LOST IN TRANSLATION: BUILDING A COMMON LANGUAGE FOR REGULATING AUTONOMOUS WEAPONS

*Marc C. Canellas and Rachel A. Haga**

There have been three UN meetings of experts in 2014, 2015, and 2016 to address autonomous weapons systems (AWS), however, little to no progress. In this article, we argue that the fundamental reason for the stalled discussions is the lack of a unifying, technical language for describing and understanding the problems posed by AWS. A unifying, technical language would address two major communications issues facing the discussants of AWS: an inability to identify the sources of the conflict and solutions that have consensus, and an inability to operationalize the regulations that are agreed upon. We propose that the language of cognitive systems engineering can be the unifying technical language to provide initial answers to the four key questions at the UN: (1) How do we define autonomy? Use the requirements for effective function allocation to develop standards for human-AWS interaction and meaningful human control. (2) What amount or quality of human control necessary for lawful use of AWS? Use function allocation's models and metrics to evaluate human-AWS interaction and enforce meaningful human control standards. (3) What would an accountability framework look like for AWS? Use the models and metrics for evaluating authority-responsibility mismatches in function allocation to address the AWS responsibility gap. (4) How do we review and certify permissible AWS? Use the human-automation issues that have been explored and addressed by function allocation to develop case studies and technical standards for human-AWS interaction.

* Marc Canellas is a Ph.D. candidate in aerospace engineering at the Georgia Institute of Technology and Rachel Haga is a research engineer at the Georgia Institute of Technology. Both are members of the Cognitive Engineering Center, a research lab that examines human-system integration in complex work environments from theoretical and methodological viewpoints, in the field and in the laboratory. Email: marc.c.canellas@gatech.edu; rachel.haga@gatech.edu.

© IEEE 2016. M. Canellas and R. Haga, "Lost in Translation: Building a Common Language for Regulating Autonomous Weapons," in IEEE Technology and Society Magazine, vol. 35, no. 3, pp. 50-58, Sept. 2016. doi: 10.1109/MTS.2016.2593218. Available at SSRN: <https://ssrn.com/abstract=2927663>

TABLE OF CONTENTS

Rise of Autonomous Weapons Systems	2
Call for International Regulations of Autonomous Weapons Systems	2
The Language Barrier	3
A. Identifying Sources of Conflict and Consensus	3
B. Operationalizing Regulations	4
Addressing the Four Issues of Autonomous Weapons Systems through Cognitive Systems Engineering	5
A. How do we define autonomy? Should we?	6
B. What amount or quality of human control is necessary for lawful use of AWS?	6
C. What would an accountability framework look like for AWS?	9
D. How do we review and certify permissible autonomous weapons systems?	9
The Importance of a Technical Language	10
Cognitive Systems Engineering: The Language of Autonomous Weapons Systems Regulations	11
Bibliography	12

RISE OF AUTONOMOUS WEAPONS SYSTEMS

Autonomous weapons systems (AWS) are already here. Though some of the colloquial names for AWS may suggest science fiction (killer robots [1], [2], terminators [3], and cyborg assassins [3]), these systems are anything but fiction. Since the 1970s the U.S. Navy's 'Phalanx' Close-In Weapon System has been capable of "autonomously performing its own search, detect, evaluation, track, engage and kill assessment functions" against high-speed threats such as missiles, ships, aircraft, and helicopters [4]. Not limited to the U.S., Germany has developed a similar land vehicle defense system, the Active Vehicle Protection System, which has a reaction time of less than 400ms when launching fragmentation grenades against incoming missiles [5].

AWS are possible due to the convergence of new technology supply and well-established military demand [6]. The drivers of military demand can be summed up as force multiplication, expanding the battle-space, extending the warfighters' reach, and casualty reduction [7]. As for technology supply, over the past three decades, sensors and transmitters have decreased in cost while increasing in functionality. As a result, AWS sit at the intersection of novel automation capable of making decisions without humans and established lethal weapons.

CALL FOR INTERNATIONAL REGULATIONS OF AUTONOMOUS WEAPONS SYSTEMS

With the rise of AWS, there has been an increasingly sophisticated and coordinated response from states and non-governmental organizations calling for international regulations. The impetus for an international discussion began in 2010, when the International Committee for Robot Arms

© IEEE 2016. M. Canellas and R. Haga, "Lost in Translation: Building a Common Language for Regulating Autonomous Weapons," in IEEE Technology and Society Magazine, vol. 35, no. 3, pp. 50-58, Sept. 2016. doi: 10.1109/MTS.2016.2593218. Available at SSRN: <https://ssrn.com/abstract=2927663>

Control (ICRAC) published the Berlin Statement which expressed concern for the loss of human control in the governance of lethal force and the conduct of war.

From this statement, in just a few years, two distinct sides have coalesced: those in favor of banning AWS (pro-ban) and those in favor of allowing and regulating AWS (pro-regulation). Pro-ban supporters argue that AWS without human control will never be capable of adhering to the laws of armed conflict (e.g. [1], [2], [8–10]) whereas the pro-regulation supporters argue that AWS are already here and *are* capable or, at least, *can be* capable of adhering to the laws of armed conflict – potentially even better than humans (e.g. [6], [11–15]). The debate has escalated rapidly, resulting in the 121 states parties of the UN Convention on Certain Conventional Weapons (CCW)¹ hosting three consecutive years of Meeting of Experts on Lethal Autonomous Weapons Systems in 2014, 2015, and 2016.

Although no regulations have been agreed upon, there has been a general consensus that international regulations of AWS are needed [16].² As summarized by Crootof [16], there are many benefits to international regulations of AWS: channeling research and state practice by limiting and defining what lawful options are available, improving the likelihood that AWS will be used in accordance with the law, ensuring different states' actions are evaluated using the same standard, among many others.

THE LANGUAGE BARRIER

Despite the three consecutive years of meetings, the discussions of AWS at the UN have made little to no progress on the four key issues identified at the first meeting in 2014: (1) how to define autonomy, (2) the amount or quality of human control necessary for lawful use of AWS, (3) how to establish an accountability framework for AWS, and (4) how review and certify permissible AWS [17], [18]. Importantly, these four issues encapsulate the breadth of what international regulations would have to address. So until the discussions can make progress on these four issues, there will be no international regulations even though there is consensus for regulations.

The fundamental reason for the stalled discussions is the lack of a unifying, technical language for describing and understanding the problems posed by autonomous weapons systems. A unifying, technical language would address two major communications issues facing the discussants of AWS: an inability to identify the sources of the conflict and solutions that have consensus, and an inability to operationalize the regulations that are agreed upon.

A. Identifying Sources of Conflict and Consensus

With so many nations, non-governmental organizations, and outside experts discussing their opinions and proposals, it is no surprise the identifying sources of conflict and consensus has been a consistent problem for the discussions of AWS. One particularly difficult discussion has been

¹ The UN CCW is the international treaty that bans or restricts the use of land mines, blinding lasers, and other weapons that have the potential to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately. Currently a total of 121 States are parties to the CCW with a further five having signed but not yet ratified. [16]

² There are a few skeptics of the benefits of international regulations who argue that states do not need to take any formal action and that the law of armed conflict will naturally evolve over time to address ambiguities [26].

© IEEE 2016. M. Canellas and R. Haga, "Lost in Translation: Building a Common Language for Regulating Autonomous Weapons," in IEEE Technology and Society Magazine, vol. 35, no. 3, pp. 50-58, Sept. 2016. doi: 10.1109/MTS.2016.2593218. Available at SSRN: <https://ssrn.com/abstract=2927663>

the issue regarding the amount or quality of human control necessary for lawful use of AWS, more well known as 'meaningful human control' [19]. In previous work [20], we analyzed two proposed definitions of meaningful human control from Article 36 [21] and ICRAC [22]. Even though both organizations are in favor of banning fully-AWS and mandating meaningful human control for lawful use of AWS, the prose of the definitions make it entirely unclear what types of human control they agree and disagree on. Article 36 required a human operator to have "adequate contextual information" whereas ICRAC required "full contextual and situational awareness." Article 36 required "a positive action by a human operator" to initiate attack whereas ICRAC required "active cognitive participation in the attack."

Ultimately, in order to converge on a single set of regulations, the various proposals from different groups must be laid out in such a way to be easily compared. As more people join the conversation, they continue to bring new arguments and proposals. While the breadth could potentially make the AWS regulation discussions very rich, to date many of the proposed arguments seem to be talking past one another. New proposals should modify or build off of current proposals, highlighting the conflicts and differences. The lack of a common language with technical specificity makes it unclear what conflicts need to be addressed and what consensus can be built upon in the future.

B. Operationalizing Regulations

Once conflicts have been addressed and a consensus has been converged upon, there remains the issue of operationalization: how do we write the international regulations such that they are implementable and certifiable? A summary of the progression of meaningful human control from ethical, to regulatory, to technological language shows that more work is needed to operationalize proposals. Many have made the ethical arguments that having an AWS make the decision to kill a human violates the human's right to life (only humans can make the necessary qualitative assessments), right to a remedy for unlawful action (AWS cannot be held directly accountable for their actions), and the principle of dignity (inanimate machines which cannot truly understand the value of life) [8]. These concerns were translated into regulatory language by requiring a mandate that there always be meaningful human control over autonomous weapons systems so that the human will be making the qualitative assessments, be held accountable for actions, and be valuing the human life [21]. Specifically, the proposed regulations specify that the human operator must always have 'a means for the rapid suspension of attacks' and 'sufficient time for deliberation' [22]. Only when the mandate gets to the technologist for implementation is it realized that the proposed regulations outlaw numerous weapons systems that are already used without controversy [19], are not operationalized enough to be implemented [23], nor would they likely ensure meaningful human control if they were implemented.

If the proposals for meaningful human control – or any other regulations – were implemented in their current state, without being operationalized, it would actually undermine the intentions of the initial ethical concerns that started the discussions. The ambiguity would leave many of the design decisions, such as how to interpret what exactly is meant by 'adequate contextual information' or 'sufficient time for deliberation' in the hands of each government or manufacturer. Not only would this result in an inconsistent application of the regulation as it is highly unlikely that any two designers, much less two nations, would interpret them in the same way, it leaves the

door open for an intentionally malicious interpretation to the letter of the law, while ignoring the spirit. Furthermore, there are no systematic ways to certify systems based on these proposals.

ADDRESSING THE FOUR ISSUES OF AUTONOMOUS WEAPONS SYSTEMS THROUGH COGNITIVE SYSTEMS ENGINEERING

We propose that the language of cognitive systems engineering is uniquely poised to be the unifying, technical language of the meeting of experts on autonomous weapons systems. The four key issues of definitions, human control, accountability, and certification for human-autonomous systems may be relatively new to the international community, but, these issues have been topics of research for cognitive systems engineers for over 30 years [24]. There are two general synergies between cognitive systems engineering and the regulatory discussions of autonomous weapons systems. First, autonomous weapons systems fit within the established domain of cognitive systems engineering which is focused on complex, sociotechnical, and safety-critical systems that are dependent on human-technology interaction to perform successfully,³ including military test and evaluation [25], aviation [26], and human space operations [27]. Second, whereas regulatory discussions are looking for methods of analysis, design, and evaluation, cognitive systems engineering already has standards and techniques for modeling and measuring performance of complex, sociotechnical, and safety-critical systems like autonomous weapons systems.

Developing international regulations of autonomous weapons is clearly a difficult problem but the stalled discussions should be the motivation for looking to new perspectives that make it easier to identify the sources of the conflict and solutions that have consensus, and to operationalize the solutions that are supported. We believe that to make progress on the four key issues of definitions, human control, accountability, and certification, discussants at the UN should be utilizing the synergetic potential of cognitive systems engineering:

1. **How do we define autonomy? (Should we?)** Using the requirements for effective function allocation to develop standards for human-AWS interaction and meaningful human control.
2. **What amount or quality of human control necessary for lawful use of AWS?** Using function allocation's models and metrics to evaluate human-AWS interaction and enforce meaningful human control standards.
3. **What would an accountability framework look like for AWS?** Using the models and metrics for evaluating authority-responsibility mismatches in function allocation to address the AWS responsibility gap.
4. **How do we review and certify permissible AWS?** Using the human-automation issues that have been explored and addressed by function allocation to develop case studies and technical standards for human-AWS interaction.

³Definition adapted from <http://cec.gatech.edu/> and [28].

A. How do we define autonomy? Should we?

The most obvious question poised to those looking to regulate AWS is “What makes an autonomous weapons system autonomous?” This seemingly simple question is actually incredibly complex. The most referenced definition, a three-level classification, was proposed in 2012 by the U.S. Department of Defense (DoD) [28]. However, the DoD definition of AWS is not internationally agreed upon. A number of other definitions have been proposed [16], [19].

The difficulty of defining the ambiguous concept of autonomy has led some nations to suggest that the CCW should not seek a consensus on the definition at this point because discussions are impeding the discussions [17]. The difficulty of defining autonomy is by no means limited to the AWS community; defining autonomy has plagued the cognitive systems engineering discipline for decades. However, despite the head start, the discipline has similarly failed to converge on a single definition of autonomy. Even the most well-cited classifications, such as Parasuraman et al.’s 10 ‘Levels of Autonomy’ [29], are now acknowledged to be limited, problematic, and worth discarding altogether [30], [31].

The reason that the cognitive systems engineering community has moved away from an isolated focus on autonomy is that the focus had created seven so-called “myths of autonomy” [31]: (1) autonomy is unidimensional; (2) levels of autonomy are a useful scientific grounding for the development of autonomous system roadmaps; (3) autonomy is a widget or discrete component that can be added or removed from a system without affecting the rest of the system; (4) autonomous systems are autonomous, in that they are capable of performing competently in every task and situation; (5) full automation obviates the need for human-machine collaboration; (6) as machines acquire more autonomy, they will work as simple substitutes (or multipliers) of human capability; and (7) full autonomy is not only possible but always desirable.

In light of the myths of autonomy, our response to “how should we define autonomy” is that we shouldn’t define autonomy without defining the human-automation team. This holds for regulators and designers alike, because only by understanding the human-automation team can the effectiveness of the autonomous system be measured.

B. What amount or quality of human control is necessary for lawful use of AWS?

As a response to the focus on autonomy, one of the few convergent themes at the UN CCW meetings has been that AWS should be designed to ensure meaningful human control (MHC) [23], an organizing principle “that those who plan or decide on an attack have sufficient information and control over a weapon to be able to predict how the weapon will operate and what effects it will produce in the context of an individual attack, and thus, to make the required legal judgements” [21]. Similar to the discussions of defining autonomy which have been too automation-centered, the discussions of meaningful human control have been too human-centered.

The amount or quality of human control necessary for lawful use of AWS is dependent on the design of the human-AWS team. To address this dependency, cognitive systems engineering has shifted to a holistic view of human-automation teams through function allocation which determines how to allocate work within teams of human and automated agents (e.g. a human

operator and an AWS) [30]. Function allocation is a deep and growing field in which researchers have presented requirements for effective function allocation [30], methods for modeling function allocation [32] and metrics for evaluating function allocation [33]. The five requirements for effective function allocation defined by Feigh and Pritchett [30] introduce a new way to think about regulating the human-AWS teams:

1. Each agent must be allocated functions that it is capable of performing.
2. Each agent must be capable of performing its collective set of functions.
3. The function allocation must be realizable with reasonable teamwork.
4. The function allocation must support the dynamics of the work.
5. The function allocation should be the result of deliberate design decisions.

The most important characteristic of these requirements is that they are performance-based, not ethics-based. The allocation of functions can span the full spectrum of weapons operations from completely removing the human operator, to a human completely integrated into every phase of the decision making process, to any combination of human-AWS teamwork in-between. If the overall goal of the team was to uphold the laws of war, the requirements do not overtly prescribe whether or where a human operator should be within the team, just that the laws of war should be upheld. From these requirements, research programs could and should be developed to identify what types of effective function allocations for human-AWS teams are capable upholding the laws of war.

In designing human-AWS teams such that the operations are lawful, there are many potential problems that must be accounted for in design. Table 1 relates the most well-known problems with human-automation teams to the design considerations that can attempt to alleviate or even solve the problems. The problems and design solutions were adapted from Feigh and Pritchett's [30] review of the function allocation literature which has explored human-automation interaction through ethnographies, human-subjects studies, and computer simulations.

To the question of "what amount or quality of human control is necessary for lawful use of AWS," the function allocation response is: the necessary amount or quality of human control is that which enables each agent within the team to be able to perform their task or tasks with reasonable work within the dynamics of the scenario.

Table 1. Problems and Design Solutions for Human-Automation Teams

Potential Problems	Design Considerations
Brittle automation: where emergencies must be handled by humans and there is little support for off-nominal conditions	Design functions around both the automation and the human's capabilities.
High workload spikes during off-nominal situations [34] and excessively low workload during normal operations in between the spikes leading to humans becoming out of the loop [34], [35].	
Leftover allocation (automate as many functions as technology will permit, and assume the human will pick up whichever functions are leftover) often results in a human assigned to monitor automation or the environment for conditions beyond which the operate; functions in which humans are ineffective [36]	
Gap between authority and responsibility: automation is assigned to execute a function in an operational sense but human will be held accountable in an organizational and legal sense for its outcome. [37]	Consider authority-responsibility
Without being able to assess whether automation is correct, humans often overtrust or undertrust the automation [38]	
Automation and/or human unable to anticipate each other's information needs and provide information at useful, non-interruptive times [39]	Consider automation as a team member.
Clumsy automation which interrupts its team members because it cannot implicitly sense information about whether other team members would benefit from an interruption [40]	
In complex work environments where many functions interdependent and coupled, some couplings may be hidden. Unaccounted-for couplings can result in insufficient coordination, idling, and workload accumulation.	Dynamic analysis of the physical environment. Function allocation should support how the work environment is managed by the agents.
Adaptive function allocation (dynamically changing function allocation) can aggravate environmental unpredictability.	Examine the tradeoff between maintaining predictability for the human versus applying complex automated capabilities and dynamically allocating functions [41]

C. What would an accountability framework look like for AWS?

One of the main concerns with AWS is the apparent ambiguity of who should be responsible for the deaths caused by AWS [1], [2], [8], an issue which has been termed the responsibility gap [13]. There have been claims that the use of AWS would inevitably result in unacceptable responsibility gaps (e.g. operators and commanders, programmers, and manufacturers escaping liability) [1]. Despite these claims, most nations at the UN were confident in their ability to develop an accountability framework for AWS [17]. Furthermore, Müller argues that we already live with responsibility gaps in civilian life and that we should focus on narrowing the responsibility gaps, not establishing an untenable requirements of responsibility in all cases [13].

Narrowing responsibility gaps is already a focus of cognitive systems engineers. Specifically, function allocation is used to model and measure responsibility gaps in proposed and implemented human-autonomous systems by characterizing authority-responsibility mismatches. Authority describes which functions an agent is asked to perform and responsibility describes which outcomes an agent will be accountable for in an organizational, regulatory or legal sense [30]. A synthesis of the function allocation literature makes a clear argument for a general regulation of autonomous weapons systems: “the responsibility for the outcome of a function must be considered relative to the authority to perform it” [33].

One well-studied mismatch comes from the modern commercial airline cockpits where the human flight crew “maintain responsibility for the safety of the flight, even as the autopilot and autoflight systems exercise significant authority over important actions within the aircraft control and trajectory management functions” [32]. In the interest of minimizing the number of mismatches, Pritchett et al. [32], [33] developed conceptual models of various function allocations between the human flight crew and the autopilot systems and counted the number of mismatches. Then, to understand the effect of the various function allocations within different contexts, the models were computationally simulated to determine which one minimized human workload.

Those concerned about the AWS responsibility gaps ought to use these methods developed in the cognitive systems engineering literature to identify what functions should be allocated to humans in order to establish the desired responsibilities.

D. How do we review and certify permissible autonomous weapons systems?

At the 2016 UN CCW meeting on AWS, almost every nation that addressed weapons reviews, endorsed the idea that the review process was critical in ensuring the lawful use of autonomous weapons [17]. In developing such procedures for reviewing and certifying permissible AWS, the experts should leverage the extensive modeling and measuring techniques for function allocation which have been used extensively to analyze human-automation interaction.

There are two main categories of models for understanding the effectiveness of the distribution of work between agents: static work models and dynamic simulations [32]. Static work models assess the distribution of work in a context-independent manner whereas dynamic simulations focus on how the coupled interactions between agents is affected by the context of the environment, usually through computational simulations. Importantly these models should be developed through significant analysis of the work domain via qualitative and quantitative studies

of the operations (most likely, cognitive work analysis).

We will focus our discussion on two static work models as they are more applicable to the AWS discussion at this time because we are not aware of any computational analyses of human-AWS interaction.⁴ One static work model is the abstraction hierarchy which provides an understanding of how an agent may be able to view work at different levels of abstraction, ranging from a detailed descriptions of a specific task (manage decision to engage) to mission goals (ensure safety of the convoy) and modeling teamwork as defined by function allocation and team design. A second model is to identify and list the functions allocated to different agents within a team (human operator and AWS) [32], a method we refer to as function allocation form. In previous work [20], we used function allocation form to identify and list the distribution of allocated functions between the human operator and AWS the three classes of AWS as defined by the DoD [28] and two definitions of MHC [22], [42]. Our analysis of the function allocation form showed that if either MHC standard were adopted in the laws of armed conflict, all levels of the DoD's AWS would be illegal because of one specific difference in the definitions: whether the human (required by MHC) or the AWS (as classified by DoD) has control over engaging targets.

Once the models are completed, they can be evaluated using at least eight different metrics [33]: 1) workload, 2) stability of the work environment, 3) mismatches between responsibility and authority, 4) incoherency in function allocations, 5) interruptive automation, 6) automation's boundary conditions, 7) function allocations limiting human adaptation to context, 8) and mission performance. Theoretically, it would be possible to evaluate the currently proposed definitions of AWS and MHC using these metrics to determine how the function allocations actually relate to the performance of the human-AWS team.

THE IMPORTANCE OF A TECHNICAL LANGUAGE

Our argument for using a technical language for discussing AWS is not new nor is it without detractors. When discussing the definitions of AWS, Scharre and Horowitz [43] argued that autonomy should be described along three dimensions: the human-machine command-and-control relationship, the complexity of the machine, and the type of decision being automated. They concluded their descriptions by stating, "These are all important features of autonomous systems, but they are different ideas, and often people tend to mix them together."

Crootof [16] rebutted these three dimensions of autonomy, stating that, "evaluating autonomy along these different axes introduces a confusing and unnecessary level of particularity... Although these gradations may be extremely useful in research and development, they are unnecessarily precise for a legal document that is ultimately concerned with regulating weaponry that might independently exercise lethal force."

While more work should be done to address any confusing aspects of technical explanations of AWS, we argue that particularity and precision are *absolutely necessary* for understanding and regulating AWS. Just a cursory review of the incidents and accidents related to human-automation team problems can show that well-intentioned designs cause serious harm [34], [38], [44].

⁴ There are computational analyses of function allocation in other complex work domains such as pilot-autopilot interaction in commercial aviation [37], [38].

© IEEE 2016. M. Canellas and R. Haga, "Lost in Translation: Building a Common Language for Regulating Autonomous Weapons," in IEEE Technology and Society Magazine, vol. 35, no. 3, pp. 50-58, Sept. 2016. doi: 10.1109/MTS.2016.2593218. Available at SSRN: <https://ssrn.com/abstract=2927663>

Cognitive systems engineering and its related disciplines have spent decades trying to unmask the causes of the harm and, as shown in the paper, developed some important techniques for designing effective teams such that harm is mitigated. This is why the fifth requirement for effective function allocation explicitly states that function allocation (or cognitive systems engineering, generally) should be a part of the design process. Therefore, if the ethicists and regulators want designers to address meaningful human control or responsibility gaps in their designs, then the regulations will eventually need to be translated into particular and precise technical language like cognitive systems engineering.

Another benefit of adopting technical languages is that they are often agnostic to the ethical challenges being debated. In a ‘pro-regulation world’, where a fully AWS is legal, regulating their design according to the requirements of function allocation would still work because the AWS would be required to perform its functions and support the dynamics of the work – including upholding the laws of war. Alternatively, in a ‘pro-ban world’ the arguments of requiring meaningful human control can be integrated into the requirements. For example, should the UN adopt a proposal of meaningful human control such that only human operators can perform the function of engaging targets, then the functions should be allocated to the human and AWS such that both can perform their functions with reasonable teamwork within the dynamics of the work.

More broadly, if the regulators do not understand the real complexities of the systems they are trying to regulate then it is easy for those who are constructing and fielding the technology to, either deliberately or unintentionally, discover loopholes in the regulations. This concern has been termed the “pacing problem” to reflect the growing gap between the pace of science and technology and the lagging responsiveness of legal and ethical oversight society relies on to govern emerging technologies [45]. This pacing problem is not unique to AWS as it is affecting the areas of biotechnology, genetic testing, nanotechnology, computer privacy, and many others. However, for AWS in particular, our legal and ethical discussions cannot shirk the complexities of the human-AWS teams because that could result in insufficient regulations potentially enabling the use of fully AWS under the guise of meaningful human control.

COGNITIVE SYSTEMS ENGINEERING: THE LANGUAGE OF AUTONOMOUS WEAPONS SYSTEMS REGULATIONS

The language of cognitive systems engineering, its principles, models, and measures, have been developed over decades to address the concerns over designing, analyzing, and evaluating human-autonomous teams. Our drive for showing how the cognitive systems engineering language can address the four key issues of UN CCW discussions of AWS (defining autonomy, human control, accountability, and review and certification) is our worry that the international community may pass regulations of AWS that do not result in the intended consequences. If an insufficient definition of MHC is adopted, AWS could be fielded which pass the regulatory tests but which only have a façade of human control determining life or death – leaving the human operator responsible for actions they could not be expected to control. Furthermore, if an AWS is incorrectly classified and insufficiently regulated, the determinants of responsibility for the actions of the AWS will be too ambiguous – fulfilling the concern that operators, programmers, or manufacturers who should be found responsible will escape liability.

We intend for this paper to differentially affect two sets of communities. First, we hope that it will provide the experts, participants, and researchers that are already part of the AWS debates and discussions with new language for articulating their ethical concerns and proposals for regulations. Using cognitive systems engineering as a common language should enable identification of the sources of the conflict and solutions that have consensus, and operationalize the regulations that are agreed upon. Second, we hope that the cognitive systems engineers⁵ will see the value of their perspectives and expertise to the ethical and legal communities, where regulating human-automation teams are becoming both more important and more difficult – and thus start to engage in the debates and discussions themselves.

There are many avenues for future work at the intersection of technology regulation and cognitive systems engineering. For the specific problem of human-AWS teams we intend to identify what function allocations are necessary for a human-AWS team if the direct human operator is responsible for upholding the laws of war. Moreover, the synergy between cognitive systems engineering and the AWS discussions are only an example of the synergies between cognitive systems engineering and the regulation of robotics and automated systems generally. We believe there is potential for the formal integration of cognitive systems engineering as a partner to the growing field of “robot law” [46].

BIBLIOGRAPHY

- [1] B. L. Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*. Human Rights Watch, 2015.
- [2] B. L. Docherty, *Losing Humanity: The Case Against Killer Robots*. Human Rights Watch, 2012.
- [3] D. Garcia, “The Case Against Killer Robots: Why the United States Should Ban Them,” *Feogn Affairs*. Foreign Affairs: Council on Foreign Relations, May-2014.
- [4] U. S. Navy, “United States Navy Fact File: MK 15 - PHALANX CLOSE-IN WEAPONS SYSTEM (CIWS).” U.S. Navy, Nov-2013.
- [5] DIEHL, “AVePS: Active Vehicle Protection System - Protection of medium and heavy armored vehicles against RPGs and ATGMs.” Apr-2016.
- [6] K. Anderson, D. Reisner, and M. C. Waxman, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*, vol. 90, pp. 386–411, 2014.
- [7] G. E. Marchant, B. Allenby, R. Arkin, E. T. Barrett, J. Borenstein, L. M. Gaudet, O. Kittrie, P. Lin, G. R. Lucas, R. Oâ€™Meara, and others, “International governance of autonomous military robots,” *The Columbia Science and Technology Law Review*, vol. 12, no. 7, pp. 272–315, 2011.
- [8] B. L. Docherty, *Shaking the Foundations: The Human Rights Implications of Killer Robots*. Human Rights Watch, 2014.
- [9] A. Krishnan, *Killer robots: legality and ethicality of autonomous weapons*. Ashgate Publishing, Ltd., 2009.
- [10] N. Sharkey, “Death strikes from the sky: the calculus of proportionality,” *Technology and Society Magazine, IEEE*, vol. 28, no. 1, pp. 16–19, 2009.
- [11] R. Arkin, “Lethal Autonomous Systems and the Plight of the Non-combatant,” *AISB*

⁵ This includes all those related to cognitive systems engineering regardless of their labeled discipline: psychologists, engineers, designers, and scientists, etc.

- Quarterly*, no. 137, 2013.
- [12] R. C. Arkin, "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture," Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, 2011.
 - [13] V. C. Müller, "Drones and Responsibility: Legal, Philosophical, and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons," E. D. Nucci and F. S. de Sio, Eds. London: Ashgate, 2016, pp. 1–16.
 - [14] V. C. Müller and T. W. Simpson, "Killer Robots: Regulate, don't ban," *Blavatnik School of Government Policy Memo*, pp. 1–4, 2014.
 - [15] T. W. Simpson and V. C. Müller, "Just War and Robots' Killings," *The Philosophical Quarterly*, vol. 66, no. 263, pp. 302–322, Apr. 2016.
 - [16] R. Crotoft, "The Killer Robots Are Here: Legal and Policy Implications," *Cardozo Law Review*, vol. 36, pp. 1837–1915, 2014.
 - [17] C. Ford and C. Jenks, "The International Discussion Continues: 2016 CCW Experts Meeting on Lethal Autonomous Weapons," *Just Security*, Apr. 2016.
 - [18] S. Knuckey, "Governments Conclude First (Ever) Debate on Autonomous Weapons: What Happened and What's Next," *Just Security*, May 2014.
 - [19] M. C. Horowitz and P. Scharre, "Meaningful Human Control in Weapons Systems: A Primer," Center for a New American Security, Mar. 2015.
 - [20] M. C. Canellas and R. A. Haga, "Toward Meaningful Human Control of Autonomous Weapons Systems through Function Allocation," in *IEEE International Symposium on Technology and Society (ISTAS 2015)*, 2015.
 - [21] "Key areas for debate on autonomous weapons systems," Article 36, May 2014.
 - [22] D. Garcia, "ICRAC statement on technical issues to the 2014 UN CCW Expert Meeting," in *CCW Meeting of Experts on Lethal Autonomous Weapons Systems*, 2014.
 - [23] "The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward," United Nations Institute for Disarmament Research (UNIDIR), 2014.
 - [24] L. G. Militello, C. O. Dominguez, G. Lintern, and G. Klein, "The Role of Cognitive Systems Engineering in the Systems Engineering Design Process," *Systems Engineering*, vol. 13, no. 3, pp. 261–273, 2010.
 - [25] J. Ockerman, D. Roberts, A. Sciarretta, D. Folds, S. McKee, R. M. Reynolds, and R. Grier, "Opportunities for Human Factors Measures in Military Operational Test and Evaluation," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2010.
 - [26] F. H. Hawkins and H. W. Orlady, *Human Factors in Flight*. Ashgate Publishing Co., 1993.
 - [27] M. J. Miller, K. M. McGuire, and K. M. Feigh, "Decision Support System Requirements Definition for Human Extravehicular Activity using Cognitive Work Analysis," *Journal of Cognitive Engineering and Decision Making*, 2016.
 - [28] U. S. D. of Defense, "Directive 3000.09: Autonomy in Weapon Systems," United States of America: Department of Defense, Nov. 2012.
 - [29] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, May 2000.
 - [30] K. M. Feigh and A. R. Pritchett, "Requirements for Effective Function Allocation A Critical Review," *Journal of Cognitive Engineering and Decision Making*, vol. 8, no. 1, pp. 23–32, 2014.

- [31] J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods, "The Seven Deadly Myths of "Autonomous Systems," *IEEE Intelligent Systems*, vol. 13, pp. 2–9, 2013.
- [32] A. R. Pritchett, S. Y. Kim, and K. M. Feigh, "Modeling Human–Automation Function Allocation," *Journal of Cognitive Engineering and Decision Making*, vol. 8, no. 1, pp. 33–51, 2014.
- [33] A. R. Pritchett, S. Y. Kim, and K. M. Feigh, "Measuring Human-Automation Function Allocation," *Journal of Cognitive Engineering and Decision Making*, vol. 8, pp. 52–77, 2014.
- [34] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.
- [35] M. R. Endsley and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 2, pp. 381–394, 1995.
- [36] R. W. Bailey, *Human performance engineering: A guide for system designers*. Prentice Hall Professional Technical Reference, 1982.
- [37] D. D. Woods, "Cognitive technologies: The design of joint human-machine cognitive systems," *AI magazine*, vol. 6, no. 4, p. 86, 1985.
- [38] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, no. 2, pp. 230–253, 1997.
- [39] E. E. Entin and E. B. Entin, "Measures for evaluation of team processes and performance in experiments and exercises," in *Proceedings of the 6th International Command and Control Research and Technology Symposium*, 2001, pp. 1–14.
- [40] K. Christoffersen and D. D. Woods, "Advances in human performance and cognitive engineering research," E. Salas, Ed. Bingley, UK: Emerald Group, 2002, pp. 1–12.
- [41] C. A. Miller and R. Parasuraman, "Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 49, no. 1, pp. 57–75, 2007.
- [42] "Killer Robots: UK Government Policy on Fully Autonomous Weapons," Article 36, Apr. 2013.
- [43] P. Scharre and M. C. Horowitz, "An Introduction to Autonomy in Weapons Systems," Center for a New American Security, 2015.
- [44] T. B. Sheridan and R. Parasuraman, "Human-automation interaction," *Reviews of human factors and ergonomics*, vol. 1, no. 1, pp. 89–129, 2005.
- [45] G. E. Marchant, B. R. Allenby, and J. R. Herkert, *The growing gap between emerging technologies and legal-ethical oversight: The pacing problem*, vol. 7. Springer Science & Business Media, 2011.
- [46] R. Calo, M. Froomkin, and I. Kerr, *Robot Law*. Edward Elgar Publishing, 2016.