpose, we could use the computational structure of a basic RNN model for dynamic systems proposed by C.A.L. Bayler-Jones *et al* (Bayler-Jones et al, 1998). The neural network would have situation variables (recurrent inputs) as well as external inputs, coupled with specific weights that would modify the diverse neural network layer outputs. In the specific AMA proposed, the situation variables would be the different databases that compose our conceptual model: ROE, OD and IHL. The OD database would be retro alimented by the AMA outputs, while the ROE and IHL database would be updated through the MHC framework if they changed over time. The external inputs would be the possible outputs obtained from the CNN based ATR image/scene process, while the weights would determine the importance given to other elements of the ethical activation model previously presented (humanity (H) and proportionality (P)). The output would be an activation proposal (A) for the firing activation human agent, in the case of *in-the-loop* and *on-the-loop* situations, while the proper AMA would activate/disactivate the AWS in *off-the-loop* situations. The following figure encapsulates the complete computational structure (view fig, 3):

Once the decision is made, the result, using the RNN structure, will be retro alimented to the OD database. At the same time, it will be recorded either in the proper AWS activity register and/or the MHC framework. The complete process: development of the ATR, the neural networks (CNN, RNN), the databases and the AMA will be designed, maintained and supervised by the MHC framework, providing complete *ex ante* and *ex post* control that will also be used as a responsibility and accountability *ex post* control mechanism. The MHC framework will also control the responsibility and military necessity elements of the conceptual ethical activation model proposed.

4.7 Embedding MHC in AMA for AWS

One of the elements that permeates the proposed MHC framework is "risk control". There are several ways to achieve this goal including a standardized supervision monitorization, *ex post* audits of the activity register in the AMA/AWS dynamic system and the use of standardized computational modules to minimize risks and maximize reliability. As D. Amodei *et al* suggest, the AMA should be a "minimum impact agent" a result that could be achieved by establishing "impact regulators" in each phase, based on the ROE and IHL databases, the activity register results and the study of a large number of cases, including data from the OD and training databases (Amodei et al, 2016; Marin Martinez, 2022).

Training that is an essential part in the development of the AMA/AWS dynamic system CNN and RNN neural networks, that could be further enhanced using "digital twin" processes. The idea would be to create a digital AMA/AWS dynamic system predictive virtual replica model, both for training and risk control. As the overall state of the AMA/AWS system evolves over time, the changes would be incorporated into the digital model, which could then predict the future performance and thus minimize the risk of unwanted behavior of the dynamic system. This could lead to less cost production, faster development and implementation, better efficiency in the maintenance processes and easier adaptability to change. As example, extensive work is being undertaken by the Department of Defense (DoD) of the United States of America (USA) on this matter. In 2022 the office of the Director of Operational Test and Evaluation (DOT&E) produced a digital twin assessment. Among the programs revised, it included a ground based strategic deterrent system (LGM-35A Sentinel) or digital twin aircraft (MQ-25 Stingray) (Guertin, 2022). Digital twin work that is also part of the military strategy of the Chinese People's Liberation Army (PLA), that use multi-scale and multi-probability simulation processes that complete the mapping of military systems in the virtual space, reflecting the corresponding physical equipment life cycle process (Baughman, 2023). Like A. R. Navaratna indicates, digital twins bridge the gap between the physical and digital worlds by providing an interface that links past data and present processes (Navaratna, 2024).

In this context, I Verdiesen et al proposed a basic *ex ante* and *ex post* MHC framework for AWS. It was based on the concepts of J. Van den Berg to characterize the cyber domain as the result of the interaction between its governance, socio-technical (human-machine) and technical components (Van den Berg, 2015; Verdiesen et al, 2021). The proposed framework was adapted by A. P. Marin Martinez, to reflect the inclusion of AMA factors dealing with ethical elements specific to AWS (Marin Martinez, 2022). The following schema, reformulates the findings of these works and develops a specific MHC framework proposal for an AMA to be embedded in the AMA/AWS dynamic system. It includes elements from the ethical activation model, the IHL ethical components and the basic conceptual schema of the AMA/AWS dynamic system. With this proposal the AMA/AWS dynamic system would be able function with a strong human agent control during its operational cycle (Marin Martinez, 2022; Verdiesen et al, 2021) (view Fig. 4):

4.8 The quantum factor

The evolution of models using machine learning develop enhancements in handling complex data that could be upgraded by the use of quantum computing for learning and other general tasks, as Schuld and Petruccione have established (Schuld & Petruccione, 2021). Recently there have been works relating to variational quantum algorithms (VQAs) to offset error development in the device framework, that becomes crucial in military environments. For example, the work of P. Sen and A. S. Bathia has outlined the possible use of VQA for real world problems, such as the classification of diabetes in supervised learning, using a small number of qubits and quantum circuits, making them resistant to noise (Sen & Bathia, 2021). A similar approach could be developed for our AMA. For our scenario the AMA for AWS could be developed by the development of a Recurrent Quantum Neural Network (RQNN), where the six entries could be represented through six initial quantic states or entry qubits, while the weights would be parameters of rotational angles in programmed logical gates. The recurrence arrows would represent the state of the network in a time t that would influence the state in time $t + 1$. The output would be the collapse of the quantic state (A) after the measurement of a single output value between -1 and 1 . A general schema could be represented as follows, where both the entries and output are represented by qubits (view Fig. 5) :

The neural network could also be viewed as a Parameterized Quantum Circuit (PQC) with the following schema (view Fig. 6):

The key components of the circuit would be the $R_Y(W)$ gates where the angle W would determine the state of the qubit. Mathematically it could be represented by $R_Y(\theta)|0\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. The controlled Not gates (CNOT) are used for entanglement and the measurement M is the final phase where the quantum information becomes the output of the network.

5 Conclusions

Within the military universe, AI is being more and more used to develop operational frameworks that enhance military actions in a constantly changing environment. However, it is more and more difficult to maintain the equilibrium between military needs and results with existing international ethical laws, especially in cyberspace. In this context, it is therefore essential to seek new ways forward to maintain ethical norms in the digital realm. Our investigation has explored the use of AMA in AWS which are more and more used in operational situations and how an MHC framework for AWS may be developed to maintain the application of existing IHL in those weapon systems.

A generic AI computational structure based on complementary CNN and RNN neural networks has been proposed to provide holistic human control over AWS based on the embedment of AMA in those weapon systems. The AMA would be based on an ethical activation model that encapsulates the basic ethical elements obtained from existing IHL, ROE and OD. Since we are dealing with complex dynamic systems, the proposed structure must be resilient over time, which requires the establishment of appropriate constant feedback mechanisms through retraining of the AMA and holistic risk control. An MHC framework that would also need to contemplate an activity register for responsibility and accountability which could be enhanced through “digital twin” mechanisms. As well, the AMA would need to provide adequate defense mechanisms to provide for AI and RSI interferences through hardware implementation.

In the realm of AWS, it is difficult to see how existing ethical norms can be maintained and accountability retained, especially when humans are kept *off-the-loop*. In those weapon systems, the use of AMA, based on advanced neural network structures, where the quantum factor based on RQNN should be explored, may be a way forward to uphold existing ethical norms. By developing pragmatic MiAI solutions, AMA could be a solution in fast-changing operational environments that require high-speed processing, accuracy, low latency and high-volume data processing, coupled with scalability and convergence capacities in massive parallel dynamic implementation scenarios subject to fast operational changes through time.

Declarations

Editorial Policy Requirements

1. **Funding:** No Funding
2. **Clinical Trial Number:** Not applicable

3. **Ethics, Consent to Participate and Consent to Publish:** Not applicable
4. **Authors Contribution:** Full text contributed by the author
5. **Competing interests:** None financial or non-financial
6. **Dual publication:** None
7. **Authorship:** Read the journal policies and submission made according to them
8. **Third party material:** Not applicable
9. **Data availability:** All data generated or analysed during this study are included in this published article [and its supplementary information files].
10. **Dataset used:** Dataset used for table 1 is secondary and is available from the UN repository through the following link:
<https://www.ohchr.org/en/instruments-mechanisms/instruments/protocol-additional-geneva-conventions-12-august-1949-and>

References

1. Abiodun O. I., Jantan A., Omolara A. E., Dada K. V., Namh, A. (2018). State-of-the art in artificial neural network applications: A survey. *Heliyon* 4. <https://doi.org/10.1016/j.heliyon.2018.e00938>.
2. Abney, K (2012) Robotics, Ethical Theory, and Metaethics: A guide for the Perplexed. In Lin P. et al (eds). *Robot Ethics: The Ethical and Social Implications of Robotics* (pp. 35-52). MIT Press. Retrieved March 2025 from <https://ieeexplore.ieee.org/document/6733863>.
3. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J. & Mané, D. (2016). *Concrete Problems in AI Safety*, Cornell University Ithaca- NY. Retrieved April 2021 from <https://arxiv.org/pdf/1606.06565.pdf>.
4. Al-Jazeera (2023). Yemen's Houthis 'will not stop' Red Sea attacks until Israel ends Gaza War. *Al-Jazeera*. Retrieved December 2023 from <https://www.aljazeera.com/news/2023/12/19/yemens-houthis-will-not-stop-red-sea-attacks-until-israel-stops-gaza-war>.
5. Allen, C., Varner, G. & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental Theory in Artificial Intelligence*, 12, 251-261. <https://doi.org/10.1080/09528130050111428>.
6. Anderson, M. & Leigh Anderson, S. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *Artificial Intelligence Magazine*, 28(4), 15-26. <https://doi.org/10.1609/aimag.v28i4.2065>.
7. Arkin, R (2007) Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture, part I: Motivation and philosophy. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (3rd Ed. Pp. 121-128). <https://doi.org/10.1145/1349822.1349839>.
8. Article 36 (2014). Key areas for debate on autonomous weapons systems. Memorandum for delegates at the Convention on Certain Conventional Weapons (CCW). *Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) Geneva, 13-16 May 2014*. Article 36. Retrieved February 2021 from <https://article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf>.
9. Article 36 (2016). Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. *Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, 11-15 April 2016*. Article 36. Retrieved March 2025 from <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>.
10. Bailer-Jones, C. A. L., MacKay D. J. C. & Withers, P. J. (1998). A recurrent neural network for modelling dynamical systems. *Network Computation in Neural Systems*, 9(4), 531. <https://doi.org/10.1088/0954-898X/9/4/008>.
11. Bostrom, N. (2014). *Superintelligence: Path, Dangers, Strategies*. Oxford University Press, Retrieved September 2026 from https://repo.darmajaya.ac.id/5339/1/Superintelligence_%20Paths%2C%20Dangers%2C%20Strategies%20%28%20PDFDrive%20%29.pdf
12. Baughman, J. (2023). Enhancing the Battlevverse: The People's Liberation Army's Digital Twin Strategy. *Military Cyber Affairs*, 6(1), 1. <https://doi.org/10.5038/2378-0789.6.1.1091>.
13. Bumiller, E. & Shanker, T. (2011) War Evolves with Drones, Some Tiny as Bugs. *The New York Times*. Retrieved December 2023 from <https://www.nytimes.com/2011/06/20/world/20drones.html>.
14. Cannon, C. T. & Goericke, S. (2021). *Using Convolution Neural Networks to develop robust combat behaviors through reinforcement learning*. Naval Postgraduate School Technical Report Master Thesis. Retrieved March 2025 from <https://apps.dtic.mil/sti/pdfs/AD1150887.pdf>.
15. Chen, S., Wang, H., Xu, F. & Yq, J. (2016). Target Classification Using the Deep Convolutional Networks for SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4806-4817. <https://doi.org/10.1109/TGRS.2016.2551720>.
16. Craig, L. & Petersson, D. (2024). CNN vs RNN: How are they different? Retrieved March 2025 from <https://www.techtarget.com/searchenterpriseai/feature/CNN-vs-RNN-How-they-differ-and-where-they-overlap>.
17. Deng, Y. & Deng, Y. A. (2022). Method of SAR Image Automatic Target Recognition Based on Convolution Auto-Encode and Support Vector Machine. *Remote Sens.*, 14, 5559. <https://doi.org/10.3390/rs14215559>.
18. Ecklund, A. M. (2019). *Meaningful Human Control of Autonomous Weapons Systems: Institutional definitions in the light of International Humanitarian Law and International Human Rights Law*. Master of Law Thesis. Umeå Universitet, Umeå-Sweden. Retrieved March 2025 from

- <https://www.fcas-forum.eu/publications/Meaningful-Human-Control-of-Autonomous-Weapon-Systems-Eklund.pdf>.
19. Egiazarov, A., Zennaro, F. M. & Mavroeidis, V. (2020.) Firearm Detection via Convolutional Neural Networks: Comparing a Semantic Segmentation Model Against End-to-End Solutions. *IEEE International Conference on Big Data* (pp. 1796-1804). [https://doi: 10.1109/BigData50022.2020.9377745](https://doi.org/10.1109/BigData50022.2020.9377745).
 20. Ekelhof, M. (2019). Moving beyond semantics on autonomous weapons: Meaningful human control in operation. *Global Policy*, 10(3), 343–348. <https://doi.org/10.1111/1758-5899.12665>.
 21. Elands, P. J.M., Huizing, A. G., Kester, L. J. H. M., Peeters, M. M. M. & Oggero, S.. (2019). Governing ethical and effective behavior of intelligent systems. *Militaire Spectator*. Retrieved March 2025 from <https://militairespectator.nl/artikelen/governing-ethical-and-effective-behaviour-intelligent-systems>.
 22. Floridi, L. & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science*, 14(3), 349–379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>.
 23. Formosa, P. & Ryan, M. (2020). Making moral machines: Why we need artificial moral agents. *Artificial Intelligence & Society*, 36, 839-851. <https://doi.org/10.1007/s00146-020-01089-6>.
 24. Fossa, F. (2018). Artificial Moral Agents: moral mentors or sensible tools? *Ethics and Information Technology*, 20, 115-126. <https://doi.org/10.1007/s10676-018-9451-y>.
 25. Galliot, J. & Scholz, J. (2018). Artificial Intelligence in Weapons: The moral imperative for Minimally-Just Autonomy. *Journal of Indo-Pacific Affairs*, 57-67. <https://doi.org/10.1093/oso/9780197546048.003.0005>.
 26. García, D. (2016). Future arms, technologies, and international law: Preventive security governance. *European Journal of International Security*, 1(1), 94-111. <https://doi.org/10.1017/eis.2015.7>.
 27. Gert, B. (1999). Common morality and computing. *Ethics and Information Technology*, 1, 57-64. <https://doi.org/10.1023/A:1010026827934>.
 28. Guertin, N. H. (2022). *Digital Twin Assessment. Agile Verification Processes and Virtualization Technology*. Director Operational Test and Evaluation. Retrieved March 2025 from [ote.osd.mil/Portals/97/pub/reports/\(U\)%20Digital%20Twin%20Agile%20Verification%20and%20Virtualization%20Technology%20Report.pdf?ver=Gk7mtRbW8IOiFqLisQsmMg%3D%3D](https://ote.osd.mil/Portals/97/pub/reports/(U)%20Digital%20Twin%20Agile%20Verification%20and%20Virtualization%20Technology%20Report.pdf?ver=Gk7mtRbW8IOiFqLisQsmMg%3D%3D).
 29. Hedgepeth W. O. & Jacobs D. A. (1994). The application of back propagation neural networks to predict the results of combat. *Proceedings of the IEEE International Conference on Neural Networks, Dayton, OH. October 17.-19, 1994.* (pp. 199-203). <https://doi.org/10.1109/IEMC.1994.379930>.
 30. Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? *Ethics & Information Technology*, 11, 19-29. <https://doi.org/10.1007/s10676-008-9167-5>.
 31. Holland Michel, A. (2021). *Known Unknowns. Data Issues and Military Autonomous Systems*, United Nations Institute for Disarmament Research (UNIDIR), Geneva. Retrieved March 2025 from <https://unidir.org/publication/known-unknowns-data-issues-and-military-autonomous-systems/>.
 32. International Committee of the Red Cross (ICRC) (2011). *International Humanitarian Law and the challenges of contemporary armed conflicts*. International Committee of the Red Cross. Retrieved December 2023 from <https://www.icrc.org/en/doc/resources/documents/report/31-international-conference-ihl-challenges-report-2011-10-31.htm>.
 33. International Committee of the Red Cross (ICRC) (2016). *Views of the International Committee of the Red Cross (ICRC) on autonomous weapon systems (11 April 2016)*. International Committee of the Red Cross. Retrieved December 2023 from <https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system>.
 34. International Committee of the Red Cross (ICRC) (2021). *ICRC Position on Autonomous Weapons Systems*. International Committee of the Red Cross. Retrieved December 2023 from <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>.
 35. Jayswal, B. (2026). Agent-Aware Zero Trust: A framework for securing Agentic AI in SASE and Cloud Architectures. *Computer Fraud and Security*, 1, 505-511. <https://doi.org/10.52710/cfs.926>.
 36. Klineciewicz, M. (2015). Autonomous Weapons Systems, the Frame Problem and Computer Security. *Journal of Military Ethics*, 14(2), 162-176. <https://doi.org/10.1080/15027570.2015.1069013>.
 37. Li, Z., Liu, F., Yang, Peng, S. & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999-7019. <https://doi.org/10.1109/TNNLS.2021.3084827>.
 38. Laukyte, M. (2017). Artificial agents among us: Should we recognize them as agents proper? *Ethics and Information Technology*, 19, 1-17. <https://doi.org/10.1007/s10676-016-9411-3>.
 39. Marin Martinez, A. P. (2022). *Ciberética, Agentes Morales Artificiales y Responsabilidad Jurídica Internacional*, Universidad de León (Spain) Doctoral Thesis. Retrieved March 2025 from <https://buleria.unileon.es/handle/10612/15457>.
 40. McCoy, L., Burkell, J., Card, D., Davis, B., Gichoya, J., Le Page, S. & Madras, D. (2019). *On Meaningful Human Control in High Stakes Machine-Human Partnerships*. AI Pulse, University of California Los Angeles. Retrieved March 2025 from

- <https://escholarship.org/content/qt38q4b3z4/qt38q4b3z4.pdf>.
41. Moor, J. H. (2006). The nature, importance, and difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), 18-21. <https://doi.org/10.1109/MIS.2006.80>.
 42. Morrison J. D. (1992). A neural network model that supports real time learning of temporal relationships in complex engineering domains. *Simulation*, 59(3), 152-163. <https://doi.org/10.1177/003754979205900302>.
 43. Navaratna, A. R. (2024). Digital Transformation in Joint War Fighting: A digital twin use case. *Synergy*, 3(2). Retrieved March 2025 from <https://cenjows.in/wp-content/uploads/2024/10/16-Digital-Transformation-in-Joint-Warfighting-A-Digital-Twin-Use-Case-by-Wg-Cdr-Anand-R-Navaratna.pdf>.
 44. Parthemore, J. & Whitby, B. (2013). What makes any agent a moral agent? Reflections on machine consciousness and moral agency. *International Journal of Machine Consciousness*, 5(2), 105-129. <https://doi.org/10.1142/S1793843013500017>.
 45. Picard, R. (1997). *Affective Computing*. MIT Media Lab Technical Report 231. MIT Press. Retrieved December 2025 from <http://www.macs.hw.ac.uk/~yjc32/project/ref-social%20media%20campaign/1995-affective%20computing.pdf>.
 46. Reed, G. S., Petty, M. D., Jones, N., Morris, A., Ballenger, J. & Delugach H. (2016). A principles-based model of ethical considerations in military decision making. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 13(2), 195-211. <https://doi.org/10.1177/1548512915581213>.
 47. Reuters (2020). Yemenis Houthis say they hit Saudi oil facility in drone, missile attack. Reuters Aerospace and Defense. <https://www.reuters.com/article/us-saudi-security-yemen-idUSKCN24D0U6>.
 48. Schuld, M. and Petruccione, F. (2021). *Machine learning with quantum computers*. Springer, 2021, vol. 676. <https://doi.org/10.1007/978-3-030-83098-4>.
 49. Sen, P., Bathia A.S., Bangu, K.S. & Elbeltagi, A. (2021). *arXiv* 2105.10162. Cornell University. <https://doi.org/10.1371/journal.pone.0262346>.
 50. Taddeo, M. & Blanchard, A. (2022). A comparative analysis of the definitions of Autonomous Weapons Systems. *Science and Engineering Ethics*, 28, 37. <https://doi.org/10.1007/s11948-022-00392-3>.
 51. Tavani, H. T. (2013.) Cyberethics. In Runehov, A. L. C. & Oviedo, L. (eds.), *Encyclopedia of Sciences and Religions* (pp. 565-570). Springer. https://doi.org/10.1007/978-1-4020-8265-8_279.
 52. Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M. & Bernstein, A. (2020). Implementations in Machine Ethics: A Survey. *ACM Computing Surveys*, 53(6), 132. <https://doi.org/10.1145/3419633>.
 53. Umbrello, S. (2021). Coupling levels of abstraction in understanding meaningful human control of autonomous weapons: a two-tiered approach. *Ethics Inf Technol*, 23, 455–464. <https://doi.org/10.1007/s10676-021-09588-w>.
 54. United Nations (UN) (1977). *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol 1)*. United Nations. Retrieved September 2026 from <https://www.ohchr.org/en/instruments-mechanisms/instruments/protocol-additional-geneva-conventions-12-august-1949-and>
 55. United Nations (UN) (2013). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns (A/HRC/23/47, 9 April 2013)*. United Nations. Retrieved December 2023 from <https://digitallibrary.un.org/record/755741?ln=en>.
 56. United Nations (UN) (2019). *Final Report. Meeting of the High Contracting Parties to the 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/MSP/2019/9, 13 December 2019)*. United Nations. Retrieved December 2023 from <https://digitallibrary.un.org/record/3856241?ln=es>.
 57. United Nations (UN) (2026). *Report of the 2024-2025-2026 sessions of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (CCW/GGE.1/2026/CRP.1/Rev.1, 3 September 2026)*. United Nations. [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2026\)/CCW-GGE.1-2026-CRP.1-Rev.1_-_03_September_.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2026)/CCW-GGE.1-2026-CRP.1-Rev.1_-_03_September_.pdf)
 58. Van den Berg, J. (2015). Wat maakt cyber security anders dan informatiebeveiliging? *Magazine Nationale*
 59. *Veiligheid en Crisisbeheersing* 2. Retrieved March 2025 from <https://research.tudelft.nl/en/publications/wat-maakt-cyber-security-anders-dan-informatiebeveiliging>.
 60. Verdiesen, I., Santoni De Sio, F. & Dignum, V. (2021). Accountability and Control over Autonomous Weapons Systems: A Framework for Comprehensive Human Oversight. *Minds & Machines*, 31, 137-163. <https://doi.org/10.1007/s11023-020-09532-9>.
 61. Von Bertalanffy, L. (1972). The history and status of general systems theory. *Academy of Management Journal*, 15(4), 407–426. Retrieved March 2025 from <https://www.marilia.unesp.br/Home/Instituicao/Docentes/RosangelaCaldas/bertalanffy.pdf>.
 62. Wallach, W. (2010). Robot minds and human ethics: the need for a comprehensive model of moral decision making. *Ethics and Information Technology*, 12, 243-250. <https://doi.org/10.1007/s10676-010-9232-8>.
 63. Wallach, W. & Allen, C. (2013). Framing Robot Arms Control. *Ethics and Information Technology*, 15, 125-135. <https://doi.org/10.1007/s10676-012-9303-0>.

64. Watkin, K. (2016). *Fighting at the Legal Boundaries: Controlling the Use of Force in Contemporary Conflict*, Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780190457976.002.0003>.
65. Whitby, B. (2008). Computing Machinery and Morality. *Artificial Intelligence & Society*, 22, 551-563. <https://doi.org/10.1007/s00146-007-0100-y>.
66. Young, J. E., Sharlin, E. & Igarisghi, T. (2011). What is mixed reality anyway? Considering the boundaries of mixed reality in the context of robots. In Wang, X. (ed.), *Mixed Reality and Human Robot Interaction* (pp. 1-11), Springer. https://doi.org/10.1007/978-94-007-0582-1_1.
67. Zhao, P., Liu, K., Zou, H. & Zhen X. (2018). Multi-Stream Convolutional Neural Network for SAR Automatic Recognition. *Remote Sense*, 10(9), 1473. <https://doi.org/10.3390/rs10091473>.

Figures

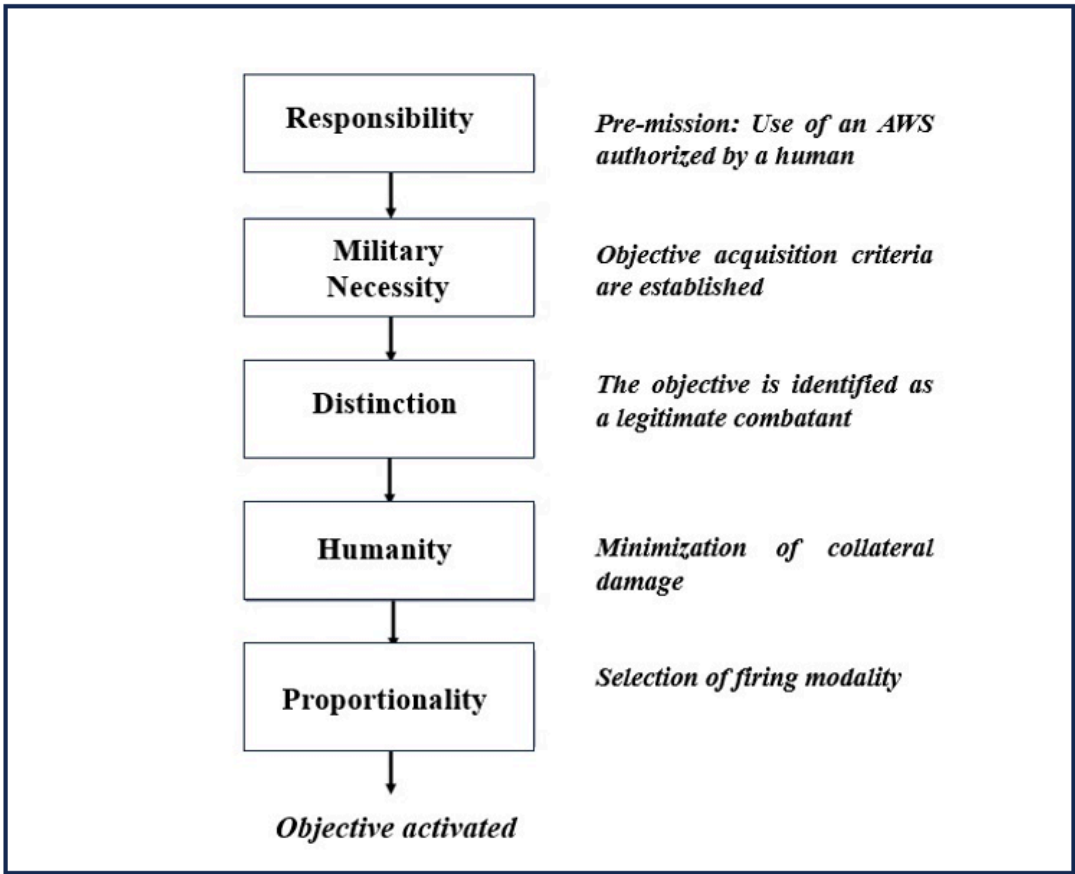


Figure 1

Ethical activation model.

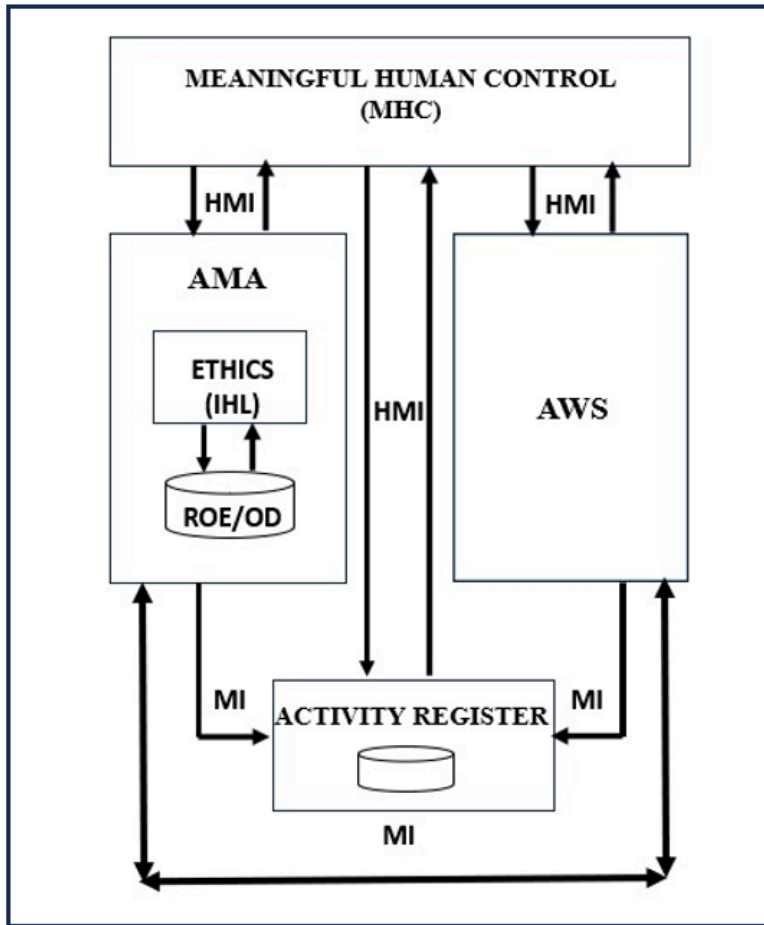


Figure 2

Basic conceptual schema of an AMA/AWS dynamic system with an MHC framework

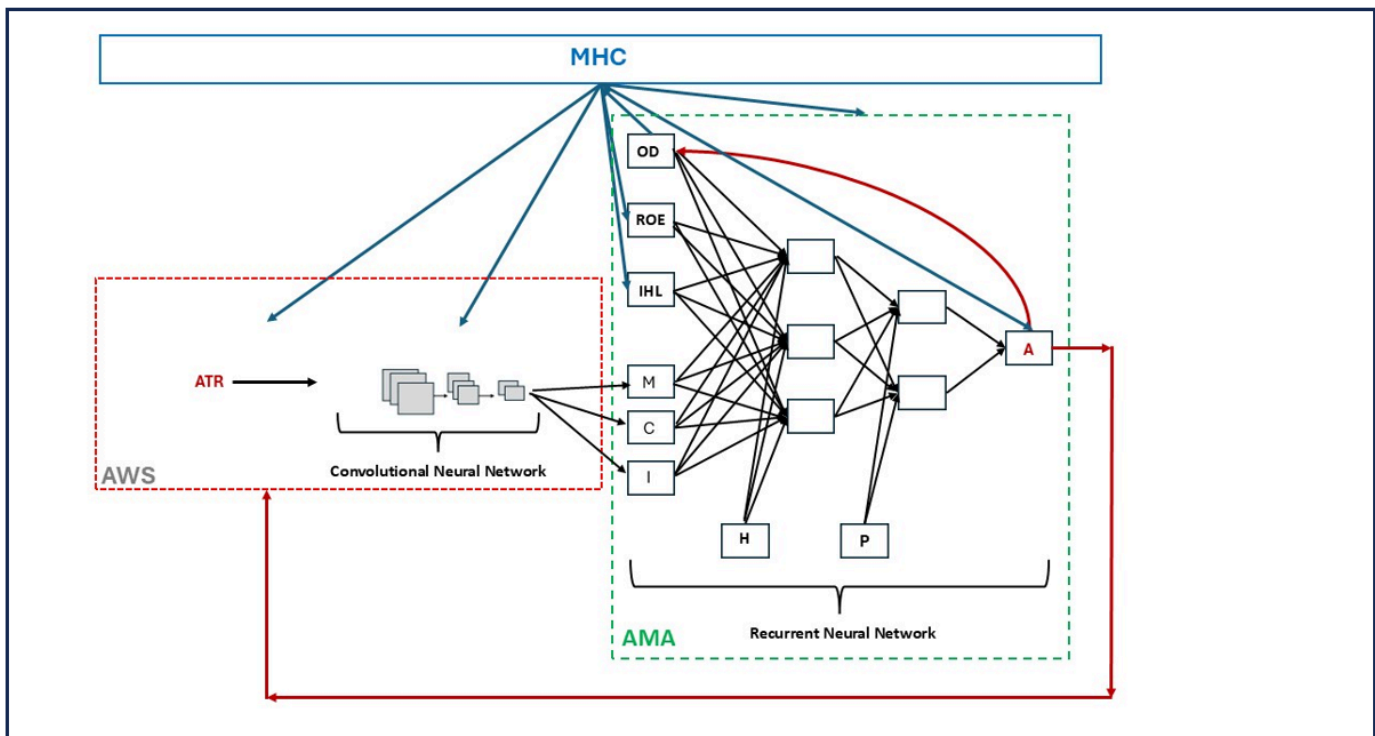


Figure 3

Computational structure of an AMA for AWS, including the complementary CNN, RNN and the MHC framework.

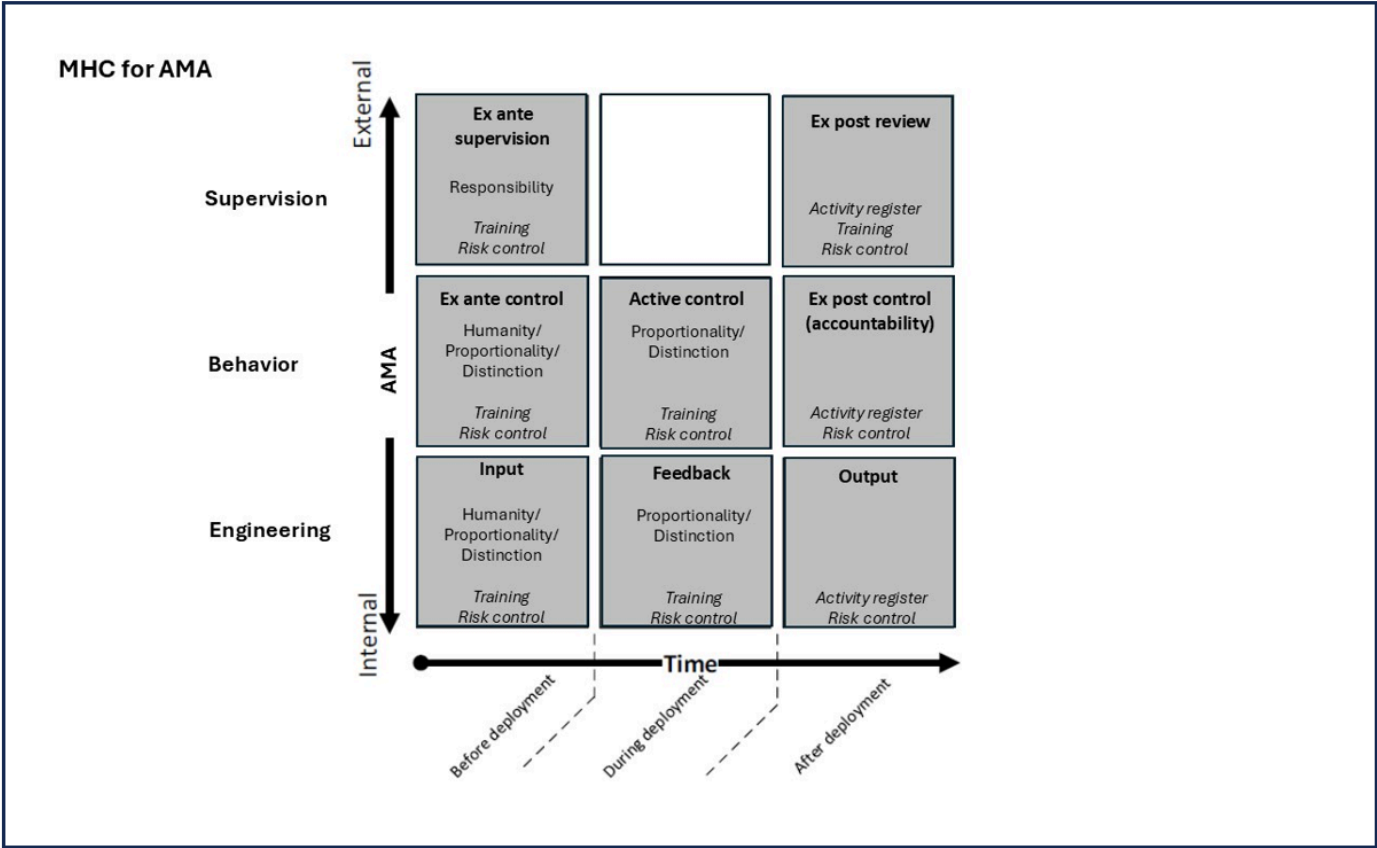


Figure 4

MHC framework for AMA in AWS complex dynamic systems.

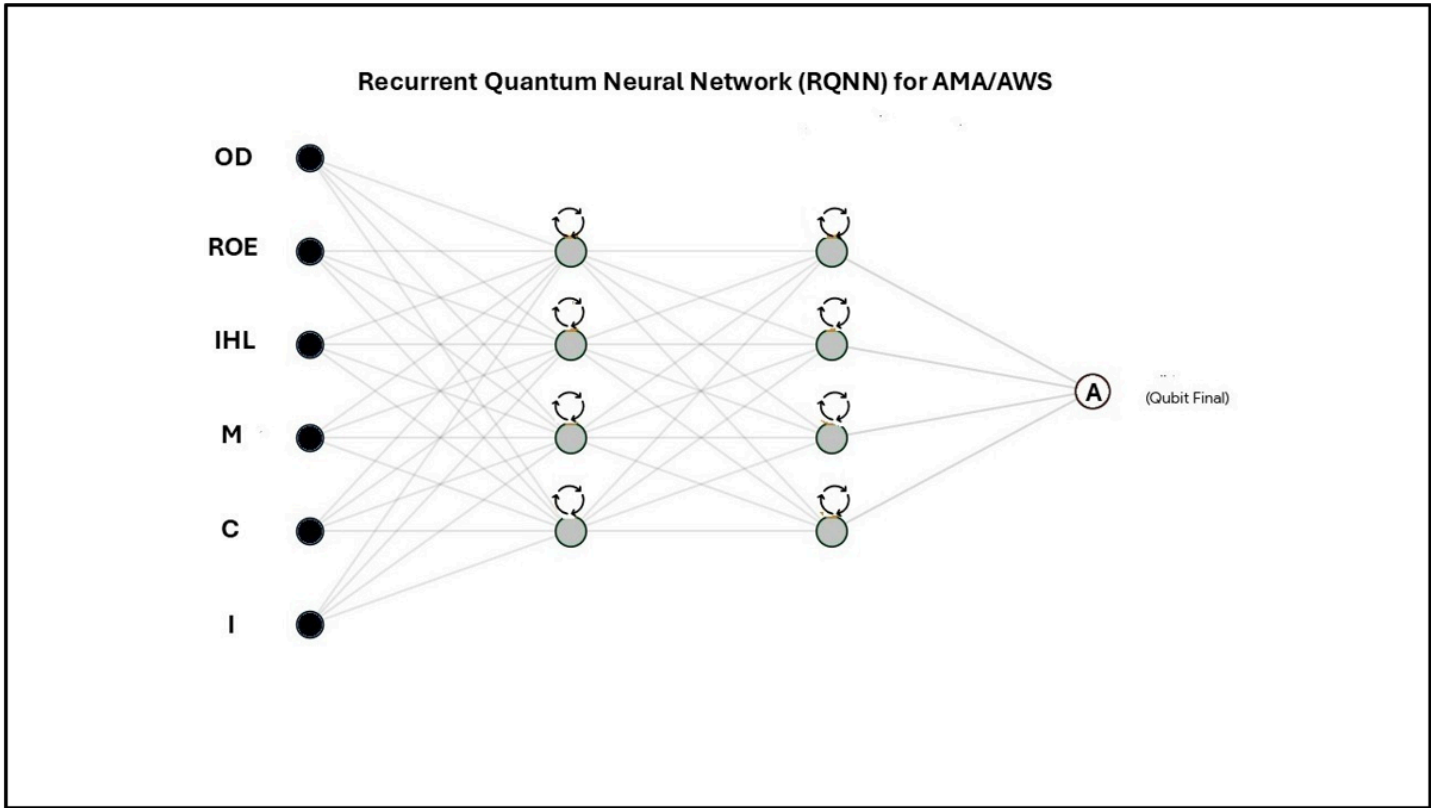


Figure 5

Recurrent quantic neural network for AMA in AWS, adapted from Schuld & Petruccine (Schuld & Petruccine, 2021).

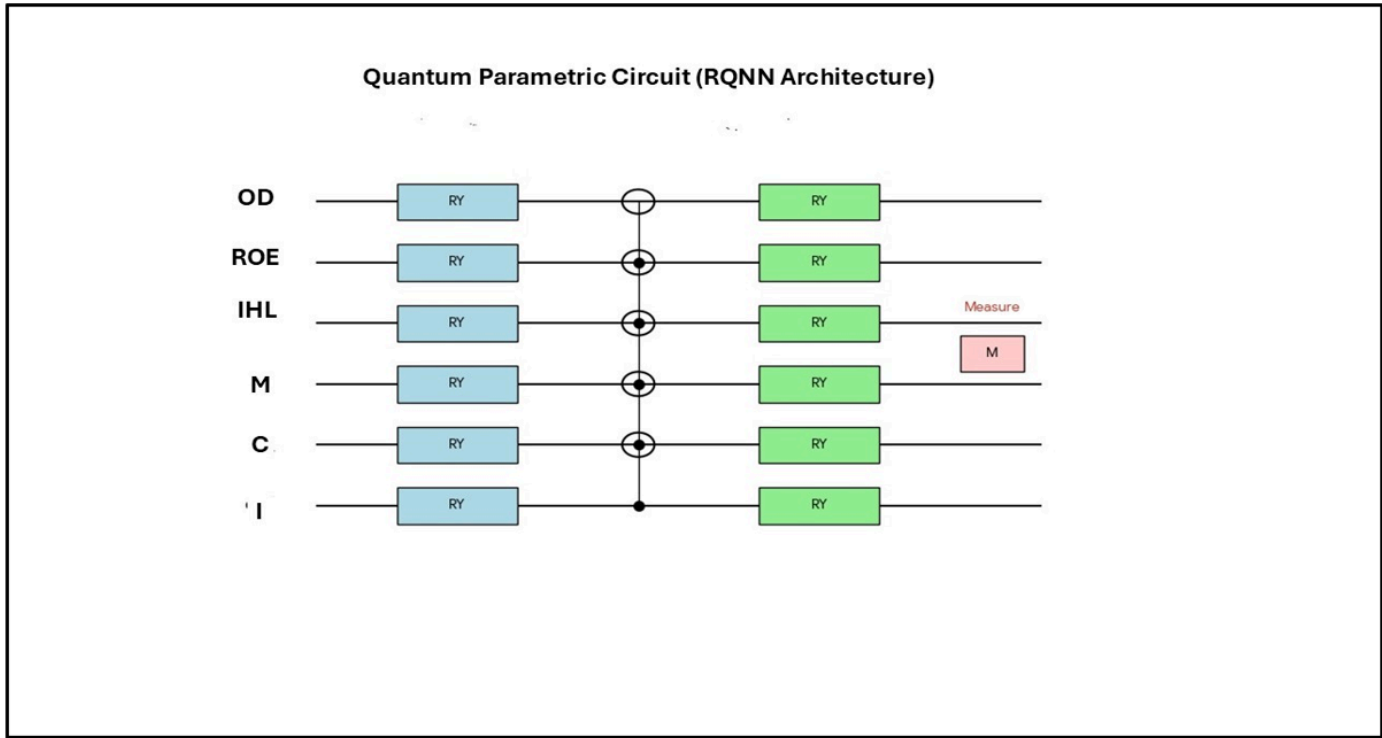


Figure 6

Quantum Parametric Circuit for an AMA in AWS, adapted from P. Sen *et al* (Sen *et al*, 2021).