# An Exploration of Maintaining Human Control in AI Enabled Systems and the Challenges of Achieving It

**Michael Boardman and Fiona Butcher**
Dstl Porton Down
Building 5 - Room G02 - ISAT E
Porton Down
SP4 0JQ
UNITED KINGDOM

mjboardman@dstl.gov.uk

## *ABSTRACT*

*Future operating environments are likely to require the effective integration of humans and artificial intelligence (AI) enabled systems within decision-making processes. As these AI-based systems are adopted it will become increasingly important to ensure that appropriate human control is maintained for big data analytics and decision support systems. Meaningful human control (MHC) can be described as the ability to make timely, informed choices to influence AI-based systems that enable the best possible operational outcomes. There are a number of factors contributing to MHC including freedom of choice and sufficient human understanding of the situation and system. Therefore, to ensure appropriate application of MHC future agile command and control system designs will need to give due consideration to human judgement and decision making requirements. Over the last year the NATO HFM-ET-178 has explored the broad range of human factors issues associated with MHC and has agreed a subset of these to investigate in a follow-on NATO workshop planned for 2020. Although led by the HFM panel, future activities will be intensely cross-panel to produce comprehensive solutions. This paper discusses the importance of maintaining human control within military, AI enabled systems. It is based on work conducted over a 12 month period by members of the NATO HFM178 Exploratory Team with the intent of encouraging future linkages across NATO panels and activities.*

## 1.0 INTRODUCTION

This paper discusses this issues involved in maintaining appropriate human control within military, Robotic, Intelligent and Autonomous (RIA) Systems and other technologies utilising Artificial Intelligence (AI). It is based on work conducted over a 12 month period by members of the NATO HFM178 Exploratory Team [1] with the intent of encouraging future linkages across NATO panels and activities.

The Exploratory Team included chairs: Mark Draper (USAF USA) and Jurriaan van Diggelen (TNO Netherlands); team members: Marcel Baltzer (Fraunhofer Germany), Michael Boardman (Dstl UK), Fiona Butcher (Dstl UK), Robert Shively (NASA USA) and Rogier Woltjer (FOI Sweden) and mentors: Frank Flemisch and Adelbert Bronkhorst.

The use of Robotic, Intelligent and Autonomous Systems within both military and civil applications is becoming increasingly common and it likely that this trend will accelerate as the technologies mature and become more affordable. These systems could be physically embodied as in the case of autonomous vehicles or workplace robots, or exist only as software agents that autonomously process data at speeds, levels of accuracy or for durations that humans are not capable of, provide support to decision makers, or even make decisions in their own right. While these technologies offer potentially significant benefits the use of advanced automation, autonomous systems, AI and Machine Learning raises many complex socio-technical, legal, moral and ethical questions. These concerns are often addressed though the inclusion of human oversight and control of the system. Human control can be achieved through a variety of means ranging from supervising the system and taking over complete control when required through to exercising control as part of a Human Machine Team (HMT).

The demands of the future operating environment in terms of operational tempo and increasing information availability and a desire to reduce the risk to personnel is likely to lead to the more widespread adoption of sophisticated automation, autonomous, and AI based systems within defence applications. However, the physical, informational and political complexity of the future battlespace will require human judgement, adaptability, flexibility and active involvement in decisions that are outside the capabilities of the technology in order to deliver operationally effective, legal and resilient capabilities. Therefore the effective integration of humans and artificial intelligence (AI) enabled systems within data analysis, planning, decision-making and delivery of effects will become increasingly important.

The use of AI, machine learning and robotics is more advanced in many civilian systems than in comparable military applications; as such there are significant lessons that can be learnt from their development, use and evolving legislation. For example in the development of autonomous cars, “controllability” proved to be a critical consideration combining complex technical, ergonomic, legal, moral and organisational factors. Some of these lessons might well be transferred to the military domain; however, there are significant differences between military and civil applications and requirements. While the environment and scenarios of use may be relatively predictable in the civil domain, RIA systems operating in a military context must: withstand adverse and rapidly changing conditions, accommodate imperfect and potentially erroneous data sources, protect against deliberate adversarial actions, deal with ethical implications surrounding the use of force, and remain within bounds set by international humanitarian law (IHL) and rules of engagement. Moreover, military systems must be able to function safely and effectively under a wide range of highly dynamic environments and use cases that are hard to predict or anticipate during the design phase. They must also be resilient to failure and to complex, uncertain and unpredictable events and situations where the dynamics of the military domain necessitate complex judgements regarding acceptable actions based on rules of engagement, international law and judgements over legality, proportionality and risk. Because of this the maintenance of Human Control through a combination of specification, design, training, operating procedures, and assurance processes is seen as critical in many, if not all military systems.

This paper examines:

- The nature of human control within systems utilising AI or Machine Learning technology
- The factors that can contribute to or impede human control in various future military applications of AI-based systems.
- The complexity of ensuring that human control and accountability is delivered and maintained within future systems and Systems of Systems

The paper concludes by highlighting gaps identified for further research and is intended to initiate further discussions and debates in this critical area for defence.

## 2.0 BACKGROUND

The initial objective of the NATO Exploratory Team was to understand what was meant by *Meaningful Human Control (MHC)* within systems utilising AI or Machine Learning technology and develop a shared working definition of what it is.

To guide the development of this understanding MHC was examined from multiple perspectives, these included:

- Authority, responsibility and accountability in the chain of command;
- The type of application in which AI technologies might be used;
- The type of interaction between humans and AI-based systems, such as: human oversight and intervention (e.g., ‘human-on/over-the-loop’), the use of AI-based systems as support tools, Human Machine Teaming (HMT).

Examining MHC from these perspectives highlighted to the NATO Exploratory Team the broad range of factors and complexities of different application areas that need to be taken into account when providing a definition. It also identified the complex legal and political environment associated with the use of AI and Autonomy within systems that are capable of delivering lethal effects.

The concept of MHC has recently gained attention as an important concept during the 2016 expert meetings organised by the UN Convention on “Certain Conventional Weapons” (CCW) and the use of Lethal Autonomous Weapon Systems (LAWS). While the concept of MHC has become linked to autonomous weapons it can be applied much more generally to a broader set of applications [12], including AI-based military applications for planning, decision support and Intelligence, Surveillance, Target, Acquisition & Reconnaissance (ISTAR). This is because these applications may be used within the processes that result in the decision to take actions that may result in the loss of life and/or have legal or moral considerations that require human involvement and therefore is essential maintaining human accountability. Human control can also be seen as a critical component of safeguarding the performance and resilience of such systems as the adaptability and flexibility provided by active human involvement can address situations outside of the capabilities of the technology alone.

Because of the role that MHC can play in delivering effective systems and also the wide range of system types that contribute to the intelligence collection, decision-making, planning and situational understanding that precedes the use of lethal and non-lethal effects it was decided that the ET should examine a wider range of systems and not just those delivering lethal effect.

## 3.0 THE NATURE OF MEANINGFUL HUMAN CONTROL

In order to develop a definition of MHC it was necessary to explore the components that comprise it. It was recognised that many of the issues being raised in relation to human control within Autonomous and AI enabled systems are not new, but share many parallels with those associated with automated systems. For example the following design principles for automated command and control systems were published by NASA in 1995 [2]:

- To command effectively, the human operator must be involved.
- To be involved, the human operator must be informed.
- The human operator must be able to monitor automated systems.

- The automated system must also be able to monitor the human operator.
- Automation systems must be predictable.
- Each of the elements of the system must have knowledge of the other's intent.

This identifies some key themes which feature in many recent papers and guidance documents relating to autonomy [3, 4, 5], such as: human awareness of system state and ability to predict system behaviour, the importance of the user maintaining situational awareness, the need for two-way understanding between human and machine, and shared understanding of each other's objectives. However, there are some concepts that don't feature in this NASA guidance such as *trust* and the concept of the machine acting as a member of a *team*[1] rather than as a *tool*. These two aspects are particularly important in understanding how human control is enabled and how it can be lost.

There are various automation and autonomy taxonomies [6, 7, 8], which provide various insights into the locus and nature of control within automation and autonomous systems, one such example are the levels of automation of decision and action selection [9]; see Table 1.

**Table 1 - Levels of automation of Decision and Action Selection [9]**

| **Low** |
|---|
| 1 The computer offers no assistance; human must take all decisions and actions. |
| 2 The computer offers a complete set of decision/action alternatives from which the human selects and executes. |
| 3 The computer narrows the selection down to a few from which the human selects and executes. |
| 4 The computer  suggests one alternative that the human executes |
| 5 The computer suggests one alternative executes it if the human approves. |
| 6 The computer suggests one alternative and allows the human a restricted veto time before automatic execution. |
| 7 The computer executes automatically, then necessarily informs the human. |
| 8 The computer executes automatically and informs the human only if asked. |
| 9 The computer executes and informs the human only if it, the computer, decides to. |
| 10 The computer decides everything, acts autonomously, ignores the human. |
| **High** |

Whereas these Levels of Automation have proven useful as a broad characterization of options in many applications (such as autonomous vehicles), we consider them as too simplistic for the MHC debate.

[1] The notion of the machine as being referred to as a member of a team is not intended to anthropomorphise the technology, but convey the types of interaction and collaborative activity between human and machine.

Based on a review of the available literature and consideration of a range of potential applications of AI in military systems it was identified that human control over a system is *complex* and *rarely discrete* (simply present of absent), rather it is *dynamic* and *multi-dimensional* and the level of control required is highly situation dependent, and that it is not bound to one particular moment in time. A proposed set of dimensions or characteristics that contribute to human control was developed by the ET. These are described in Table 2.

**Table 2 - Proposed dimensions/characteristics of human control**

| Dimension of Human Control | Example |
|---|---|
| **The human has freedom of choice** | This dimension captures the degree to which the human user can choose from all of the possible courses of action available. At one extreme the human has complete, unconstrained freedom to choose any course of action, including detrimental or illegal ones with no influence or assistance from the system. At the other extreme the machine has constrained the human to a single course of action which they have no freedom to deviate from.<br>The constraining of human freedom of choice can be as a result of:<br>• Direct effects – The system does not make a particular course of action or element of system functionality available to the user.<br>• Indirect effects – Information is inaccessible or incomprehensible by the user preventing an informed, free choice.<br>• Cultural effects - The organisational culture is such that selecting a course of action other than one recommended by the system carries potential blame or users feel that poor outcomes can be blamed on AI recommendations.<br>• Workload effects – high workload prevents users from processing information to make a free choice and has to rely on the system to make the decision for them in totality or in part.<br>It should be noted that not all decisions and actions require complete freedom of choice; in fact it may not be desirable in some situations due to the speed at which decisions must be made or where workload means that humans are unable to process all of the information available to them and consider all of the potential courses of action or options. |

| | |
|---|---|
| **The human has ability to impact the behaviour of the system** | This dimension captures the extent to which the user is provided with the functionality to change the behavior of the system. This could be in real time, or in advance through the setting of bounds or constraining allowable actions and behaviors.<br>At one extreme the user has complete control over system behavior e.g. manual control over a UAV, ability to override an autonomous system.<br>At the other extreme there is no functionality built into the system to allow the user to change the behavior of the system. |
| **The human has time to decide to engage with the system and alter its behaviour.** | This dimension captures the temporal aspect of user interactions with the system i.e. does the system allow the user sufficient time to process information, make decisions and impact on its behavior if required.<br>At one extreme no time constraints are placed on the user to make decisions or act on the system.<br>At the other extreme the system processes information and executes courses of actions at a speed beyond the capabilities of the user. This has been referred to in some quarters as machine speed warfare.<br>There are situations where it is not feasible or desirable for the human to be “in-the-loop” e.g. defensive systems, so they have to be involved “before-the–loop” e.g. by setting constraints on action, to ensure that human control is maintained |
| **The human has sufficient situation understanding** | This dimension describes the extent to which the human has sufficiently accurate situational understanding to make an informed choice.<br>At one extreme the user is presented with a completely accurate situational representation of the real world.<br>At the other extreme the user is provided with no situational information at all. |

| | |
|---|---|
| **The human has sufficient system understanding** | This dimension captures the degree to which a human has a sufficient understanding of the system state, in order to understand the provenance, quality and accuracy of the information and the rationale of the decisions and recommendations made.<br>At one extreme the user has a complete and accurate understanding of the system state, capabilities, the information that has informed the actions that the system takes and/or its behavior, and how the system uses information to determine its actions.<br>At the other extreme the user has no understanding of the system state or how it is making decisions. |
| **The human is capable to predict the behaviour of the system and the effects of the environment (physical and information)** | This dimension captures the extent to which the user is able to predict how the system will behave in different circumstances.<br>At one extreme the user has a complete and accurate understanding of how the system will behave in response to any given different inputs and/or conditions.<br>At the other extreme the user has no understanding of how the system will behave in any situation. |

## 3.1 Factors that could influence human control of Defence AI-based Systems

As a result of a review of the available literature and applying it to the defence context the ET determined that there were multiple dimensions and concepts wrapped up within the notion of *Meaningful Human Control*. As a result the ET felt that it was more helpful to consider Human Control of AI enabled systems more broadly under the notion of Appropriate Human Control. This recognises that the degree and nature of human control is dynamic, highly context dependent and made up of multiple factors which are more or less important depending on the situation and nature of the system and potential benefits and consequences provided by the system.

As a result the ET proposes two complimentary and interlinked components of *Human Control* within AI enabled and Autonomous systems; these are: *Meaningful Human Control* and *Effective Human Control* (see Figure 1).

***Meaningful Human Control*** **(MHC)** encompasses the following elements:

- **Legal** – MHC is required to prevent violation of International Humanitarian Law and maintenance of accountability of military action (in both physical and information domains).

- **Moral / Ethical** – MHC provides the moral and ethical dimension to military decision making and military action (in both physical and information domains).

***Effective Human Control*** **(EHC)** is required because humans and machines working towards a common goal can be better than either working in isolation. Effective human control within human machine teams can be an:

- **Enabler of Performance/effectiveness** – EHC (human involvement in decision making) is an enabler of improved operational outcomes

- **Enabler of Risk Reduction** – EHC (human involvement in decision making) reduces risk of undesirable outcomes and is a critical enabler of system resilience.

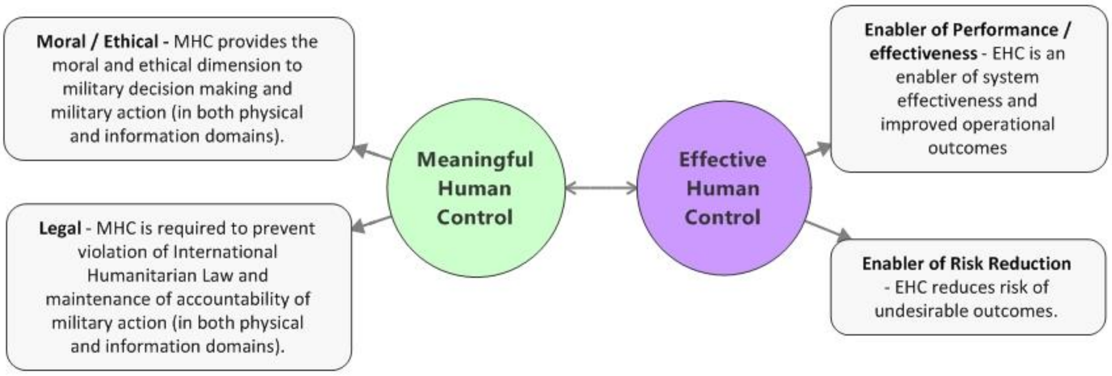
**Moral / Ethical -** MHC provides the moral and ethical dimension to military decision making and military action (in both physical and information domains).
**Legal -** MHC is required to prevent violation of International Humanitarian Law and maintenance of accountability of military action (in both physical and information domains).
Meaningful Human Control
Effective Human Control
**Enabler of Performance / effectiveness -** EHC is an enabler of system effectiveness and improved operational outcomes
**Enabler of Risk Reduction** - EHC reduces risk of undesirable outcomes.

**Figure 1 Meaningful and Effective Human Control**

These two aspects need to be considered together as a focus on one at the expense of the other may have negative outcomes. For example a focus on effectiveness and optimising system performance could inadvertently impact on meaningful control and vice versa excessive involvement of users in some decision types could lead to delays, biases or errors that could impact on operational effectiveness.

## 3.2 A Description of Meaningful Human Control

As a result of the investigations conducted and in order to avoid potential political sensitivities associated with the evolving LAWS debate it was decided that a working description, rather than a formal definition would be more helpful in order to support further study.

Based on the factors identified the following description of HC was agreed upon by the members of the ET:

> *'Humans have the ability to make informed choices in sufficient time to influence AI-based systems in order to enable a desired effect or to prevent an undesired immediate or future effect on the environment.'*

## 3.3 Achieving Meaningful and Effective Human Control

The ET also considered how systems might be designed to deliver an appropriate level of Human Control. This included consideration of both the processes and stakeholders involved in the fielding of systems that enable MHC and EHC so that suitable guidance and support can be developed.
In order to field systems that deliver Effective and Meaningful Human Control there are a wide range of activities that will need to contribute, including:

- National or Organisational Policy

- Systems specification

- Systems design

- Systems V&V
- Training of Users
- Training of AI and ML
- Systems of Systems (SoS) integration
- C2 process development
- Interoperability
- Operational Use
- After action review and lessons learned

The span of these activities will require the involvement of multiple stakeholder groups to ensure that appropriate human control is delivered by fielded systems. See Table 3.

**Table 3 – Stakeholder groups that contribute to ensuring appropriate human control in fielded systems**

| Stakeholder Groups | Role in Delivering Appropriate Human Control in Future Systems |
|---|---|
| **Policy Makers** | • Set policy, standards and doctrine to facilitate EHC and MHC across NATO systems. |
| **R&D / Scientific Community** | • Identify and fill knowledge gaps in the field of EHC and MHC.<br>• Conduct research to provide evidence to support the specification and design of systems that support MHC and EHC. |
| **System Acquirers** | • Specify, contract and accept AI enabled systems that support EHC and MHC. |
| **System Designers** | • Conduct and analysis, use Human Centred Design approaches such as those described in ISO 9241 210 [0], and apply best practice to systems design and testing to develop systems that support the delivery of EHC and MHC. |
| **Organisational Users** | • Integrate AI enabled systems within their wider system of systems and organisational structures such that they enable EHC and MHC. |
| **End users** | • Train, support and employ human machine teams that deliver EHC and MHC.<br>• Develop operating procedures and practices to deliver EHC and MHC. |

These activities and stakeholders conducting them will need to be supported by evidence based guidance, tools, standards and policies including:

- Enablers and Tools:

    - Common language and terms.
    - Models of MHC and EHC.
    - Tools to assess and measure MHC and EHC within Systems / SoS and organisations.
    - EHC risk analysis tools.
    - Evidence base for MHC and EHC and associated research literature.
    - HC Use Cases and examples.
- Guidelines and Standards:
    - Human Systems Integration (HSI) and acquisition approaches for MHC and EHC.
    - MHC and EHC Best Practice/Exemplars.
    - Human Machine Teaming Guidance.

## 3.4 Use of Human Control Dimensions to Support the Delivery of Appropriate Human Control within Fielded Systems

An initial step in this process was to identify whether the dimensions identified align to proposed approaches to assuring human control within Autonomous and AI enabled systems. A mapping of the ET dimensions of human control to the iPRAW requirements for human control [11] was conducted to determine the extent to which these aligned and how the two approaches might support the design of these types of system, see Table 4.

**Table 4: Application proposed dimensions of ET developed Dimension of human control to iPRAW's Requirements for Human Control in the Use of Force [11].**

| | **Situational Understanding** | **Intervention** |
|---|---|---|
| **Control by Design (Technical Control)** | iPRAW<br>Design of systems that allows human commanders the ability to monitor information about environment and system | iPRAW<br>Design of systems with modes of operation that allow human intervention and require their input in specific steps of the targeting cycle based on their situational understanding |
| | Application of Dimensions of Human Control<br>• The system design allows the Human to develop sufficiently accurate situation, and system awareness/understanding to identify risks to violating IHL and/or unacceptable moral, ethical or operational outcomes.<br>• The system design allows the Human to predict the behavior of the system and its effects on the environment (physical and information). | Application of Dimensions of Human Control<br>• The system design allows the Human to impact on the behaviour of the system in time to prevent an undesirable act (violating IHL and/or unacceptable moral, ethical or operational outcomes). |

| | | |
|---|---|---|
| **Control in Use (Operational Control)** | iPRAW<br>Appropriate monitoring of the system and the operational environment | iPRAW<br>Authority and accountability of human operators, teammates and commanders; abide by IHL |
| | Application of Dimensions of Human Control<br>• The Human has sufficiently accurate situation, and system awareness/understanding to identify risks to violating IHL and/or unacceptable moral, ethical or operational outcomes.<br>• The Human is able to predict the behaviour of the system and its effects on the environment (physical and information). | Application of Dimensions of Human Control<br>• The Human is able to exercise freedom of choice and has the ability to affect system behaviour during use to ensure that accountability and adherence to IHL are maintained.<br>• Training with systems allows users to understand and predict system behaviours across different situations in order to avoid undesirable outcomes or failures to comply with IHL.<br>• Organisational culture does not indirectly impact on freedom of choice and willingness to question system behaviours and actions. |

### 3.5 Complexity of Human Control

Considering the nature of human control within single user - single systems can be complex in its own right, but this complexity increases significantly when wider systems of systems aspects are considered. Interactions between multiple AI enabled systems, Humans, and Human Machine Teams are likely to lead to complex and unanticipated emergent behaviours and systems properties, see Figure 2. This complexity increases the potential for Human Control to be lost with associated detriments in system performance and with risks of accountability, legal and moral issues arising. There is also a risk that Human Control at an individual system level is lost at a system of systems level.

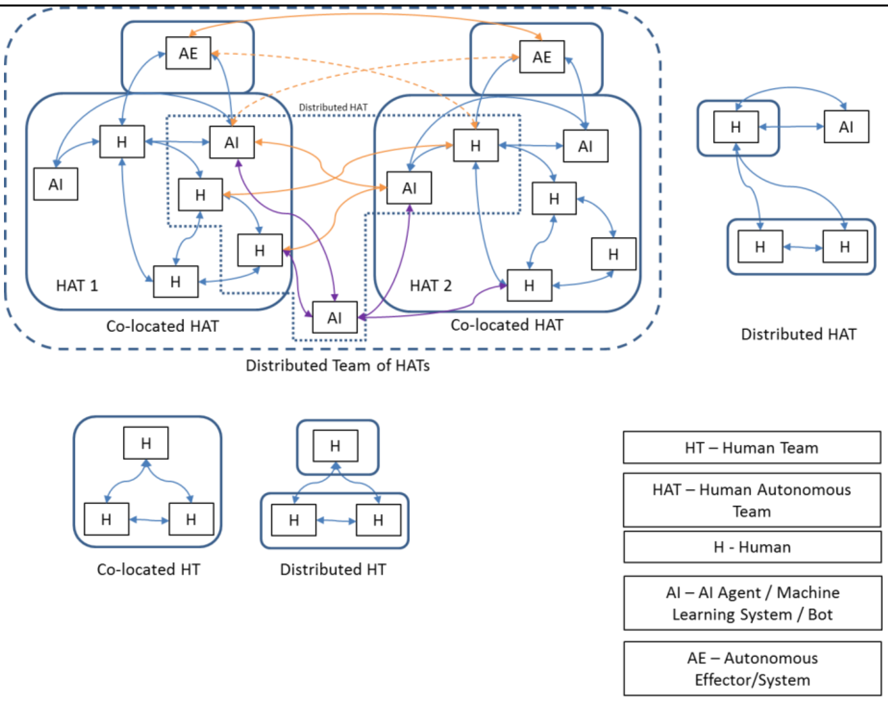

**Figure 2 Complexity of multiple interacting elements in achieving, maintaining human control.**

An added complexity is that systems may vary in functionality and potential impact on operational outcomes depending on system state, warfighting posture, rules of engagement, and wider the environmental context (e.g. where communications are constrained or denied). The breadth of potential system behaviours across all of these system behaviours these systems states needs to be understood and considered both within design and planning for deployment to ensure that Appropriate Human Control is maintained and human accountability is not lost within this complexity.

## 3.6 What are the Challenges and Complexities for NATO Regarding Human Control of AI Enabled Systems?

Bringing together the exploratory work conducted the ET identified a number of gaps in knowledge and complexities surrounding human control which require further investigation and research. Including the following:

- Further exploration and validation of the dimensions of human control.
- Methods of establishing the appropriate level of human control required by a system within a specific context and what factors affect this, such as as: the speed/tempo required to achieve effect; confidence (trust) in the output/behaviour of the AI Enabled System; potential consequence of action; the level of risk that is considered acceptable; potential operational benefits.
- How human control can be measured/evaluated.
- Whether appropriate human control can be guaranteed at design time, or is it something that must be continuously managed after deployment.
- Methods for delivering human control in complex systems of systems.
- Methods for delivering human control in multi-national or multi-agency organisations.
- Methods for tracing human accountability through complex systems of systems and identifying accountability gaps.

To address some of these gaps the ET recommended that a new RTG should be established together with a dedicated workshop to further investigate the most pressing influencing factors regarding HC including:

- Organisational Considerations of HC (including team training, agile C2 structures).
- HF guidelines to achieve and maintain HC for all NATO AI applications.
- Systems Engineering methods (including TEV&V for learning systems) to support HC
- Adversary tactics to counter/undermine HC (methods and mitigations).
- MHC for complex socio-technical system of systems (emergent properties, HC propagation).
- Legal, ethical, political, and public perception of HC over AI-based systems.

The proposal for a dedicated workshop in this area was accepted and a workshop (HFM-322 Meaningful Human Control of AI-based Systems: Key Characteristics, Influencing Factors and Design Considerations) will be held in 2020. Although led by HFM, the workshop is intended to be cross-panel and inter-organisational in order to produce comprehensive solutions which draw on the wide range of relevant research being conducted across NATO. Because MHC spans many scientific topic areas across several disciplines but will focus on particularly on:

- Human-autonomy interaction
- Explainable AI
- Teaming research/ distributed intelligence/team structures and roles
- Shared dynamic mental models, maintenance of common ground
- Dynamic task allocation
- Directability and predictability with AI
- Observability/transparency and trust
- Organisational influences and processes
- Joint human-machine decision making/biases
- Joint human-machine learning/training
- Accountability and MHC
- Ethics and morality
- Complexity research

## 4.0 CONCLUSIONS

- Human control over AI enabled systems is complex and rarely discrete (simply present of absent), rather it is dynamic and multi-dimensional and the level of control required is highly situation and time dependent.
- While MHC has become linked to autonomous weapons it can be applied much more generally to a broader set of military applications where AI based systems are used, such as planning, decision support and ISTAR systems. This is because these applications may be used within the processes that result in the decision to use lethal effects and/or have legal or moral considerations that require human involvement and therefore is essential maintaining human accountability. Human control can also be seen as a critical component of safeguarding the performance and resilience of such systems as the adaptability and flexibility provided by active human involvement can address situations

outside of the capabilities of the technology alone.

- The concept of Appropriate Human Control in AI enabled systems may be more useful to support AI enabled systems design than Meaningful Human Control. But determining the degree of Appropriate Human Control within a system requires consideration of both: Meaningful Human Control (associated with legal, moral and ethical aspects of human control) and Effective Human Control (associated with Human Machine Interaction and Human Machine Team performance and system effectiveness). These two aspects need to be considered in unison as a focus on one at the expense of the other may have negative outcomes.
- Achieving and Assuring human control and accountability in complex Systems of Systems and multi-national organisations will be challenging. Therefore, understanding and managing the risks to achieving appropriate human control can only be achieved through analysis of all aspects of the system design and interactions within it (human-human, human-machine, machine-machine and multiple teams of human machine teams).
- Multiple stakeholder groups and processes will be involved in ensuring that appropriate human control is delivered in fielded AI enabled systems.

## 5.0 RECOMMENDATIONS FOR NEXT STEPS

The ET recommended that a dedicated workshop should be held to further investigate the most pressing influencing factors regarding MHC. This proposal was accepted and a workshop will be held in 2020.

Given the breadth of AI related research being conducted that may impact on the achievement of MHC it is recommended that an integrating Team should be formed to serve as the centre of a dynamic hub-and-spoke structure between teams so as to enhance cross-team communications, leverage efforts, and maximise overall progress towards integrated MHC solutions across NATO.

## 6.0 REFERENCES

[1] TECHNICAL ACTIVITY PROPOSAL HFM-ET-178 Meaningful Human Control Over AI-based Systems (2018).

[2] Charles E. Billings (1991) Human-Centered Aviation Automation: Principles and Guidelines. NASA Technical Memorandum 110381.

[3] NATO INDUSTRIAL ADVISORY GROUP NIAG SG.233 FINAL REPORT ON HUMAN MACHINE TEAMING (2019). NIAG-D(2019)0006.

[4] Andrew P. Williams and Paul D. Scharre (eds) (2015) Autonomous Systems Issues for Defence Policymakers. ISBN 9789284501939.

[5] Joint Concept Note 1/18: human machine teaming (2018) UK MOD Development, Concepts and Doctrine Centre.

[6] Endsley, M.R. & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42, 462-492.

[7] Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30, 286–297.

[8] Save, L. and Feuerberg, B. (2013) Designing Human-Automation Interaction: a new level of Automation Taxonomy.

[9] Sheridan, T.B., & Verplank, W. (1978). Human and Computer Control of Undersea Teleoperators. Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT.

[10] Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (2019) ISO 9241-210:2019.

[11] iPRAW (2017) “Focus On” Report No. 1 :Focus on Technology and Applications of Autonomous Weapons.

[12] Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: a philosophical account. Frontiers in Robotics and AI, 5, 15.