ISTOCK/SHOCK

*The CCW Debate and Its Implications for VSD*

# Meaningful Human Control of Lethal Autonomous Weapon Systems

Thea Riebe, Stefka Schmid, and Christian Reuter

The debate on the development and deployment of lethal autonomous weapon systems (LAWS) as an emerging technology is of increasing importance, with discussions stalling and technological development progressing. Monitoring the progress of increasingly autonomous weapons systems in civilian and military use (1) as well as regulating possible autonomous systems early on is demanded by civil society actors, like the Campaign to Stop Killer Robots and the International Committee of the Red Cross (ICRC), while nation states follow a variety of interests and strategies, showing little room for consensus on central terms and questions (2), (3). This article therefore sheds light on the work of the Group of Governmental Experts (GGE) of the UN Convention of Certain Conventional Weapons (CCW). The CCW, offering an arena for international cooperation, has dedicated itself to the purpose of finding common ground with respect to an understanding of LAWS, as well as to the necessary degree of human control. From an ethical perspective, the concept of Meaningful Human Control (MHC) supports a

*Digital Object Identifier 10.1109/MTS.2020.3031846*
*Date of current version: 2 December 2020*

human-centric approach. Several IEEE projects, series and publications [4] are dedicated to this prioritization, especially regarding civilian use. As autonomous technology is increasingly at the center of contemporary military innovations, questions of (human) agency and responsibility in warfare have become even more pressing [5]. As stressed by the United Nations Institute for Disarmament Research (UNIDIR), the concept of MHC may prove useful in the context of development and use of (semi-) autonomous weaponry [6].

Acknowledging the need for a multidisciplinary approach [7], we present our analysis of the respective CCW discourse as a first step towards answering the question of how to ensure MHC in the interaction with LAWS. Asking for factors supporting the implementation of MHC, we look for values and underlying, more abstract discourses which may converge with the idea of MHC. The identification of such values and dominant discursive narratives may contribute to a better understanding of the political, ethical, legal, and technological requirements for MHC, a concept which has been introduced in both political and legal debates to allow for improved regulation of the use of force in armed conflicts [8]. The regulation of certain conventional weapons considered particularly injurious or indiscriminative, such as landmines and blinding lasers, is part of the protocols of the UN CCW. However, in the past, the effectiveness of the CCW has been called into question in light of Cold War politics [9], consensual decision-making, or difficulties in advancing arms control in the context of humanitarian and military arguments [10]. Beside these difficulties, the CCW has helped set relevant norms of arms control in the past and may do so with regards to the regulation of LAWS [11].

First, we give an overview of related work and identify research gaps. In the following, we elaborate our theoretical perspective, inspired by Value-Sensitive Design (VSD). Further, we offer insights into the research design, i.e., our sample of 43 CCW documents and discourse analysis as the method used. Subsequently, the results of the analysis are presented. Our focus lies on the CCW's prevalent understanding of autonomy and LAWS as well as on control, particularly concerning MHC in human-computer interaction. These conceptual clarifications are necessary foundational work for retaining human control of LAWS. Then, we give an overview of identified values with respect to the interaction with LAWS technology, putting them into relation with MHC as well as with each other. To gain a deeper understanding of the crucial values, we illustrate correlations between respective values and dominant discourses, which may not always support an implementation of MHC. The work is concluded with a formulation of implications, a discussion of results, and an outlook.

## Related Work

Scholars of various disciplines are dedicated to LAWS and the question of human control. Most work focuses on autonomous weapons from a legal perspective with regard to international humanitarian law (IHL). Crootof [12], focusing on the applicability of international humanitarian law and accountability, reflects on the inherent imprecision of the concept of MHC, while stressing the need to interpret the evolving norm as convergent with existing international law. Anderson [13]–[15] elaborates on the applicability of the law of armed conflict on emerging technologies and explicitly argues against a ban of LAWS. Walker-Smith [16], concentrating on MHC, criticizes a human-biased view by pointing out that autonomous weapon systems can potentially limit lethal human behavior. Walsh [17] notices a concomitant shift of accountability with the development of LAWS, increasingly including designers and programmers. Yet, these legal debates are often characterized by repetitive arguments, echoing difficulties regarding legal accountability while stressing the merit of existing international humanitarian law principles.[1]

Ethical perspectives stress that the technological advances towards more autonomous functions and systems, which are increasingly interacting with humans, need to be designed with more attention towards responsibility [18], [19], control [20], and effects on human dignity [21], thus arguing to limit the possible consequences on the lives of affected people [22], [23]. Scholars of ethics agree that LAWS run counter the principle of human dignity as only humans can be moral agents that can take life-affecting decisions and be held accountable [21], [24]. Sharkey [25] introduces a classification system of human supervisory control, which is adapted by Amoroso and Tamburrini [24], putting it into relation with the concept of MHC, as well as by Suchman and Weber [26], who focus on autonomy of human-machine configurations. Ekelhof [27], [28] asks for the feasibility of operational implementation of MHC, suggesting other ideas like "distributed control" to be more practical. It can be identified that the discourse has moved towards discussing autonomous functions, especially regarding the target selection and target engaging process [24], [29].

Scholars of international relations and strategic security [30]–[33] have discussed the strategic consequences of autonomy. On one side, autonomous systems are a central part of the network-centric warfare (NCW) doctrine, which plans for autonomous weapon systems to increasingly assist humans as human-machine teaming [34], [35]. However, the autonomy of armed systems is increasing

[1]In contrast, Sehrawat [86] takes a minor position in the debate, arguing IHL principles to be ineffective for regulating LAWS.

the pressure or even the likelihood of a first strike due to reduced conflict threshold, which might destabilize nuclear deterrence. However, others like Cummings [36] refer to the advantages of automated systems over humans to carry out attacks due to the likelihood of human error under the stressful conditions of combat.

To further the progress of work towards a limitation of harm, it is important to take a closer look at computer science and engineering, especially robotics. Linked publications are frequently interested in the development of semi-autonomous drones [37], [38], machine learning techniques [39], and human-computer interaction. Often focusing on civilian environments, some scholars pay special attention to disruptive situations [40] or warfare technology [41]. Still, many engineering studies are interested in optimizing automatic or autonomous processes and robotics, disregarding ethical questions or highlighting the potential of LAWS [42]–[45].

**Ethically, lethal autonomous weapons run counter the principle of human dignity as only humans can be moral agents that can take life-affecting decisions and be held accountable.**

From the perspective of machine ethics, Canellas *et al.* [46] investigate the "mismatch between authority and responsibility in an exemplar military scenario (which) can still plague the human-AWS interactions." The interaction of the autonomous system with its complex environment during mission-related tasks is another challenge for researchers. Thus Hägele *et al.* [47] introduce a real-time environmental situation risk assessment approach to improve the safe situational behavior of the autonomous system. Beside the value of safety, Chmielewski [48] tries to incorporate non-Western values and stresses the need for an ethical evaluation of the use of LAWS, referring to IEEE's "Ethical Considerations in Artificial Intelligence and Autonomous Systems" [49]. Others focus on norm change initiated by countries of the Global South [50] or gendered perspectives on autonomous weapon systems [51], [52]. The cognitive engineering approach [53] by Canellas *et al.* [46] is one of the few works that has analyzed different understandings of MHC and concrete options, realized in human-computer interaction. The authors highlight implications for function allocation to autonomous systems vis à vis human operators, derived from definitions of MHC [46]. Yet, interested in establishing MHC in warfare human-machine interaction, they disregard important questions regarding software and interface design and take a less critical stance by assuming definitions to be exogeneous.

Value-Sensitive Design (VSD) as a "theoretically grounded approach to the design of technology" serves to investigate the discourse on LAWS and thus helps to fill these gaps [54]. Friedman *et al.* [54] define values as "what a person or group of people consider important in life." Usually, a VSD design process consists of three types of investigations: conceptual, empirical, and technical. For this paper, only the conceptual investigation is of relevance. Such an investigation aims at understanding the interests and conflicts across the stakeholders' debates within their cultural and strategic contexts [54]. Moreover, it proposes approaches to mitigate conflicts and prioritize values in trade-off situations. Within the scope of this study, VSD is thus used to understand the stakeholders' interests and values towards the control of autonomous weapon systems. VSD in the context of autonomous weapons systems has also been used by Asaro [55], taking important work on autonomy by Cummings [56], [57] into account, referring to her concept of automation bias and the VSD-study of the cruise missile Tactical Tomahawk [58]. Thornton *et al.* [59], de Sio and van den Hoven [20], as well as Umbrello [19] use the VSD method on autonomous vehicles and AI, arguing for human-centered approaches like MHC in civilian innovation as well. Our analogous approach tries to grasp challenges of today's discussion of LAWS.

This work is a contribution to the field of IT peace research [60], [61] as well as natural science/technical peace research [62], [63][2] and sheds light on technology's normative and social effects in crisis and conflict. While Verbruggen *et al.* have dedicated several papers to autonomous weapon systems, following a more reflective and open research path [64], this work contributes by incorporating VSD. So far, only one contribution focusing on LAWS and VSD exists: Verdiesen asks for moral values which are important to military personnel and the public, disregarding the merit for a more critical, deconstructivist stance [65]–[67]. Finally, this work aims at offering interdisciplinary social scientific approaches, like Pugliese's [68], and technical perspectives, like Arkin's [42], towards challenges for international security.

[2]As Altmann states, "(n)atural-science/technical research for peace, international security, arms control and disarmament is applied research with the intention to support the political processes of preventing war, reducing armament, building confidence and diverting financial and human resources from military to civilian purposes, in particular for solving urgent global humanitarian problems". It "is a broad field of research that deals with the role of natural scientific and technical possibilities in the context of war and peace, armament and disarmament" [62].

## Theoretical Background

We chose to analyze MHC with regard to LAWS from a perspective of the Value-Sensitive Design (VSD) approach. Shedding light on LAWS' inscribed attributes, we follow the VSD approach by Friedman *et al.* (69). VSD yields theoretical and methodological implications by assuming more or less abstract values to be reflected in interfaces or software and thus indicate the need of interpretative work (70). Thereby, we understand values in affirmation of the definition of Friedman *et al.* (54) as norms or standards assessed by a collective while neglecting values as norms by individuals. These values, also supporting a common understanding by VSD scholars, do not necessarily have to be explicitly moral values (54). Following this approach, we consider the design process to be especially relevant with respect to the interaction between human operators and LAWS, an assumption that is already prevalent in debates about the regulation of autonomous weaponry (20), (46), (57).

In this work, we focus on the first and fundamental step of the three-pronged iterative approach, i.e., conceptual investigations of LAWS (58). Here, we search for values incorporated into LAWS, which may be competing against each other due to different stakeholder positions across the CCW arena (69). Thus, we pose the question: "What values can be derived for MHC from the stakeholders' discourse in the 2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems?"

Hence, we are interested in approaching MHC in relation to its empirical context, constituted by various perspectives in the CCW debate, which reflect different values, and broader discursive narratives. Understanding the expert's group debate as well as the diverse notions of MHC also allows the formulation of implications or priorities in technology development as well as for the regulation and control of increasingly autonomous weapon systems. Answering our research question, we investigate the CCW GGE's divergent conceptual understandings of MHC of LAWS (first gap). The divergent understandings of human-LAWS interaction can be deconstructed by choosing the theoretical perspective of VSD (second gap) as well as a thorough the investigation of CCW-relevant socio-technological values and their interrelatedness (third gap). Our analysis of influential discourses and values leads to the formulation of implications for the design of LAWS (fourth gap).

## Research Design

As we are interested in the question of how to retain MHC of LAWS, we focused on the Group of Governmental Experts on LAWS, which meet in the forum of the Convention on Certain Conventional Weapons. These experts constitute the most relevant international body dedicated to understanding respective human-machine interaction in a military context, and it is the main organizational forum for the conceptual debate regarding autonomy and control with respect to lethal weaponry. To answer the question of how MHC may be achieved within this forum, we choose a discourse analytical approach, grasping mindsets and conceptualizations of the stakeholders. Our sample and method of analysis are presented in the following.

> **To still be able to hold humans responsible, a special focus on human machine "touchpoints" is needed.**

### *Data Collection*

We concentrated on documents of the *2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)* (GGE on LAWS), thus restricting our sample to working papers and statements by states as well as non-governmental actors, which were formulated in the course of the first meeting, taking place from April 9 to 13, 2018, and the second meeting, held from August 27 to 31, 2018 (71). In sum, we analyzed 43 documents (see Table 1). The number of member country statements needed to be reduced, not at least due to repetitiveness of content and to reflect actors' relative dominance in the discussion. Thus, we included strong positions of U.S. and European countries such as the U.K. As Western countries' statements were more frequently represented and accessible, our selection of country statements tried to reflect a certain unequal distribution of participation. China's and Russia's positions were reflected in the working papers they submitted. Special attention was given to the U.S. due to their prevalence in the development of LAWS and network-centric warfare. To reflect transnational work done in the arena of the GGE on LAWS, we included statements by debate-steering non-governmental actors, again mirroring the diversity of positions as well as dominance in the discourse. We also regarded positions of military-relevant industry partners and marginalized critique by whistleblowers. This was useful to gain more insight into discursive narratives and actors' positions, serving as points of reference for participants of the CCW expert group. While the number of documents by non-state actors may seem high in comparison to GGE documents, the latter were considerably longer and offered more in-depth content. All sources are accessible online.

### *Data Analysis*

For the analysis of data, we used the open-source software of R, in particular RQDA (72), allowing for a qualitative analysis of text files. Following our theoretical assumptions of discursivity, we conducted a discourse analysis, inspired by Fairclough's Critical Discourse Analysis of communicative events (73). Textual documents, like the working papers and statements by states

**Table 1. Overview of Sample**

| Title (abbreviated) | Parties | Number and Type of Document |
|---|---|---|
| CCW/GGE.2/2018/WP.1<br>CCW/GGE.2/2018/WP.2<br>CCW/GGE.2/2018/WP.3<br>CCW/GGE.2/2018/WP.4<br>CCW/GGE.2/2018/WP.5<br>CCW/GGE.2/2018/WP.6<br>CCW/GGE.2/2018/WP.7 | United Kingdom<br>Estonia and Finland<br>France<br>United States of America<br>Brazil<br>Australia<br>Austria, Brazil, Chile | 7 CCW working papers of the first (April'18) meeting |
| CCW/GGE.1/2018/WP.1<br>CCW/GGE.1/2018/WP.2<br>CCW/GGE.1/2018/WP.3<br>CCW/GGE.1/2018/WP.4<br>CCW/GGE.1/2018/WP.5<br>CCW/GGE.1/2018/WP.6<br>CCW/GGE.1/2018/WP.7 | Venezuela, Non-Aligned Movement<br>Argentina<br>Poland<br>United States of America<br>Intern. Committee of the Red Cross<br>Russia<br>China | 7 CCW working papers of the second (August'18) meeting |
| LAWS6b<br>LAWSGeneralExchange | Brazil, Estonia, Pakistan, UK, Israel<br>Australia, EU, Canada, Germany-France, Republic of Korea | 10 country statements (5 on agenda item 6b, first meeting, 5 of the general debate) |
| Guidelines for the human control of weapon systems<br>International law and the standard of human control in weapon systems<br>Statement to the Convention<br>AI in Weapons: The Moral Imperative for Minimally-Just Autonomy | N. Sharkey (International Committee for Robot Arms Control, ICRAC)<br>T. Chengeta (South African Research Chair in Intern. Law)<br>Campaign to Stop Killer Robots<br>J. Scholz & J. Gaillott (Trusted Autonomous Systems Defence Cooperative Research Centre & University of South Wales, Canberra) | 4 CCW documents by non-governmental parties |
| Report of the 2018 session | UN CCW GGE | 1 summary report of both meetings |
| List of Participants, LoP Addendum | UN CCW GGE | 2 informative notes |
| DoD Directive Number 3000.09<br>AI Next Campaign Explainable Artificial Intelligence<br>Safer, smarter, faster: An interview with Gen. James McConville | US Department of Defense<br><br>US Army | 4 documents of specific US position (1 DoDD-Directive, 2 DARPA websites, 1 army interview) |
| Google says it won't use AI for weapons<br>The Wired Interview: iRobot CEO Colin Angle<br>Lockheed Martin and Drone Racing League Launch Groundbreaking AI Innovation Challenge<br>The new recruitment of AI experts to significantly strengthen Samsung's AI R&D capabilities<br>AI is the next step for robots – A Conversation with Nicolas Boudot<br>US Drone Program Whistleblower Explains Why She Spoke Out<br>Brandon Bryant: "Ramstein is absolutely essential"<br>Don't fear my robots, says the Boston Dynamics founder who makes machines that drive the internet wild | CNN<br>The Wired<br>Lockheed Martin<br><br>Samsung<br><br>Dataconomy<br><br>Global Research<br><br>Das Erste/NDR<br>Business Insider | 8 newspaper articles (interviews) and press releases |

on autonomous weapons systems, may contribute "to the construction of systems of knowledge and meaning" (73). For the qualitative analysis of text files, codes were developed abductively. The codes are derived from the documents and also take into consideration the existing literature regarding network-centric warfare. Network-centric warfare conceptualizes an technology-oriented strategy of warfare being particularly time- and space-oriented, forming a decentralized network of synchronizing entities and man-machine teaming (34), (35). Codes were also grouped by the three code categories humans and LAWS interaction, socio-technological values and embedding discourses, the latter comprising military discourse as an important sub-category of LAWS-embedding discourses. Hierarchical relationships were noted in the respective memos and are visualized in this work (see Figure 1). Because the derivation of inter-code relationships proved costly, we chose to rely

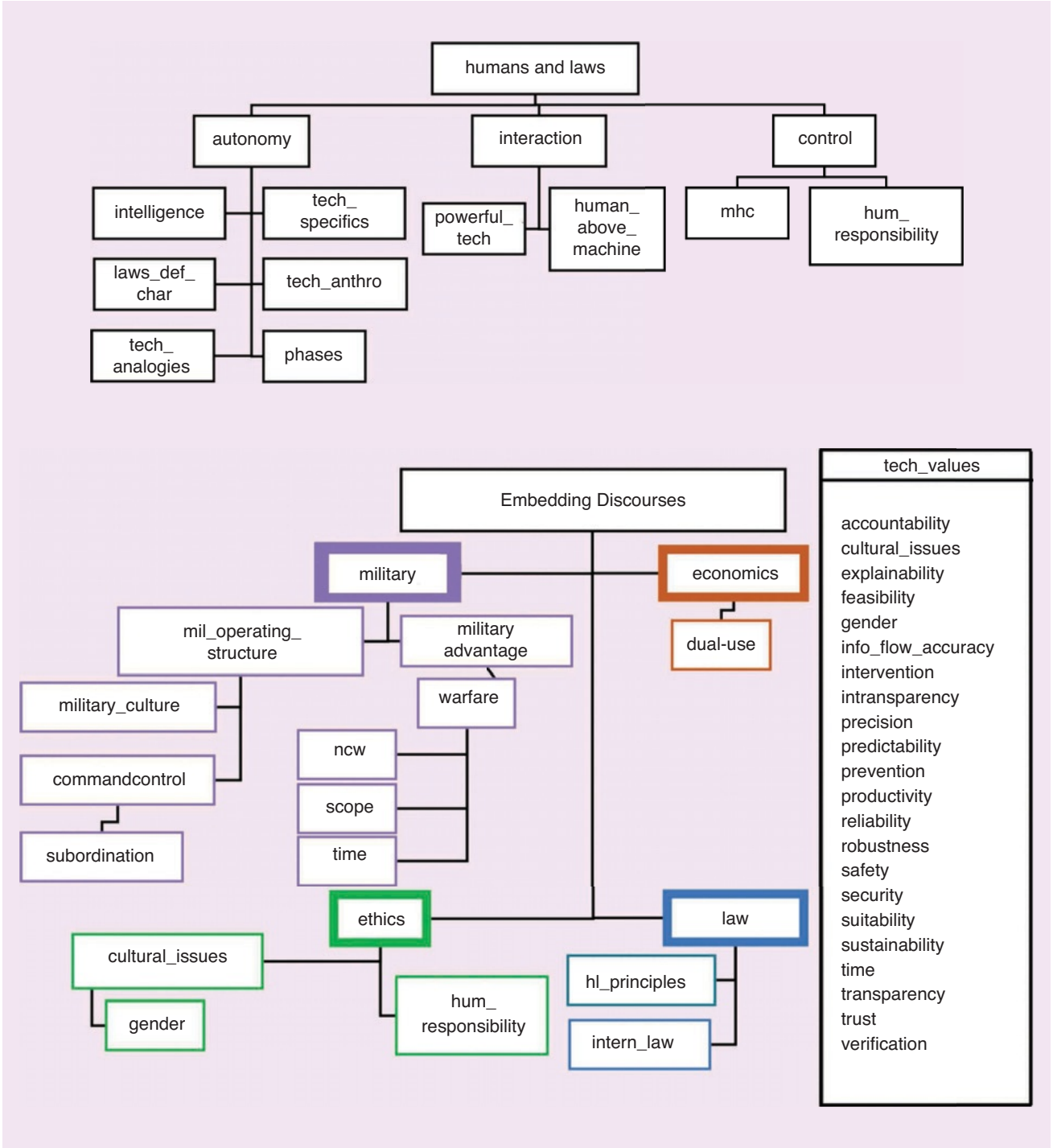


**FIGURE 1.** Code categories "humans&LAWS," "Embedding discourses" and "tech_values."

on the R text mining package tm [74] as well as related packages to grasp relationships of proximity of word stems. To look for associations of word stems, the text corpus was split into 230 files. Punctuation and numbers were removed, and a rather simple algorithm reduced words by cutting of suffixes.

## Results

### *Autonomy and LAWS*

To gain a deeper insight of the dominant conceptual and empirical understanding of autonomy and LAWS, we created "autonomy" as a code for covering all instances in the CCW discourse in which the respective signifier was named and its meaning temporarily fixed. We also coded phrases describing defining characteristics of LAWS with "laws_def_char." To approach not only what is conceptually envisaged with respect to autonomy and LAWS but also how these concepts may materialize, we created the codes "tech_specifics" and "tech_analogies." These four codes all fall under the code category humans and LAWS interaction.

**The identification or detection of targets is another critical function which is to be distinguished from the ultimate selection of a target.**

We retrieved 168 codings for "autonomy," and while these were often made with references to humans not being "in the loop" anymore or the general problem of defining and differentiating between autonomy and its various degrees, one can notice a recurrent focus on autonomy as a function. The following references were found: "using a weapon with autonomous functions" (WP4-US), "(s)ystems with advanced artificial intelligence and enhanced autonomous functions" (GenEx-AU), that "its understanding changes with shifts in the technology frontier, and different functions of a weapons system could have different degrees of autonomy" (SummaryReport), that "(a)utonomy is in and of itself a function" (WP5-BR), that "some semi-autonomous machines can have highly autonomous critical functions while highly autonomous machines can have no or limited autonomy in critical functions" (SummaryReport).

Thus, within the CCW debate, it becomes a clear objective to focus on autonomy of such critical functions. As critical functions, "different parts of the targeting cycle" (SummaryReport) are to be taken into consideration. Especially, functions of selecting and attacking targets are looked at critically. Out of 168 codings regarding "autonomy," 56 were specifically referring to selecting and targeting as critical functions. This is illustrated in documents by various actors, pointing out that "(t)he ICRC has correctly noted that the main and perhaps what states should be concerned about characteristic of AWS is that they have autonomy in the critical functions of selecting targets and attacking without human intervention" (SARCIL). The United Kingdom also stressed that "many participants (had) call(ed) for human control of 'critical functions,' often specifically referring to 'select and engage,' but it was unclear precisely what these terms mean" (WP1-UK).

Looking at empirical references to technological specifics, it is referred to "(p)attern matching algorithms (…) used for target selection (WP2-ES/F) or weapons with *advanced sensors such as millimetric wave radar (that) can operate beyond visual range* (for example, the AIM 120 Advanced Medium-Range Air-to-Air Missile (AMRAAM)) or engage multiple targets from a single platform (such as Hellfire or Brimstone guided missiles)" (WP1-UK; own emp.).

In general, sensors are often named as crucial features of LAWS: "Computers can enable machines to respond to inputs from sensors through an application of the algorithms or other processes with which they have been programmed" (WP4-US); "(e)ach munition is equipped with heat and radar sensors which can scan a 200m diameter area. If a target is detected, the warhead is activated; otherwise it self-destructs" (SummaryReport). These statements, as well as others such as – "(t)he projectile has sensors that allow it to identify the target that the human operator intends to hit, and computers and guidance systems that allow it to select and engage that target" (WP4-US) – yield two implications. On the one hand, the CCW discourse treats the detection or identification of a target as part of selecting it, constituting them as one process of selection. Within the CCW discussion, the International Committee of the Red Cross noted that a "weapon system with autonomy in its critical functions" is one "that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention" (WP5-BR, ICRC). Thus, detection or identification is seen as a synonym or subcategory, respectively, for selecting. This already implies the significance of an autonomous identification of a target with respect to its actual, fixating selection.

On the other hand, it becomes clear that the identification or detection of targets is another critical function which is to be distinguished from the ultimate selection

of a target. The interpretation of the statements indicates that detection is a critical function belonging to the targeting cycle, preceding the actual selection and subsequent attack of a target. Numerous statements point to the detection and identification of a target, while most times they do not offer explanatory details: "Both primary sensors (laser scanners, millimeter-wave radars, hyperspectral imaging, etc.) and signal processing Algorithms" (WP2-ES/FI); "the Lightweight Counter Mortar Radar can identify indirect fire threats by automatically detecting and tracking shells and backtracking to the position of the weapon that fired the shell" (WP4.2-US).

Besides these functions belonging to the targeting cycle, CCW participants also pointed to other tasks that may be categorized as autonomous. These are supportive tasks like "cyberattack warning, supply chain logistics" (DARPA_2) or automatization of (US) Department of Defense "business processes, such as security clearance vetting or accrediting software systems for operational deployment" (DARPA_2), for accident prevention, i.e., ground collision (WP4-US), or logistical calculations (WP1-UK).

In general, references that were made with respect to autonomy of certain critical functions stressed repetitively that a weapon system might consist of autonomous as well as non-autonomous parts and functions, e.g., "weapon systems that have been deployed still require human remote authorization to launch an attack (even though they may identify targets autonomously)" (ICRC/WP5-BR). At the same time, CCW participants did not omit referring to autonomy's different degrees. Often, autonomy of a system meant automatization of processes, but references to artificial intelligence and machine learning, indicating self-learning capabilities and independence, were also prevalent (WP2-ES/FI). No references regarding specific types of deep learning algorithms or the like were made.

### *Human-Computer Interaction*

To understand the dominant and respective marginal counter-discourses on human-LAWS interaction across the CCW debate, we created and organized the codes in a continuum with technological anthropomorphization[3] and MHC at its extremes (see Figure 2). The first stage defines human-LAWS interaction to be dependent on technology's "behavior," treating technology as an essential being with human-like features (e.g., intelligence or decision-making authority). In contrast, MHC, at the other end of the continuum, implicitly assumes a hierarchical relationship between humans and technology, with humans having legitimate authority over technology as an instrument. We only marked phrases specifically referring to MHC with the respective code, while all other statements regarding control (without further conceptualization as meaningful) where marked accordingly (i.e., "control").

**Adherence to principles of international humanitarian law are a dominant reference point for evaluation.**

The code "powerful_tech" was used to grasp statements that indicated perceiving technology (and technological development) as a driving force while not anthropomorphizing it. In cases where perceptions of "both sides" (technology versus human) were used, they weighted each other out to "interaction," reflecting a rather equal relationship. This code was naturally also used when actors explicitly referred to interaction as a relationship. The code "human_above_machine" refers to a hierarchical relationship of humans, yet not necessarily being always in control. It is important to note that this process heavily focused on linguistic formulations,

[3]It would have been plausible to put "autonomy" as the contrary extreme to "MHC." Yet, this code was used massively across documents, including reflections on definitory issues and demands for constraining LAWS. Thus, we created a code to reflect anthropomorphized illustrations of technology.

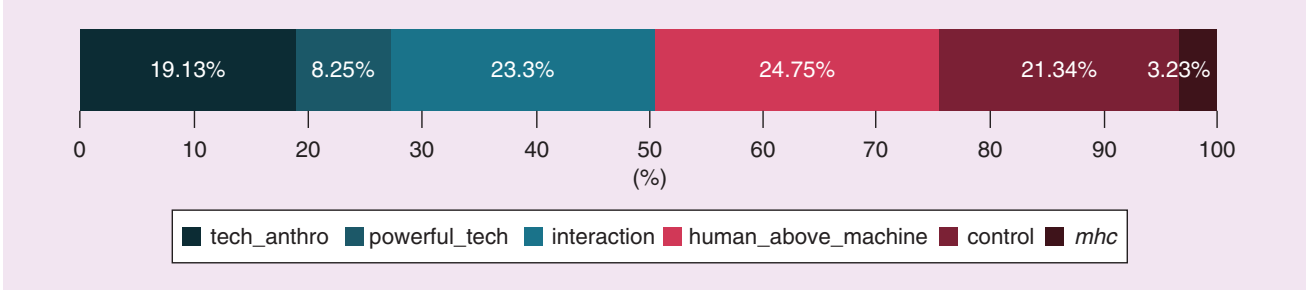


**FIGURE 2.** Continuum of human-LAWS relationship (100% = 1,176 codings).

assuming that language shapes actors' perceptions of human-LAWS interaction (3).

Speaking in relative terms, statements indicating hierarchical relationships of humans and subordinated technology were found almost twice as often as statements supporting the autonomy of technology (49% versus 27%). At the same time, "interaction," as a rather neutral or non-hierarchical position, made up roughly a quarter (27%) of the respective human-LAWS interaction statements. This indicates that perceiving humans to be superior to technology is the dominant view within the CCW GGE forum. These relative frequencies yield several implications: First, it becomes clear that explicit MHC-related statements are very rare (in absolute as well as relative terms). This reflects that despite its initial trendiness as an ethical-legal buzzword (6), (75), CCW GGE participants are rather uninterested in highlighting MHC. In cases of referring to MHC, it is either done by a non-governmental participant like the Campaign to Stop Killer Robots (GGE_CTSKR) or the ICRC, elaborating on the concept in a more detailed manner (WP5.2-ICRC). When used by states, MHC is framed to be important merely by (potential target) states which are dedicated to a ban of LAWS, like Pakistan or Brazil. In such cases, stronger normative terms were used: "The task for the GGE now should be to ascertain the scope and extent of human control necessary to address the various concerns associated with LAWS to ensure that it is meaningful" (6b-PAK); "(t)he proposal that humans retain 'meaningful control' over LAWS seems to us the most promising avenue to explore" (6b-BR). Other nation-states' representatives do not seem similarly interested in the MHC concept, disregarding the concept considering its level of abstraction and necessity: "(I)t becomes difficult to provide a technical statement of meaningful human control. (…) To be meaningful, human control does not necessarily have to be exercised contemporaneously with the delivery of force. (…)" (WP2-ES/FI).

**Control of autonomous systems matters, however, with divergent implications on how, what (regarding the function and situation), and in which circumstances.**

The U.S. also relativizes the conceptual impact of MHC, as "an operator might be able to exercise meaningful control over every aspect of a weapon system, but if the operator is only reflexively pressing a button to approve strikes recommended by the weapon system (…)" (WP4-US). Pointing out its lack of value because of its conceptual unclarity functions as a devaluation mechanism of the MHC concept across the CCW discourse: "'Control' and 'judgment' are, however, flexible terms, even when qualified by adjectives such as 'meaningful' (…), (…) used (…) to signify different things" (6b-Estonia). Most decisively stated, MHC "or a similar notion" is not considered "to reflect a new or emerging norm of international law" (6b-Estonia), while one participant seems to regret "exploring the scope of meaningful human control in the delegation of decisions to intelligent machines *instead* of what we must regulate on LAWS" (GenEx-KO, own emp.).

Nevertheless, the widespread references to human control stress the respective participants' underlying interest in a hierarchical relationship, with human operators controlling LAWS to varying degrees. As laid out in the following section, "Socio-Technological Values of LAWS," such control may be associated with various values, with the latter being regarded as defining characteristics of the concept of MHC. Thus, even though actors show little support of the concept, they widely refer to technological requirements defining the concept of interest (75).

Third, while dominantly retaining human-focused arguments within the CCW debate, there are still a considerable number of statements carrying a supportive notion of technology, i.e., perceptions of technology as a societal driving force or as essentially being. Phrases that were coded as "tech_anthro" or "powerful_tech," respectively, show such instances: "The development of artificial intelligence (AI) should be seen as a logical process in computing science," while demanding that "the discussions on LAWS must reflect the undeniable direction of technological development" (WP2-ES/FI). Interest in the support of LAWS and related AI technology is further formulated, declaring "the system would be capable of defining and thereby deciding the ultimate goals of its functioning, very much like humans do" (WP2-ES/FI), peaking in the following explicit statement for implementation: "The level of abstraction of computing keeps getting higher and higher, leading towards increasing possibilities for various levels of machine autonomy. Past experience has shown that once new technology proves to work, society quickly adopts it, and later its use becomes the accepted norm" (WP2-ES/FI).

While remarks that were coded by "interaction" may also imply *characterizations* of technology, they also suggest a more reflective awareness of interactional relationships between humans and LAWS as well as related effects: "(T)he way humans use machines and interact with them is changing (…) (because) (i)n

complex systems the human role will have various postures in relation to the machine" (WP2-ES/FI).

### *Socio-Technological Values of LAWS*

The NGO Article 36 has issued a briefing within the CCW forum (75), in which it points to the question of value-sensitive design while identifying certain "key elements" for MHC: 1) predictability, reliability, and transparency of technology; 2), information accuracy regarding planned outcome, operation, function of technology, and context of use; 3) "timely human action and a potential for timely intervention," and 4) "accountability to a certain standard" (75). We assume such elements of MHC to be implementable into technology. Our analysis reveals more important values and looks for mutually tense and supportive relationships, respectively, to put them into context. Thus, we identified 23 values and looked for frequencies and locations of occurrences (see Figure 3).

The codes "accountability," "cultural_issues," "gender," "scope," "time," and "verification" were not always used for the explicit description of technology, but also with respect to general procedural issues. Yet, such phrases implied that participants found the respective characteristic very important. Therefore, they are included as relevant socio-technological values that are reflected in the conceptualization of LAWS and interaction. The other value-grasping codes were linked directly to technological requirements.

Most crucial are the codes "time," "predictability," and "reliability" (77, 74, and 70 codings). Additionally, references coded as statements regarding "productivity" (n = 64), "accountability" (n = 58), "explainability" (n = 53), or "safety" (n = 50) could be made out at a relatively high frequency, while CCW documents surprisingly referred to issues of (human) "intervention" (n = 21) or an accurate flow of information (n = 14) only at a relatively moderate rate. Yet, not every technological requirement considered necessarily has a supportive impact with respect to the establishment of MHC. Thus, we screened arguments pointing out MHC or human control.

For example, a statement formulated by Brazil stresses that "meaningful human control can only be achieved if the role of the human (…) is such as to ensure (…) *the capacity to intervene and override machine functions when operationally possible*" (6b-BR, own emp.). Furthermore, the non-governmental actor ICRAC defines "*sufficient time* for deliberation on the nature of targets (…) (as one of the) necessary conditions for meaningful human control of weapons" (ICRAC-WP3, own emp.).

Besides having the possibility to intervene — also with respect to the issue of time — France, perhaps taking a less critical stance on LAWS compared to ICRAC, points out with regards to "(d)eveloping autonomy and human-machine interaction" that (t)he human command must be aware of and be able to assess system

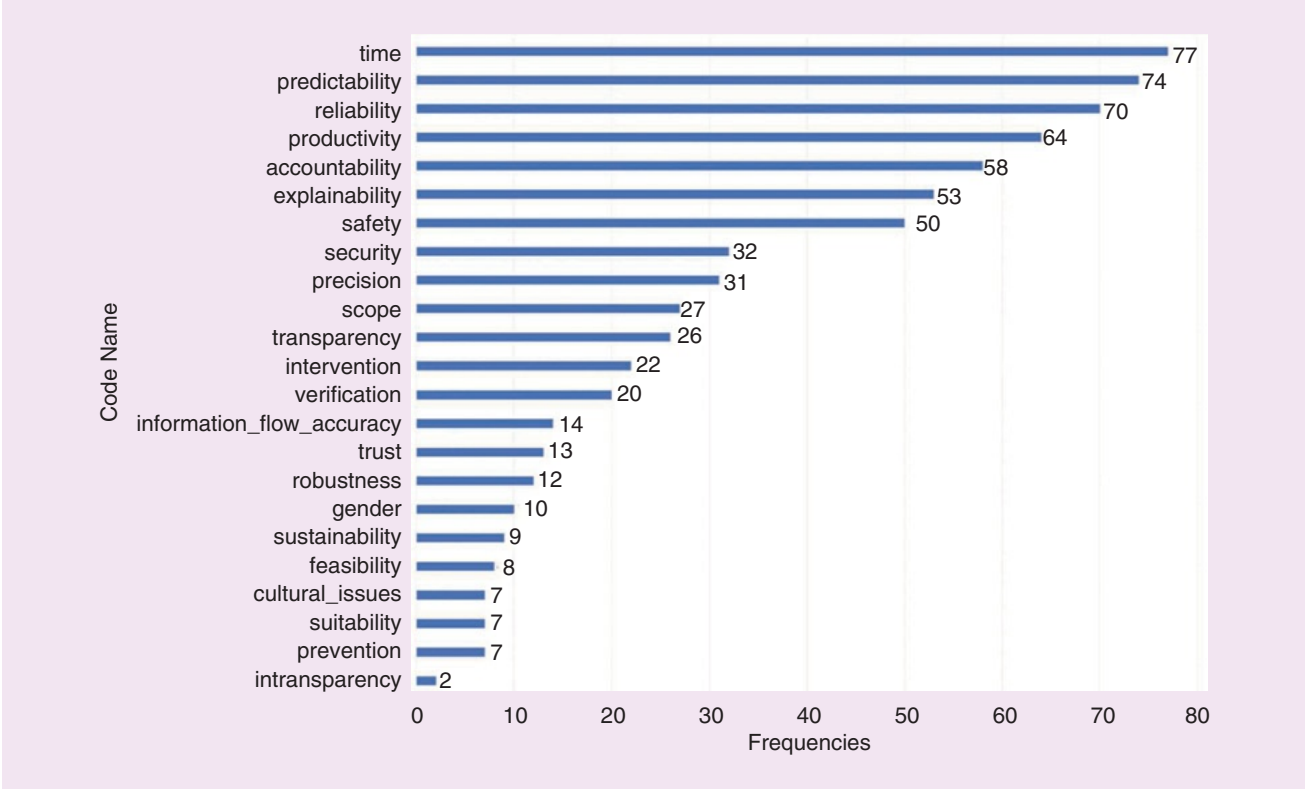


**FIGURE 3.** Occurrences of socio-technological values.

*reliability* and *predictability*" (WP3-FR, own emp.). Assuming reliability and predictability as well as issues of "time" and "intervention" to be foundational for MHC, we further analyzed the CCW discourse with respect to associations between these values. As RQDA did not offer a convenient approach for testing inter-code relationships (76), we alternatively proceeded with text mining and a search for word associations. A word cloud (see Figure 4), visualizing all word stems which occurred at least 70 times, reflects similarity of codes and relevant words.

Consequently, we screened the CCW corpus for associations occurring at least by 30% with respect to predictability.[4] This revealed first that "predict" is sometimes associated with aspects relating to the question of MHC. For example, "agency" (56% of all occurrences of "predict"), "dignity" (46%), "human control" (43%), "moral" (34%), or "ethical" (30%) are associated with predictability. Second, testing for associations also showed that predictability is often (78%) associated with reliability. Similarly, reliability is often associated across CCW documents with control ("software control," "human control") or agency (42%), "trustworthy" (70%) as well as "knowledge" (43%) or "consequence" (38%). At the same time, "intervention" may be associated by at least 30% with "overriding" (46%), "capacity" (38%), "conscious" (37%) and "constrain" (38%).

As the preceding coding process revealed, "time" was not only coded referring to statements directed at values, but also with respect to elaborations on warfare (as well as less substantially laden phrases like "in the meantime," "at the same time"). Thus, there are many


**FIGURE 4.** Cloud of word stems occurring at least 70 times.

[4]As we conducted a word stem analysis, we eliminated numbers, punctuation, and capitalization. At the same time, words were reduced to their stems. Thus, any associations presented here, are the respective co-occurrences of stems (e.g., "predict" instead of predictability or predictable and "effici" instead of efficiency or efficient).

words associated with "time," among which some like "faster" (36%), "able" (31%), or "deliberately" (30%) indicate the issue of time pressure and necessity of time to intervene or decide, respectively. Yet, besides pointing to enough time as a requirement for human control of LAWS, associations of "time" also already reflected the signifier's relevance of the current military discourse (see section, "Influencing Discourses.").

Another group of values named across CCW documents did not seem to support the actual implementation of MHC. Among these was precision, which is associated by 53% with efficiency, and military terms like "troop" (52%), "military" (33%) and "Afghanistan" (48%), "Syria" (33%), "Iraq" (30%), "Libya" (48%). At the same time, precision was comparatively often used along empirical examples of LAWS like "torpedo" (38%) and negatively-laden words like "cruel" (48%) or "deteriorating" (33%). Similarly, efficiency was sometimes associated with the same words as well as "easier" (47%), "cost" (34%), or "kill" (34%). Screening for associations also suggests efficiency-related phrases not be formulated alongside issues of control, ethics, and international humanitarian law.

The third group of codes within the category of values seemed to be neutral towards MHC, as the coding process suggested. Values like "feasibility," "suitability," and "sustainability" were present across documents, yet the respective substantial understanding of these features did not become clear straightforwardly. Looking at other word occurrences along feasibility raised the possibility of it constituting an empty signifier (77), rather reproducing cost-effective ratio instead of proactively supporting MHC (cf., 34% "rule," 31% "advantageous," 32% "commercial," 40% "applicable," 31% "certified").

### *Influencing Discourses*

As our work carries discourse theoretical assumptions, i.e., of discourses reproducing mindsets and practices, we checked for codes referring to overarching societal topics or thematic areas. Figure 5 reveals dominant trains of thought functioning as points of reference throughout argumentative structures.

"Ethics" as a structuring discourse of MHC was coded almost 400 times across CCW documents. The frequency shows how ethical vocabulary serves as a dominant frame for discussion. The relevance of ethics as a guiding discourse is also illustrated considering respective sub-codes "hum_responsibility" (n=121) or "HL_principles" (n=205), with the latter clarifying the close link between ethics and law ("intern_law" n=349). While Pakistan states that "its (LAWS') use should be considered unethical and unlawful" 6b-PAK), paralleling law and ethics, other remarks stress the normativity of law due to its causal links to ethics by respectively as well as actively

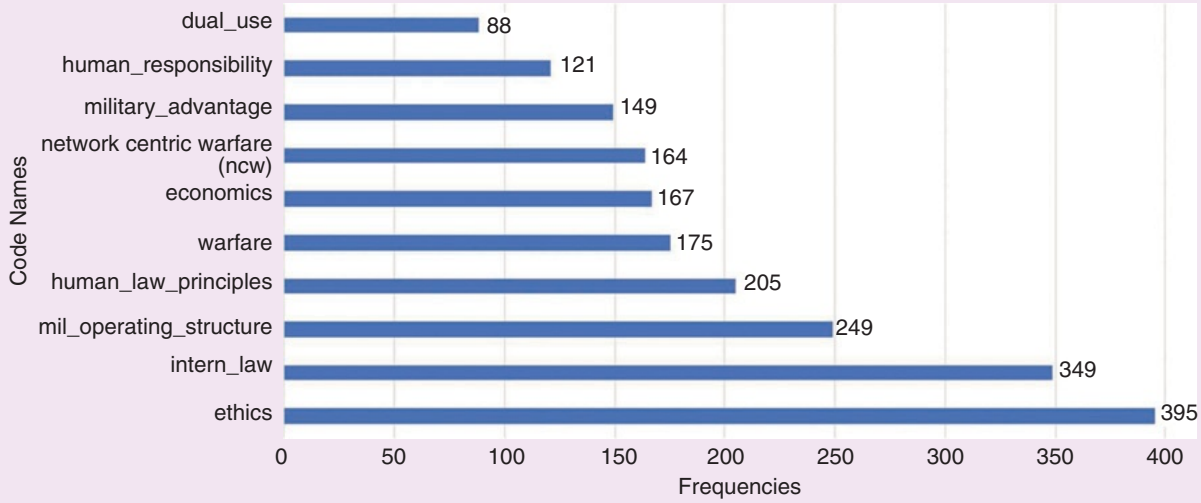


**FIGURE 5.** Reflections of relevant discourses.

demanding vocabulary: "All weapons systems *must comply* with International Law in general and, in particular, with *International Humanitarian Law* and with International *Human Rights* Law" (WP5-BR, own emp.) and "(w)e *believe* that any legal discussion about LAWS *should* centre on compliance with International *Humanitarian* Law, including the *obligation for all States to ensure* the lawfulness of their weapons, means and methods of warfare. *These should be our constant reference points*" (GenEx-CA, own emp.).

Besides ethics and law, we identified "economics" ($n$ = 167) and the military discourse consisting of military signifiers ("mil_operating_structure" $n$ = 249, "warfare" $n$ = 175) and indicating military logic ("military_advantage" $n$ = 149) to function as structuring forces. While "dual_use" ($n$ = 88) constitutes a sub-code of the economic discourse due to its recurring references to R&D in the civilian sector, the code "NCW" ($n$ = 164) is grasping all phrases related to the so-called network-centric warfare. However, the code "NCW" is formed the military discourse, it is also influenced by economics [34], [35].

Thus, we categorized thematic reference points by four subjects, forming two interrelated pairs: law and ethics as well as military discourse and economics. These pairs are not only illustrated in the text documents (see Figure 6), but also inspired by NCW-literature [34]. When looking at the word stem associations of "law," we did not find any immediate reference to ethics. Yet, there are associations with "humanitarian" (50%), "principle" (45%), "mean" (31%), and "remind" (34%), indicating the cognitive work of ethics. When testing for word associations of "ethics," it is revealed that "ethic" is very often used along "predictability" (50%) and "unexpect" (32%), thereby indicating that ethical arguments are associated with the value of predictability or vice versa. Furthermore, "ethical" is associated across the CCW corpus with "disobey" (50%), "nonacceptable" (50%), and "question" (32%), all reflecting on the option of intervention regarding decision-making (37% "decision"). At the same time, "ethics" is also associated with the value of sufficient time (32% "acute").

Associations of "military" did not point to any of the dominantly stated values. Yet, besides references to weaponry ("laser-guided," "microcomputer," "bomb," "radar") or military vocabulary ("battlefield," "submarine"), it is mostly associated with "civilian" (44%), "harm" (35%), "accuracy" (34%), and "collateral" (33%). These word stems point to the humanitarian principle of distinction as well as the technological value of "precision" (or "accuracy"). Both associations of "command" and "advantage" neither refer to values like "predictability," "reliability," "time" or "intervention." Rather, they are also associated with military vocabulary as well as with relatively diffuse values like "appropriate" (35%, ~"suitability"), "initially defined" (41%), "feasibility" (31% with "advantage"),

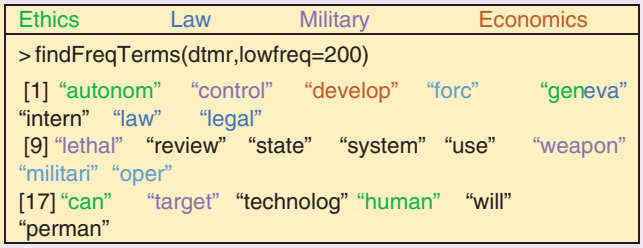


**FIGURE 6.** Extract of terms occurring at least 200 times & major discourses.

and again with precision (i.e., "accuracy" 51%). Associations of "efficiency" neither revealed any reflections of values related to MHC. Instead, paragraphs deal with "cost" (34%), "war" (58%), "military" (67%), but also with "precision" (53%) and "quick" (30%), reflecting current *NCW*-ratio.

## Conflicts of Interest in Implementation of MHC

Our analysis revealed a variety of new findings, some of which reside on a higher level of abstraction while others are more closely tied to existing practices and materiality. Independent from their theoretical-practical position, our results yield various implications with regards to the question of conflicts of interests regarding the implementation of MHC:

a) Time, predictability, and reliability, associated with each other to varying degrees, are credited to be the most important values incorporated into emerging technologies.
b) Accountability, explainability, and potential for intervention are recognized to support MHC.
c) In contrast, efficiency and precision are suggested to have an ambiguous relationship with affirmative values.

The following section discusses the results with regard to the ongoing debate on the regulation of LAWS. Additionally, it is useful to consider not only this paper's results but also existing limitations. This allows for an outlook as well.

### *Implications*

Analysis of recurrent understandings of both autonomy and LAWS across CCW GGE discussions revealed disagreement among the parties as well as a focus on the targeting cycle and its "critical functions" (see "Results: Autonomy and LAWS" section). To retain MHC, it seems helpful to reach a consensus on working definitions by concentrating on critical functions of engaging and the different steps of selecting and targeting. Second, looking at autonomy and LAWS understandings, parties may need to find some common ground with respect to empirical examples of LAWS while at the same time being aware of risks of transparency.

Approaching underlying perceptions of human-machine interactions again revealed the necessity to be conscious about *automation bias* and blinded views on technology (see "Results: Human-Computer Interaction" section). Even though the establishment of MHC relies on human-centric arguments, our analysis offers the added value of reflective positions on interacting networks. To still be able to hold humans responsible, a special focus on human machine "touchpoints" is needed. The research by Brown *et al.* [78] allows a first glance at the CCW GGE debate in this respect. In contrast to Verdiesen's work [67], our chosen VSD-perspective poses a first approach to the CCW GGE parties' understanding of MHC-relevant technological features.

Even though the stakeholders might not agree on Meaningful Human Control as a concept, the issue of control is at the heart of the debate [20], [21]. Regarding human control, a sufficient amount of time to intervene was found to be one of the most important technological requirements (see "Results: Socio-Technological Values of LAWS" section). To incorporate this feature of human-controlled LAWS, such weapon systems need to be designed accordingly. It is therefore critical to avoid automatization of firing when identification is seen as successfully completed. Options of multi-channel communication between operator and system are already under research [41]. Yet, with respect to assuring MHC, it is a) important to ensure the option of intervention, while b) active confirmation or denial of an attack might reduce the risk of technology-biased behavior. Thus, designing an interface offering (obligatory) yes/no-options with regards to an attack may ensure human authority. At the same time, values like reliability, predictability, information accuracy, or explainability might be carried by a usable interface, offering transparent listing of target detection processes. Situational awareness by commanders, consisting of some of the named requirements, is crucial for the meaningful human control of autonomous weapon systems [79], [80]. MHC can only be enforced with sufficient training of the human operators [24].

The expert group's debate mainly focused on the fifth phase of the targeting process, in which the mission is planned and executed, entailing the F2T2EA cycle of "find, fix, track, target, engage, and assess" [28]. As "[i]t is during this phase that the selected lethal or nonlethal means will be used[,] [...] this is the phase [the debate is] focused on the most" [28]. With respect to the detection of a target, training of deep learning algorithms should ensure both reliability and predictability by systematically recording and evaluating single steps. Regarding machine learning, it may also be necessary to discuss implications of diverse degrees of depth, i.e., representation and abstraction of layers. A focus on testing indicates, as do our findings referring to the importance of verification and validation of soft- and hardware, that procedures apart from the planning and execution phase, on which the debate has mainly concentrated, are critical to ensure predictable and reliable execution of the targeting process.

As our work on the debate's surrounding and influencing discourses shows, international humanitarian law principles are still a dominant reference point for evaluation (see "Results: Influences Discourses" section). With respect to the principles of distinction and proportionality,

several procedural measures may prove helpful. Software developers, certainly aware that surveillance, detection, or identification are crucial parts of the targeting process, need to intensify sensitivity towards biases (39).

Still ongoing is the discussion whether it is possible to design autonomous weapon systems that can adhere to the principles of international humanitarian law, e.g., to protect civilians and to ensure the use of appropriate and necessary force only (21), (81). Cummings (36) has argued with regards to the strategic target selection by commanders that in case the precision of the system performs better than the human operators, autonomous target selection follows the IHL (36). She (36) suggests that autonomous systems are potentially better in performing and preventing unnecessary suffering due to the stress-bias commanders face in a combat situation. Thus, a research culture that fosters open communication about systemic errors (e.g., in imagery analysis) as well as responsibility in R&D needs to be initiated urgently (82). This culture may be enhanced by recording design processes (83).

To summarize, control of autonomous systems matters, however, with divergent implications on how, what (regarding the function and situation), and in which circumstances. As we asked for supporting conditions of MHC within the respective CCW forum, tense associations between predictability and reliability, on the one hand, and precision and efficiency on the other hand, become visible. While precision may prove crucial for reliability and predictability, it is neither sufficient to aim solely for precision. Additionally, as reflected by the value of efficiency as well as various references to time as an important factor, developers need to weigh options for timely interventions against rapidity of actions. This supports prioritizing situational awareness of the human operator within the design process (26).

Lastly, and with respect to our finding of economic and military discourses not necessarily supporting MHC, restrictions of software development may neither be grounded in economic interests nor in a military advantage ratio. Paying special attention to human-LAWS perception among CCW participants, it is revealed that MHC heavily relies on a human-centric hierarchy. This may contradict a decentralized network-centric warfare. Yet, as commercial industries as well as the military are relevant (if not the most relevant) actors, deliberation among involved actors is necessary.

### *Limitations and Outlook*

In reacting to the legal (12) and technical (84) debates of LAWS and human control, we tried to put them into relation. Our work offers a first conduct on the question of MHC within the CCW GGE discourse. Naturally, this study yields several limitations. Parallel coding by several researchers and measurement of inter-rater reliability may improve the robustness of the findings. We presented a list of LAWS-incorporated values and sketched tendencies of MHC-supportive values and rather competitive relationships. Future work may focus on other relationships or ambivalences of values like "precision" or "time." At the same time, we tried to grasp discursive patterns, norms, and practices forming participants' perceptions of LAWS features by using VSD as a theoretical starting point. Crediting the initiative, one may consider that while the sample tried to represent existing power relations, it does not pay, in contrast to other work, special attention to racism or sexism (50), (51). Future studies may contribute by investigating these issues, in particular with regards to stakeholders' positions on MHC. Because of purposive inclusion of indirectly involved actors, the quantitative results of coding frequencies and word stem occurrences should not be taken to be absolute but rather be seen as an indication of relationships. To complement the focus on discourse, future analysis may shed light on material capabilities, long-term national interests, as well as innovation and defense politics. Our study may be accompanied by analyses comprising the entire body of the CCW expert group's documents to increase representativeness and validity of our findings, including statements of all actors and of more recent meetings. At the same time, focusing specifically on the different understandings of socio-technological values like accountability may prove valuable insights. Further, it may prove valuable to dig deeper into technological specifics, which CCW participants did not communicate in a very detailed way.

## Acknowledgment

This work was supported by the German Federal Ministry for Education and Research in the project KontiKat (13N14351), by the German Federal Ministry of Education and Research, the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE and by the Forum for Interdisciplinary Research (FIF) in the IANUS research fund for interdisciplinary research within the research project "IT research and development of concern." A two-page presentation of the motivation for this research was presented at Science Peace Security '19 in Darmstadt (85).

## Author Information

***Thea Riebe*** is a doctoral student and research associate at PEASEC, 64289 Darmstadt, Germany; email: riebe@peasec.tu-darmstadt.de.

***Stefka Schmid*** is a doctoral student and research associate at PEASEC, 64289 Darmstadt, Germany; email: schmid@peasec.tu-darmstadt.de.

***Christian Reuter*** is Professor at Technical University of Darmstadt and head of the research group "Science and Technology for Peace and Security" (PEASEC), 64289 Darmstadt, Germany; email: reuter@peasec.tu-darmstadt.de.

## References

[1] T. Riebe, S. Schmid, and C. Reuter, "Measuring spillover effects from defense to civilian sectors – A quantitative approach using LinkedIn," *Def. Peace Econ.*, pp. 1–13, 2020.
[2] E. Rosert and F. Sauer, "How (not) to stop the killer robots: A comparative analysis of humanitarian disarmament campaign strategies," *Contemp. Secur. Policy*, pp. 1–26, May 2020.
[3] M. A. C. Ekelhof, "Complications of a common language: Why it is so hard to talk about autonomous weapons," *J. Confl. Secur. Law*, vol. 22, no. 2, pp. 311–331, 2017.
[4] G. Adamson, J. C. Havens, and R. Chatila, "Designing a value-driven future for ethical autonomous and intelligent systems," *Proc. IEEE*, 2019, vol. 107, no. 3, pp. 518–525.
[5] T. Hellström, "On the moral responsibility of military robots," *Ethics Inf. Technol.*, vol. 15, no. 2, pp. 99–107, 2013.
[6] UNIDIR, "The weaponization of increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward," 2014; https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-considering-how-meaningful-human.
[7] V. Boulanin and M. Verbruggen, "Mapping the development of autonomy in weapon systems," p. 40, Nov. 2017. https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf
[8] M. W. Meier, "Lethal Autonomous Weapons Systems (Laws): Conducting a comprehensive weapons review," *Temp. Int'l Comp. LJ*, vol. 30, p. 119, 2016.
[9] S. Carvin, "Conventional thinking? The 1980 Convention on Certain Conventional Weapons and the politics of legal restraints on weapons during the Cold War," *J. Cold War Stud.*, vol. 19, no. 1, pp. 38–69, Jan. 2017.
[10] M. P. Cottrell, "Legitimacy and institutional replacement: The Convention on Certain Conventional Weapons and the emergence of the Mine Ban Treaty," *Int. Organ.*, vol. 63, no. 2, pp. 217–248, 2009.
[11] I. Bode and H. Huelss, "Autonomous weapons systems and changing norms in international relations," *Rev. Int. Stud.*, vol. 44, no. 3, pp. 393–413, 2018.
[12] R. Crootof, "A meaningful floor for 'Meaningful Human Control,'" *Temple Int. Comp. Law J.*, vol. 30, no. 1, pp. 53–62, 2016.
[13] K. Anderson and M. C. Waxman, "Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can," *Jean Perkins Task Force Natl. Secur. Law*, 2013.
[14] K. Anderson, D. Reisner, and M. Waxman, "Adapting the Law of Armed Conflict to Autonomous Weapon Systems," *Int. Leg. Stud.*, vol. 90, pp. 386–411, 2014.
[15] K. Anderson, "Why the Hurry to Regulate Autonomous Weapon Systems-But Not Cyber-Weapons?," *Temple Int. Comp. Law J.*, vol. 30, no. 17–41, 2016.
[16] B. Walker Smith, "Controlling humans and machines," *Temple Int. Comp. Law J.*, vol. 30, no. 1, pp. 167–176, 2016.
[17] J. I. Walsh, "Political accountability and autonomous weapons," *Res. Polit.*, vol. 2, no. 4, pp. 1–6, 2015.
[18] S. Umbrello and A. F. De Bellis, "A Value-Sensitive Design approach to intelligent agents," *Artif. Intell. Saf. Secur.*, no. January, pp. 395–410, 2018.
[19] S. Umbrello, "Beneficial Artificial Intelligence coordination by means of a Value Sensitive Design approach," *Big Data Cogn. Comput.*, vol. 3, no. 5, pp. 1–13, 2019.
[20] F. S. de Sio and J. van den Hoven, "Meaningful human control over autonomous systems: A philosophical account," *Front. Robot. AI*, vol. 5, no. FEB, pp. 1–14, 2018.
[21] A. Sharkey, "Autonomous weapons systems, killer robots and human dignity," *Ethics Inf. Technol.*, vol. 21, no. 2, pp. 75–87, 2019.
[22] K. A. Zweig, "Algorithmische Entscheidungen : Transparenz und Kontrolle," *Anal. Argumente Digit. Gesellschaft*, no. 338, 2019.
[23] Datenethikkommission, "Gutachten der Datenethikkommission," Berlin, 2019.
[24] D. Amoroso and G. Tamburrini, "What makes human control over weapons systems 'meaningful'?," 4, Aug. 2019; https://www.icrac.net/wp-content/uploads/2019/08/Amoroso-Tamburrini_Human-Control_ICRAC-WP4.pdf.
[25] N. Sharkey, "Staying in the loop: human supervisory control of weapons," in *Autonomous Weapons Systems: Law, Ethics, Policy*, C. Kreß, H.-Y. Liu, N. Bhuta, R. Geiß, and S. Beck, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2016, pp. 23–38.
[26] J. Weber and L. Suchman, "Human–machine autonomies," in *Autonomous Weapons Systems: Law, Ethics, Policy*, C. Kreß, H.-Y. Liu, N. Bhuta, R. Geiß, and S. Beck, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2016, pp. 75–102.
[27] M. Ekelhof, "Moving beyond semantics on Autonomous Weapons: Meaningful Human Control in operation," *Glob. Policy*, vol. 10, no. 3, pp. 343–348, 2019.
[28] M. A. C. Ekelhof, "Lifting the fog of targeting: 'Autonomous Weapons' and human control through the lens of military targeting," *Nav. War Coll. Rev.*, vol. 71, no. 3, pp. 61–95, 2018.
[29] T. Riebe, "Umgang mit Killerrobotern: Tagung, Evang. Akademie Loccum, 27.-29. Januar 2020," *Wiss. Frieden*, no. 2, 2020.
[30] M. C. Haas and S. C. Fischer, "The evolution of targeted killing practices: Autonomous weapons, future conflict, and the international order," *Contemp. Secur. Policy*, vol. 38, no. 2, pp. 281–306, 2017.
[31] F. Sauer, "Stopping 'Killer Robots': Why now is the time to Ban Autonomous Weapons Systems," *Arms Control Today*, vol. 46, no. 8, pp. 8–13, 2016.
[32] N. Schörnig, "Unmanned Systems: The robotic revolution as a challenge for arms control," in *Information Technology for Peace and Security*, C. Reuter, Ed. Wiesbaden, Germany: Springer, 2019, pp. 233–265.
[33] J. Altmann and F. Sauer, "Autonomous Weapon Systems and strategic stability," *Surviv. Glob. Polit. Strateg.*, vol. 59, no. 5, pp. 117–142, 2017.
[34] M. S. Dillon, "Network society, network-centric warfare and the state of emergency," *Theory, Cult. Soc.*, vol. 19, no. 4, 2002.
[35] A. K. Vice Admiral Cebrowski and J. J. Garstka, "Network-centric warfare: Its origin and future," 1998; https://www.usni.org/magazines/proceedings/1998/january/network-centric-warfare-its-origin-and-future.
[36] M. L. Cummings, "Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification? (Opinion)," *IEEE Technol. Soc. Mag.*, vol. 38, no. 4, pp. 20–26, 2019.
[37] A. Albers, S. Trautmann, T. Howard, T. A. Nguyen, M. Frietsch, and C. Sauter, "Semi-autonomous flying robot for physical interaction with environment," in *Proc. 2010 IEEE Conf. Robotics, Automation and Mechatronics, RAM 2010*, 2010, pp. 441–446.
[38] H. Chao, Y. Cao, and Y. Chen, "Autopilots for small unmanned aerial vehicles: A survey," *Int. J. Control. Autom. Syst.*, vol. 8, no. 1, pp. 36–44, 2010.
[39] P. Schramowski, C. Turan, S. Jentzsch, C. Rothkopf, and K. Kersting, "The moral choice machine," *Front. Artif. Intell.*, vol. 3, no. May, pp. 1–15, 2020.
[40] S. M. Adams and C. J. Friedland, "A Survey of Unmanned Aerial Vehicle (UAV) usage for imagery collection in disaster research and management," in *Proc. Ninth Int. Workshop on Remote Sensing for Disaster Response*, 2011.
[41] A. Hocraffer and C. S. Nam, "A meta-analysis of human-system interfaces in unmanned aerial vehicle (UAV) swarm management," *Appl. Ergon.*, vol. 68, pp. 66–80, 2017.
[42] R. C. Arkin, D. Lyons, J. Shu, P. Nirmal, and M. Zafar, "Getting it right the first time: predicted performance guarantees from the analysis of emergent behavior in autonomous and semi-autonomous systems," *Proc. SPIE 8387, Unmanned Syst. Technol. XIV*, 2012.
[43] A. Gray, Y. Gao, T. Lin, J. K. Hedrick, H. E. Tseng, and F. Borrelli, "Predictive control for agile semi-autonomous ground vehicles using motion primitives," in *Proc. American Control Conf. (ACC)*, 2012, pp. 4239–4244.

[44] P. Scharre, *Army of None: Autonomous Weapons and the Future of War*. Norton, 2018.
[45] R. C. Arkin, "The case for ethical autonomy in unmanned systems," *J. Mil. Ethics*, vol. 9, no. 4, pp. 332–341, 2010.
[46] M. C. Canellas and R. A. Haga, "Toward meaningful human control of autonomous weapons systems through function allocation," in *Proc. 2016 IEEE Int. Symp. Technology and Society (ISTAS)*, Mar. 2016, pp. 1–7.
[47] G. Hagele and D. Soffker, "A simplified situational environment risk and system reliability assessment for behavior assurance of autonomous and semi-autonomous aerial systems: A simulation study," in *Proc. 2017 Int. Conf. Unmanned Aircraft Systems, ICUAS 2017*, 2017, pp. 951–960.
[48] P. Chmielewski, "Ethical Autonomous Weapons?: Practical, required functions," *IEEE Technol. Soc. Mag.*, vol. 37, no. 3, pp. 48–55, 2018.
[49] IEEE, "Reframing Autonomous Weapons Systems," *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, 2016. (Online). Available: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_reframing_autonomous_weapons_v2.pdf.
[50] I. Bode, "Norm-making and the Global South: Attempts to regulate Lethal Autonomous Weapons Systems," *Glob. Policy*, vol. 10, no. 3, pp. 359–364, Sep. 2019.
[51] J. Santos de Carvalho, "A 'male' future?: An analysis on the gendered discourses regarding Lethal Autonomous Weapons," *Amsterdam Law Forum*, vol. 10, p. 41, Mar. 2018.
[52] R. Csernatoni, "The EU's technological power: Harnessing future and emerging technologies for european security BT - Peace, security and defence cooperation in post-Brexit Europe: Risks and opportunities," C.-A. Baciu and J. Doyle, Eds. Cham: Springer, 2019, pp. 119–140.
[53] K. M. Feigh and A. R. Pritchett, "Requirements for effective function allocation: A critical review," *J. Cogn. Eng. Decis. Mak.*, vol. 8, no. 1, pp. 23–32, May 2013.
[54] B. Friedman, P. H. Kahn, Jr., and A. Borning, "Value Sensitive Design and information systems," *Human-Computer Interact. Manag. Inf. Syst. Found.*, pp. 1–27, 2006.
[55] P. M. Asaro, "Modeling the moral user," *IEEE Technol. Soc. Mag.*, vol. 28, no. 1, pp. 20–24, 2009.
[56] M. L. Cummings, "Creating moral buffers in weapon control interface design," *IEEE Technol. Soc. Mag.*, vol. 23, no. 3, pp. 28–33, 2004.
[57] M. L. Cummings, "Automation and accountability in Decision Support System Interface Design," *J. Technol. Stud.*, vol. 32, no. 1, pp. 23–31, 2006.
[58] M. L. Cummings, "Integrating ethics in design through the value-sensitive design approach," *Sci. Eng. Ethics*, vol. 12, no. 4, pp. 701–715, 2006.
[59] S. M. Thornton, F. E. Lewis, V. Zhang, M. J. Kochenderfer, and J. C. Gerdes, "Value sensitive design for autonomous vehicle motion planning," in *Proc. 2018 IEEE Int. Vehicles Symp. (IV)*, 2018, pp. 1157–1162.
[60] C. Reuter, "Towards IT peace research: Challenges at the interception of peace and conflict research and computer science," *S+F (Security Peace)*, vol. 38, no. 1, pp. 10–16, 2020.
[61] C. Reuter, *Information Technology for Peace and Security*. Darmstadt: Springer Vieweg, 2019.
[62] J. Altmann, "Natural-Science/Technical Peace Research," in *Information Technology for Peace and Security - IT-Applications and Infrastructures in Conflicts, Crises, War, and Peace*, C. Reuter, Ed. Wiesbaden: Springer, 2019, pp. 39–60.
[63] C. Reuter, J. Altmann, M. Göttsche, and M. Himmel, "Natural-Science and Technical Peace Research: Definition, History and Current Work," *S+F (Security Peace)*, vol. 38, no. 1, 2020.
[64] V. Boulanin and M. Verbruggen, "Article 36 Reviews: Dealing with the Challenges posed by Emerging Technologies," Stockholm, 2017; https://www.sipri.org/sites/default/files/2017-12/article_36_report_1712.pdf.
[65] I. Verdiesen, V. Dignum, and I. Rahwan, "Design requirements for a moral machine for Autonomous Weapons," in *Proc. Computer Safety, Reliability, and Security: SAFECOMP 2018 Workshops*, 2018, vol. 11094 LNCS, pp. 494–506.
[66] I. Verdiesen, "The design of human oversight in Autonomous Weapon Systems," in *Proc. Twenty-Eighth Int. J. Conf. Artificial Intelligence (IJCAI-19)*, 2019, pp. 6468–6469.
[67] I. Verdiesen, "How do we ensure that we remain in control of our autonomous weapons?," *AI Matters*, vol. 3, no. 3, pp. 47–55, 2017.
[68] J. Pugliese, "Drones," in *Making Things International 1: Circuits and Motion*, M. B. Salter, Ed. Minneapolis, MN: Univ. of Minnesota Press, 2015, pp. 222–242.
[69] B. Friedman, P. H. Kahn, and A. Borning, "Value sensitive design and information systems," in *The Handbook of Information and Computer Ethics*, 2009.
[70] B. Friedman and H. Nissenbaum, "Bias in computer systems," *ACM Trans. Inf. Syst.*, vol. 14, no. 3, pp. 330–347, 1996.
[71] UNOG, "2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)," 2018; https://www.unog.ch/80256EE600585943/%28httpPages%29/7C335E71DFCB29D1C1258243003E8724.
[72] R. Huang, "RQDA: R-based Qualitative Data Analysis. R package version 0.2-8," 2016.
[73] M. Jørgensen and L. Phillips, *Discourse Analysis as Theory and Method*. London: Sage, 2012.
[74] I. Feinerer and K. Hornik, "tm: Text Mining Package. R package version 0.7-6," 2018.
[75] H. M. Roff and R. Moyes, "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." pp. 1–6, 2016. https://www.rdocumentation.org/packages/tm/versions/0.7-7.
[76] J. Pokorny, A. Norman, A. Zanesco, S. Bauer-Wu, B. Sahdra, and C. Saron, "Network analysis for the visualization and analysis of qualitative data," *Psychol. Methods*, vol. 23, no. 1, Nov. 2016.
[77] B. Giesen and R. Seyfert, "Collective identities, empty signifiers and solvable secrets," *Eur. J. Soc. Theory*, vol. 19, no. 1, pp. 111–126, 2016.
[78] B. Braun, S. Schindler, and T. Wille, "Rethinking agency in International Relations: performativity, performances and actor-networks," *J. Int. Relations Dev.*, vol. 22, no. 4, pp. 787–807, 2019.
[79] ICRC, "Autonomy, artificial intelligence and robotics: Technical aspects of human control," Geneva, Aug. 20, 2019.
[80] Australia, "Australia's system of control and applications for Autonomous Weapon Systems," Geneva, 2019; https://www.unog.ch/80256EDD006B8954/(httpAssets)/16C9F75124654510C12583C9003A4EBF/$file/CCWGGE.12019WP.2Rev.1.pdf; https://www.unog.ch/80256EDD006B8954/(httpAssets)/0F149C84EF8D36CEC125845C0051F592/$file/CCW_GGE.1_2019_WP.7.pdf.
[81] U.S.A., "Implementing International Humanitarian Law in the use of Autonomy in Weapon Systems," Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, Geneva, 2019; https://www.unog.ch/80256EDD006B8954/(httpAssets)/B2A09D0D6083CB7CC125841E0035529D/$file/CCW_GGE.1_2019_WP.5.pdf.
[82] J. van den Hoven, "Value sensitive design and responsible innovation," in *Responsible Innovation*, R. Owen, J. Bessant, and M. Heintz, Eds. Wiley Online Library, 2013, pp. 75–83.
[83] H. Aldewereld, V. Dignum, and Y.-H. Tan, "Design for values in software development," in *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, J. van den Hoven, P. E. VermaasIbo, and I. van de Poel, Eds. Springer Netherlands, 2015, pp. 831–845.
[84] K. H. Kindervater, "The technological rationality of the drone strike," *Crit. Stud. Secur.*, vol. 5, no. 1, pp. 28–44, 2017.
[85] S. Schmid, T. Riebe, and C. Reuter, "Meaningful Human Control of Lethal Autonomous Weapon Systems," in *SCIENCE PEACE SECURITY '19 - Proc. Interdisciplinary Conf. Technical Peace and Security Research*, 2019, pp. 196–200.
[86] V. Sehrawat, "Autonomous weapon system: Law of armed conflict (LOAC) and other legal challenges," *Comput. Law Secur. Rev.*, vol. 33, no. 1, pp. 38–56, 2017.

TS