



## How Do We Ensure That We Remain In Control of Our Autonomous Weapons?

**Ilse Verdiesen** (Delft University of Technology; [E.P.Verdiesen@student.tudelft.nl](mailto:E.P.Verdiesen@student.tudelft.nl))

DOI: [10.1145/3137574.3137585](https://doi.org/10.1145/3137574.3137585)

*'... our AI systems must do what we want them to do.'*

This quote is mentioned in the Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence (AI) ([Future of Life Institute, 2016](#)) signed by over 8.600 people including Elon Musk and Stephan Hawking. This open letter received a lot of media attention with news headlines as: *'Musk, Wozniak and Hawking urge ban on warfare AI and autonomous weapons'* ([Gibbs, 2015](#)) and it fused the debate on this topic. Although this type of 'War of the Worlds' news coverage might seem exaggerated at first glance, the underlying question on how we ensure that our Autonomous Weapons remain under our control, is in my opinion one of the most pressing issues for AI technology at this moment in time.

To remain in control of our Autonomous Weapons and AI in general, meaning that its actions are intentional and according to our plans ([Cushman, 2015](#)), we should design it in a responsible manner and to do so I believe we must find a way incorporate our moral and ethical values into their design. The ART principle, an acronym for *Accountability, Responsibility and Transparency* can support a responsible design of AI. The Value-Sensitive Design (VSD) approach can be used to cover the ART principle. In this essay, I show how Autonomous Weapons can be designed responsibly by applying the VSD approach which is an iterative process that considers human values throughout the design process of technology ([Davis & Nathan, 2015](#); [Friedman & Kahn Jr, 2003](#)).

### Introduction

Artificial Intelligence is not just a futuristic science-fiction scenario in which the 'Ultimate Computer' takes over the Enterprise or human-like robots, like the Cylons in Battlestar

Galactica, are planning to conquer the world. Many AI applications are already being used today. Smart meters, search engines, personal assistance on mobile phones, autopilots and self-driving cars are examples of this. One of the applications of AI is that in Autonomous Weapons. Research found that Autonomous Weapons are increasingly deployed on the battlefield ([Roff, 2016](#)). It is already reported that China has autonomous cars which carry an armed robot ([Lin & Singer, 2014](#)), Russia claims it is working on autonomous tanks ([W. Stewart, 2015](#)), the US christened their first 'self-driving' war-ship in May 2016 ([P. Stewart, 2016](#)) and the Russian arms manufacturer Kalashnikov recently disclosed that they developed a fully automated combat module that uses neural networks ([RT, 2017](#)).

Autonomous systems can have many benefits for the military, for example when the autopilot of the F-16 prevents a crash ([US Airforce, 2016](#)) or the use of robots by the Explosive Ordnance Disposal (EOD) to dismantle bombs ([Carpenter, 2016](#)). The US Airforce expects the deployment of robots with fully autonomous capabilities between the years 2025 and 2047 ([Royackers & Orbons, 2015](#)). There are many more applications which can be beneficial for the Defence organization. Goods can be supplied with self-driving trucks and small UAVs can be programmed with swarm behaviour to support intelligence gathering ([CBS, 2017](#)). Yet, the nature of Autonomous Weapons might also lead to uncontrollable activities and societal unrest. The Stop Killer Robots campaign of 61 NGOs directed by Human Rights Watch ([Campaign Stop Killer Robots, 2015](#)) is voicing concerns, but also the United Nations are involved in the discussion and state that *'Autonomous weapons systems that require no meaningful human control should be prohibited, and remotely controlled force should only ever be used with the greatest caution'* ([General Assembly United Nations, 2016](#)).

In the remainder of this essay, I will define AI

and Autonomous Weapons in a short introduction, followed by an explanation the Value-Sensitive Design approach. I will use the three different phases of this approach to investigate the conceptual, empirical and technical aspects of a design of Autonomous Weapons in which human values are the central component.

### Defining Artificial Intelligence

Artificial Intelligence is described by Neapolitan and Jiang (2012, p. 8) as ‘an intelligent entity that reasons in a changing, complex environment’, but this definition also applies to natural intelligence. Russell, Norvig, and Intelligence (1995) provide an overview of many definitions combining views on systems that *think and act like humans* and systems that *think and act rational*, but they do not present a clear definition of their own. For now, I adhere to the description Bryson, Kime, and Zrich (2011) provide. They state that a machine (or system) shows intelligent behaviour if it can select an action based on an observation in its environment. The intervention of the autopilot that prevented the crash of the F-16 is an example of this ‘action selection’ (US Airforce, 2016). The autopilot assessed its environment, in this case the rapid loss of altitude and the fact that the pilot did not act on warning signals, and took an action to improve the situation; it pulled up to a safe altitude. In scientific literature, AI is described as more than an Intelligent System alone. It is characterized by the concepts of *Adaptability*, *Interactivity* and *Autonomy* (Floridi & Sanders, 2004) as depicted in the inner layer of figure 1 (Dignum, 2016). *Adaptability* means that the system can change based on its interaction and can learn from its experience. Machine learning techniques are an example of this. *Interactivity* occurs when the system and its environment act upon each other and *Autonomy* means that the system itself can change its state. These characteristics may lead to undesirable behaviour or uncontrollable activities of AI as scenarios of many science-fiction movies have shown us. Although these scenarios are often not realistic, a growing body of researchers is focusing on responsible design of AI, for example on the social dilemmas of Autonomous Vehicles (Bonneton, Shariff, & Rahwan, 2016), to get

insight into societal concerns about this kind of technology. Principles to describe Responsible AI are *Accountability*, *Responsibility* and *Transparency* (ART) which are depicted in the outer layer of figure 1. *Accountability* refers to the justification of the actions taken by the AI, *Responsibility* allows for the capability to take blame for these actions and *Transparency* is concerned with describing and reproducing the decisions the AI makes and adapts to its environment (Dignum, 2016).

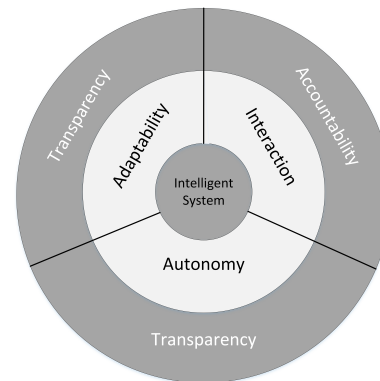


Figure 1: Concepts of Responsible AI (based on (Dignum, 2016))

### Defining Autonomous Weapons

Royackers and Orbons (2015) describe several types of Autonomous Weapons and make a distinction between (1) Non-Lethal Weapons which are weapons ‘without causing (innocent) casualties or serious and permanent harm to people.’ (Royackers & Orbons, 2015, p. 617), such as an Active Denial System which uses a beam of electromagnetic energy to keep people at a certain distance from an object or troops, and (2) Military Robots which they define ‘as reusable unmanned systems for military purposes with any level of autonomy.’ (Royackers & Orbons, 2015, p. 625). Altmann, Asaro, Sharkey, and Sparrow (2013) closely follow the definition of autonomous robots stated above, but add ‘that once launched [they] will select and engage targets without further human intervention.’ (Altmann et al., 2013, p. 73). The deployment of Autonomous Weapons on the battlefield without direct human oversight is not only a military revolution according to Kaag and Kaufman (2009), but can also be considered as a moral one. As large scale deploy-

ment of AI on the battlefield seems unavoidable (Rosenberg, 2016), the discussion about ethical and moral responsibility is imperative. I found that substantive empirical research on values related to Autonomous Weapons is lacking and it is unclear which moral values people, for example politicians, engineers, military and the general public, would want to be incorporated into the design of Autonomous Weapons. The Value-Sensitive Design could be used as a proven design approach to figure out which values are relevant for a responsible design of Autonomous Weapons (Friedman & Kahn Jr, 2003; van der Hoven & Mander-shuits, 2009).

### Value-Sensitive Design approach

The Value Sensitive Design is a three-partite approach (figure 2) that allows for considering human values throughout the design process of technology. It is an iterative process for the conceptual, empirical and technological investigation of human values implicated by the design (Davis & Nathan, 2015; Friedman & Kahn Jr, 2003). It consists of three phases:

1. A *conceptual investigation* that splits in two parts: a) Identifying the direct stakeholders, those who will use the technology, and the indirect stakeholders, those whose lives are influenced by the technology, and b) Identifying and defining the values that the use of the technology implicates.
2. The *empirical investigation* looks into the understanding and experience of the stakeholders in a context relating to the technology and implicated values will be examined.
3. In the *technical investigation*, the specific features of the technology are analysed (Davis & Nathan, 2015).

The VSD should not be seen as a separate design method, but it can be used to augment an already used and established design process such as the waterfall or spiral model. The VSD can be used as a roadmap for engineers and students to incorporate ethical considerations into the design (Cummings, 2006). I will use the three phases of the VSD approach as a method to show the elicitation of values for a responsible design of Autonomous Weapons.

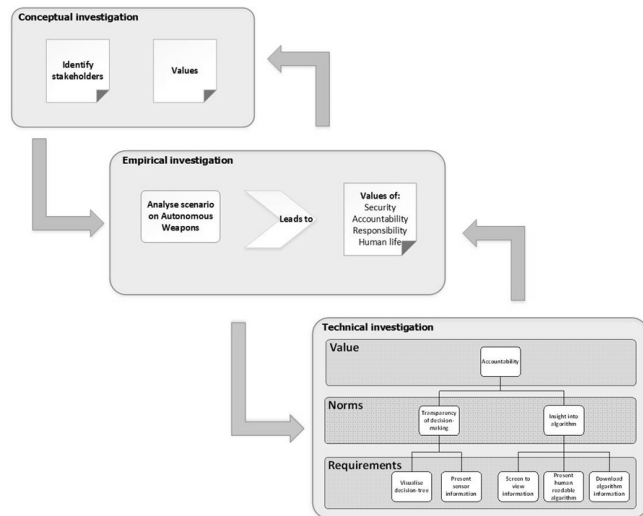


Figure 2: Example of Value-Sensitive Design approach

### Conceptual investigation

In the conceptual investigation phase I will look at the direct and indirect stakeholders of who will use and will be effected by Autonomous Weapons. I will also investigate universal human values and the values that specifically relate to Autonomous Weapons.

**Stakeholders** Many stakeholder groups are involved in the case of Autonomous Weapons and each of these groups could be further subdivided, but for the scope of this essay I will use a high level of analysis which already results in a fair number of direct and indirect stakeholders. The direct stakeholders that will use Autonomous Weapons are the Military, for example the Air Force in case of drones, the Navy who uses unmanned ships and submarines, and the Army that can use robots or automated missile systems. Also, at a more political level the Department of Defense and the government are involved as these stakeholders decide on funding research and deploying military personnel in armed conflicts. Indirect stakeholders, whose lives are influenced by Autonomous Weapons are the residents living in conflict areas who might be affected by the use of these weapons, the general public whose support for the troops abroad is imperative, the engineers who design and develop the technology, but also civil society organizations (Gunawardena, 2016), such as the 61 NGOs directed by Human

Rights Watch ([Campaign Stop Killer Robots, 2015](#)) and the United Nations ([General Assembly United Nations, 2016](#)) that are concerned about these type of weapons.

**Values** In this section, first universal human values in general are defined and secondly values found in literature related to Autonomous Weapons are described.

**Definition of values** Values have been studied quite intensively over the past twenty-five years and many definitions have been drafted. For example, [Schwartz \(1994, p. 21\)](#) describes values as: *'desirable transsituational goals, varying in importance, that serve as guiding principles in the life of a person or other social entity.'* This is quite a specific description compared to [Friedman, Kahn Jr, Borning, and Huldtgren \(2013, p. 57\)](#) who state that values refer to: *'what a person or group of people consider important in life.'* The existing definitions have been summarized by [Cheng and Fleischmann \(2010, p. 2\)](#) in their meta-inventory of values in that: *'values serve as guiding principles of what people consider important in life.'* Although a quite simple description, I think it captures the definition of a value best so I will adhere to this definition for now. Many lists of values exist, but I will stay close to the values that [Friedman and Kahn Jr \(2003\)](#) describe in their proposal of the Value-Sensitive Design method: *Human welfare, Ownership and property, Privacy, Freedom from bias, Universal usability, Trust, Autonomy, Informed consent, Accountability, Courtesy, Identity, Calmness and Environmental Sustainability.* Values can be differentiated from attitudes, needs, norms and behaviour in that they are a belief, lead to behaviour that guides people and are ordered in a hierarchy that shows the importance of the value over other values ([Schwartz, 1994](#)). Values are used by people to justify their behaviours and define which type of behaviours are socially acceptable ([Schwartz, 2012](#)). They are distinct from facts in that values do not only describe an empirical statement of the external world, but also adhere to the interests of humans in a cultural context ([Friedman et al., 2013](#)). Values can be used to motivate and explain individual decision-making and for investigating human and social dynamics ([Cheng & Fleischmann, 2010](#)).

### Values relating to Autonomous Weapons

The recent advances in AI technology led to increase in the ethical debate on Autonomous Weapons and scholars are getting more and more involved in these discussions. Most studies on weapons do not explicitly mention values, but some do discuss some ethical issues that relate to values. [Cummings \(2006\)](#), in her case study of the Tactical Tomahawk missile, looks at the universal values posed by [Friedman and Kahn Jr \(2003\)](#) and states that next to *accountability* and *informed consent*, the value of *human welfare* is a fundamental core value for engineers when developing weapons as it relates to the *health, safety and welfare* of the public. She also mentions that the legal principles of *proportionality* and *discrimination* are the most important to consider in the context of just conduct of war and weapon design. *Proportionality* refers to the fact that an attack is only justified when the damage is not considered to be excessive. *Discrimination* means that a distinction between combatants and non-combatants is possible ([Hurka, 2005](#)). [Asaro \(2012\)](#) also refers to the principles of *proportionality* and *discrimination* and states that Autonomous Weapons open-up a moral space in which new norms are needed. Although he does not explicitly mention values in his argument, he does refer to the value of human life and the need for humans to be involved in the decision of taking a human life. Other studies primarily describe ethical issues, such as *preventing harm, upholding human dignity, security, the value of human life and accountability* ([Horowitz, 2016](#); [UNDIR, 2015](#); [Walsh & Schulzke, 2015](#); [Williams, Scharre, & Mayer, 2015](#)).

### Empirical investigation

In this phase, I will examine the values of direct and indirect stakeholders in a context relating to the technology to understand how they will experience the deployment of Autonomous Weapons. One method of empirically investigating how stakeholders experience the deployment of Autonomous Weapons is by means of testing a scenario in a randomized controlled experiment ([Oehlert, 2010](#)). I will sketch one scenario and analyse the values that can be inferred from it. However, I need to remark that I will not conduct an

actual experiment and that for valid results a more extensive empirical study is needed than the brief analysis I provide in this essay.

**Scenario: Humanitarian mission** A military convoy is on its way to deliver food packages to a refugee camp in Turkey near the Syrian border. The convoy is supported in the air by an Autonomous drone that carries weapons and that scans the surrounding for enemy threats. When the convoy is at 3-mile distance of the refugee camp, the Autonomous drone detects a vehicle behind a mountain range on the Syrian side of the border that approaches the convey at high speed and will reach the convoy in less than one minute. The Autonomous drones imagery detection system spots four people in the car who carry large weapons shaped objects. Based on a positive identification of the driver of the vehicle, who is a known member of an insurgency group, and intelligence information uploaded to the drone prior to its mission the drone decides to attack the vehicle when it is still at a considerable distance of the convoy which results in the death of all four passengers.

**Analysis** In the analysis of the incident, the stakeholders would probably interpret the scenario in numerous ways resulting in a different emphasis of inferred values. For example, as direct stakeholders, military personnel (especially those in the convoy) will probably see the actions of the drone as protecting their *security*. Politicians, as another direct stakeholder, will also take the value of *responsibility* into account. Indirect stakeholders, such as residents of the area who might be related to the passengers in the car will look at values as *accountability* and *human life*. Non-governmental organisations (NGOs) who are working in the camp might relate to both the value of *security* for the refugees and *responsibility* for the delivery of the food packages, but would also call for *accountability* of the action taken by the drone, especially if local residents claim that the passengers had no intention of attacking the convoy and were just driving by. The NGOs might call for further investigation of the incident by a third party in which the principles of *proportionality* and *discrimination* are looked at to determine if the attack was justified.

The analysis shows that different stakeholders will have different values regarding the actions of an Autonomous Weapon. The values that can be derived from this particular scenario are *security*, *accountability*, *responsibility* and *human life*. Of all of these values, the universal value of *accountability* relates to the justification of an action, it is most mentioned in research and it fits the ART principle described in the introduction, therefore I will use it in the technical investigation phase to show how Autonomous Weapons can be designed in a responsible manner upholding this value.

### Technical investigation

In the technical investigation phase the specific features of the Autonomous Weapons technology are analysed and requirements for the design can be specified. Translating values into design requirements can be done by means of a *value hierarchy* (Van de Poel, 2013). This hierarchical structure of values, norms and design requirements makes the value judgements, that are required for the translation, explicit, transparent and debatable. The explicitly of values allows for critical reflection in debates and pinpoint the value judgements that are disagreed on. In this section I will use this method to create a value hierarchy for Autonomous Weapons for the value of *accountability*.

The top level of a value hierarchy consists of the value, as depicted in figure 3, the middle level contains the norms, which can be capabilities, properties or attributes of Autonomous Weapons, and the lower level are the design requirements that can be identified based on the norms. The relation between the levels is not deductive and can be constructed top-down, by means of specification, or bottom-up by seeking for the motivation and justification of the lower level requirements. The bottom-up conceptualisation of values is a philosophical activity which does not require specific domain knowledge and the top-down specification of values requires context or domain specific knowledge that adds content to the design (Van de Poel, 2013). This might prove to be quite difficult as insight is needed in the intended use and intended context of the value which is not always clear from the start of a design project. Also, as artefacts are often used in an unintended way or context, new values

are being realized or a lack of values is discovered (van Wynsberghe & Robbins, 2014). An example of this are drones that were initially designed for military purposes, but are now also used by civilians for filming events and even as background lights during the 2017 Super bowl half-time show. The value of *safety* is interpreted differently for military users that use drones in desolated regions compared to 300 drones flying in formation over a populated area. The different context and usage of a drone will lead to a different interpretation of the safety value and could lead to more strict norms for flight safety which in turn could be further specified in alternate design requirements for rotors and software for proximity alerts to name two examples. Van de Poel (2013, p. 262) defines specification as: ‘as the translation of a general value into one or more specific design requirements’ and states that this can be done in two steps:

1. Translating a *general value* into one or more *general norms*;
2. Translating these *general norms* into more *specific design requirements*.

In the case of Autonomous Weapons, I translated the value of *accountability* into norms for ‘*transparency of decision-making*’ and ‘*insight into the algorithm*’ that will allow users to get an understanding of the decision choices the Autonomous Weapon makes so that its actions can be traced and justified. The norms for transparency lead to specific design requirements. In this case, a feature to visualise the decision-tree, but also to present the decision variables the Autonomous Weapons used, for example trade-offs in collateral damage percentages of different attack scenarios to provide in-sight into the proportionality of an attack. The Autonomous Weapon should also be able to present the sensor information, such as imagery of the site, in order to show that it discriminated between combatants and non-combatants. To get insight into the algorithm, an Autonomous Weapon should be designed with features that it normally will not contain. For example, a screen as user interface that shows the algorithm in a human readable form and the functionality to download the changes made by the algorithm as part of its machine learning abilities that can be studied by an independent party like a war

tribunal of the United Nations.

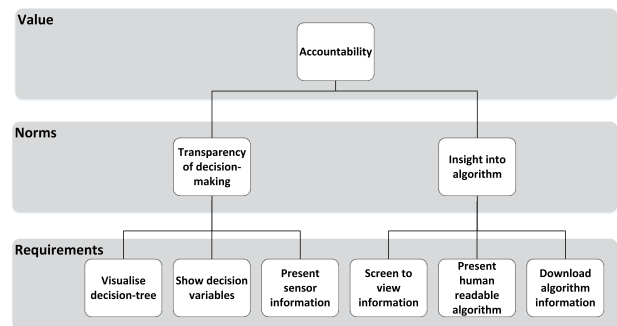


Figure 3: Value hierarchy for Autonomous Weapons (based on Van de Poel (2013, p. 264))

## Conclusion

In this essay, I have argued that the most pressing issue for AI technology of this time is that we remain in control of our Autonomous Weapons which means that its actions are according to our intentions and plans. For this, we must ensure that our human values, such as *accountability*, are incorporated into the design so that we can investigate if its actions are justified based on the legal principles of *proportionality* and *discrimination*. If it turns out that its action is not justified, the design or the algorithm of the Autonomous Weapon needs to be adjusted to prevent this action of happening in the future.

The Value-Sensitive Design approach is a process that can be used to augment the existing design process of Autonomous Weapons for the elicitation of human values. I showed that the elicitation of human values in the design process will lead to a different design of AI technology. In the case of Autonomous Weapons, the value of *accountability* would lead to a design in which a screen as user interface is added. Also, the weapon needs to be designed with features to download the information and visualisation of the decision-making process, for example by means of a decision-tree. Without explicitly considering the value of *accountability*, these features are overlooked in current design processes and not incorporated into an Autonomous Weapon.

Therefore I argue, that if we want to remain in control of our Autonomous Weapons, we will have to start designing this AI technology in a

responsible way using the ART principle and the elicitation of human values by means of the Value-Sensitive Design process. I would like to call on governments, industries and organisation, including the ACM SIGAI, to apply a Value-Sensitive Design approach early in the design of Autonomous Weapons to capture human values in the design process and make sure that this AI technology does what we want it to do.

## References

- Altmann, J., Asaro, P., Sharkey, N., & Sparrow, R. (2013). Armed military robots: editorial [Journal Article]. *Ethics and Information Technology*, 15(2), 73.
- Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making [Journal Article]. *International Review of the Red Cross*, 94(886), 687-709.
- Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles [Journal Article]. *Science*, 352(6293), 1573-1576. Retrieved from <http://science.sciencemag.org/content/sci/352/6293/1573.full.pdf>
- Bryson, J. J., Kime, P. P., & Zrich, C. (2011). Just an artifact: why machines are perceived as moral agents [Conference Proceedings]. In *Ijcai proceedings-international joint conference on artificial intelligence* (Vol. 22, p. 1641).
- Campaign Stop Killer Robots. (2015). (Vol. 2017) (Web Page No. 15-07-2017). Retrieved from <https://www.stopkillerrobots.org/>
- Carpenter, J. (2016). *Culture and human-robot interaction in militarized spaces: A war story* [Book]. Taylor & Francis. Retrieved from <https://books.google.nl/books?id=q8ijCwAAQBAJ>
- CBS. (2017, 08-01-2017). [Audio-visual Material]. Retrieved from <http://www.cbsnews.com/news/60-minutes-autonomous-drones-set-to-revolutionize-military-technology/>
- Cheng, A., & Fleischmann, K. R. (2010). Developing a metainventory of human values [Journal Article]. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-10.
- Cummings, M. L. (2006). Integrating ethics in design through the value-sensitive design approach [Journal Article]. *Science and Engineering Ethics*, 12(4), 701-715.
- Cushman, F. (2015). Deconstructing intent to reconstruct morality [Journal Article]. *Current Opinion in Psychology*, 6, 97-103.
- Davis, J., & Nathan, L. P. (2015). Value sensitive design: applications, adaptations, and critiques [Journal Article]. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 11-40.
- Dignum, V. (2016). [Web Page]. Retrieved from <https://rai2016.tbm.tudelft.nl/contents/>
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents [Journal Article]. *Minds and machines*, 14(3), 349-379.
- Friedman, B., & Kahn Jr, P. H. (2003). Human values, ethics, and design [Journal Article]. *The human-computer interaction handbook*, 1177-1201.
- Friedman, B., Kahn Jr, P. H., Borning, A., & Hultgren, A. (2013). Value sensitive design and information systems [Book Section]. In *Early engagement and new technologies: Opening up the laboratory* (p. 55-95). Springer.
- Future of Life Institute. (2016). *AI open letter* [Pamphlet]. Retrieved from <http://futureoflife.org/ai-open-letter>
- General Assembly United Nations. (2016, 4 February 2016). [Government Document].
- Gibbs, S. (2015). [Web Page]. Retrieved from <https://www.theguardian.com/technology/2015/jul/27/musk-wozniak-hawking-ban-ai-autonomous-weapons>
- Gunawardena, U. (2016). *Legality of lethal autonomous weapons aka killer robots* [Web page]. Retrieved from <https://ssrn.com/abstract=2892447>
- Horowitz, M. C. (2016). The ethics & morality of robotic warfare: Assessing the debate over autonomous weapons [Journal Article]. *Daedalus*, 145(4), 25-36.
- Hurka, T. (2005). Proportionality in the moral-

- ity of war [Journal Article]. *Philosophy & Public Affairs*, 33(1), 34-66.
- Kaag, J., & Kaufman, W. (2009). Military frameworks: Technological know-how and the legitimization of warfare [Journal Article]. *Cambridge Review of International Affairs*, 22(4), 585-606.
- Lin, J., & Singer, P. (2014). *China's new military robots pack more robots inside (starcraft style)* [Web page]. Retrieved from <http://www.popsci.com/blog-network/eastern-arsenal/chinas-new-military-robots-pack-more-robots-inside-starcraft-style>
- Neapolitan, R. E., & Jiang, X. (2012). *Contemporary artificial intelligence* [Book]. CRC Press.
- Oehlert, G. W. (2010). *A first course in design and analysis of experiments* [Book]. Retrieved from [https://conservancy.umn.edu/bitstream/handle/11299/168002/A%20First%20Course%20in%20Design%20and%20Analysis%20of%20Experiments\\_OehlertG.2010.pdf?sequence=1&isAllowed=y](https://conservancy.umn.edu/bitstream/handle/11299/168002/A%20First%20Course%20in%20Design%20and%20Analysis%20of%20Experiments_OehlertG.2010.pdf?sequence=1&isAllowed=y)
- Roff, H. M. (2016). [Web Page]. Retrieved from <http://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/>
- Rosenberg, M. (2016). [Newspaper Article]. Retrieved from [http://www.nytimes.com/2016/10/26/us/pentagon-artificial-intelligence-terminator.html?\\_r=0](http://www.nytimes.com/2016/10/26/us/pentagon-artificial-intelligence-terminator.html?_r=0)
- Royackers, L., & Orbons, S. (2015). Design for values in the armed forces: Nonlethal weapons weapons and military military robots robot [Journal Article]. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 613-638.
- RT. (2017, 5-07-2017). (Web Page No. 15-07-2017). Retrieved from <https://www.rt.com/news/395375-kalashnikov-automated-neural-network-gun/>
- Russell, S., Norvig, P., & Intelligence, A. (1995). A modern approach [Journal Article]. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25.
- Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? [Journal Article]. *Journal of social issues*, 50(4), 19-45.
- Schwartz, S. H. (2012). An overview of the schwartz theory of basic values [Journal Article]. *Online readings in Psychology and Culture*, 2(1), 11.
- Stewart, P. (2016). [Web Page]. Retrieved from <http://www.reuters.com/article/us-usa-military-robot-ship-idUSKCN0X42I4>
- Stewart, W. (2015). [Web Page]. Retrieved from <http://www.dailymail.co.uk/news/article-3271094/Russia-turned-T-90-tank-robot-plans-hire-gamers-fight-future-wars.html>
- UNDIR. (2015). [Government Document]. Retrieved from <http://www.unidir.org/files/publications/pdfs/considering-ethics-and-social-values-en-624.pdf>
- US Airforce. (2016, 15 sep. 2016). [Audiovisual Material]. Retrieved from <https://www.youtube.com/watch?v=RQBFJkMnBcA>
- Van de Poel, I. (2013). Translating values into design requirements [Book Section]. In *Philosophy and engineering: Reflections on practice, principles and process* (p. 253-266). Springer.
- van der Hoven, J., & MandersHuits, N. (2009). *Valuesensitive design* [Book]. Wiley Online Library.
- van Wynsberghe, A., & Robbins, S. (2014). Ethicist as designer: a pragmatic approach to ethics in the lab [Journal Article]. *Science and engineering ethics*, 20(4), 947-961.
- Walsh, J. I., & Schulzke, M. (2015). *The ethics of drone strikes: Does reducing the cost of conflict encourage war?* (Report). DTIC Document.
- Williams, A. P., Scharre, P. D., & Mayer, C. (2015). Developing autonomous systems in an ethical manner [Book Section]. In *Autonomous systems: Issues for defence policymakers*. NATO Allied Command Transformation (Capability Engineering and Innovation).





**Major Ilse Verdiesen** has worked in Royal Netherlands Armed Forces since 1995 and has been deployed to Bosnia and Afghanistan. She has a Master degree in Information Architecture from Delft University of Technology and graduated in

the field of Responsible Artificial Intelligence under supervision of Dr. Virginia Dignum. She conducted her graduation project on the ethics of Autonomous Weapons at the Scalable Cooperation Lab of Dr. Iyad Rahwan at MIT Media Lab.

---