Do robots equal humans? Anthropomorphic terminology in LAWS

Karolina Zawieska

Industrial Research Institute for Automation and Measurements PIAP
Al. Jerozolimskie 202, 02-486 Warsaw, Poland
(e-mail: kzawieska@piap.pl)

Summary: Given the complexity and novelty of autonomous robotic systems, obtaining an agreed-upon definition of LAWS constitutes a significant challenge. Depending on the approach, different definitions address different technical, legal, military and ethical aspects of the nature and use of LAWS. What these approaches have in common is a particular emphasis on the system autonomy and accountability. The main assumption here is that one of the reasons for robotics in general and LAWS in particular to raise widespread concerns is the human tendency to anthropomorphise robots, i.e. project human traits onto nonhuman objects. Terminology is a good example, since when we discuss both military and civil robots, we often use the same terms to describe human and robot characteristics, for example "autonomy", "intelligence" or "morality". While the human tendency to anthropomorphise is a universal "default schema" [2] that permeates all areas of human reality, the unreflective attribution of human traits to autonomous weapons systems is too risky to be accepted. Also, anthropomorphisation of LAWS may result in dehumanisation of human beings, decrease the public acceptance of robots in and outside the military domain and lead to rejection of robotic technology as a whole. In order to prevent confusion, we need to use terminology which fully addresses the difference between the anthropomorphic projections and the actual characteristics of robots. The goal is to maintain and further develop a human-centred approach towards accountability and the anthropocentric perspective on LAWS inherent to IHL [5], where the only subject accountable for the use of weapons is the human being.

While there are many different types of autonomous weapon systems, LAWS are seen here as a particular kind of robots. Different approaches towards accountability of robotics systems in general, and LAWS in particular, depend on how we define such systems. If we conceive robots and LAWS as agents, then we ascribe the human level of accountability to robots. If we define LAWS as tools, then it is the human being to be held responsible for LAWS performance. The former implies not only creating robots that are simply autonomous or intelligent but also building the systems that achieve (or surpass) *the human level* of autonomy, intelligence and so on. From this perspective, the core of robotics is anthropomorphism, i.e. the human tendency to attribute human qualities to nonhumans, be it objects, creatures or events.

In general, people anthropomorphise nonhuman objects in all cultures and societies. It has been proved that as little as motion or an illusion of intentionality in objects is sufficient to inspire anthropomorphic responses in humans [4]. Therefore, an illusion of humanlike life is particularly powerful in robots which are deliberately designed to appear and/or behave like humans. Whether it is possible to reproduce humans in artificial systems depends on how we understand what it means to be human. The underlying assumption here is that robots *appear* to be human rather than actually *are* human. This is why the key concept here is anthropomorphism, where one of the main ways to express the anthropomorphic thinking is the use of a particular terminology.

Since we often use the human qualities as a model for robot design, we also use the same terms to describe human and robot characteristics, for example autonomy or agency. Some of the terms clearly denote the machine-like nature of robot form and functions, for example "Artificial Intelligence (AI)", "Artificial Moral Agents (AMA)" or "machine learning". Most of the times, however, there is no difference between language we use to describe human and robot capabilities. This is of crucial importance as anthropomorphic terms convey specific meanings which shape our perception and thinking of robots. Depending on the robot's application, the human anthropomorphic bias can be seen as a useful mechanism which

facilitates human interaction with robots (for example, we can talk to robots) or a hindrance which leads robot users to create excessive expectations towards robots (for example, we expect robots to fully understand the human language and reasoning) [3]. The anthropomorphisation of military robots may be too risky to be accepted, an issue which has already been raised during the debate on LAWS [1, 5, 6]. The reasons why we should avoid anthropomorphism in this particular context include the following:

a) *Misleading framework of interpretation*: Human characteristics are often qualitative in nature, they vary between individuals and continue to change over time. This is why, it is not possible to successfully translate many of the human traits into engineering and computer terms. Therefore, while we use such terms as "autonomy" or "intelligence" to describe robots, they have different meanings when applied to humans. If we view the meaning of such terms as identical for humans and robots, then we realise that anthropomorphic terms often do not match the actual characteristics of robots (to the extent that the very existence of robot autonomy has been sometimes questioned). As a result, we risk assigning the human level of responsibility to robots based on the illusion of human characteristics rather than the actual humanness in robots. This is a misleading approach which hinders the understanding of LAWS in relation to not only humans but also other types of weapon.

b) *Dehumanisation of human beings*: As we ascribe more and more humanlike traits to robots where machines are supposed to equal or actually surpass humans, we increasingly objectify human beings. This is because there could be no anthropomorphisation of robots without developing a particular view of the human being. Nowadays, human mental states and body are often understood in terms of physical phenomena, where "humans" constitute a universal category of organisms to be mechanically reproduced in the machine. This is how we can claim quick success in building robots which, based on a few functionalities that resemble human traits, are seen as equal to humans or even more human than humans themselves (with higher levels of morality, better reasoning skills etc.). This has significant consequences for how we think of ourselves as humans. For example, what does it mean for us to be "autonomous" if we use the term "autonomy" to describe the robot ability to navigate from point A to B without human supervision? Thus, it is important to emphasise that the decisions we make with regards to LAWS are both a reason and a result of the entire approach towards human beings as individuals and as the human species.

c) *Unpredictable impact on the civil domain*: Autonomy has become a key requirement not only for military robots but also robotic systems used in the civil domain (for instance autonomous cars and social robots). While such systems are not directly involved in life-and-death decisions, many of their actions may influence human life and health. Thus, civil robotics too raises concerns with regard to the system autonomy and accountability, with anthropomorphisation of robots being one of the main issues. Therefore, it is important to emphasise that the decisions we make with regard to LAWS may have significant consequences for how we conceive robots and their responsibilities in the civil domain. Also, the excessive anthropomorphisation of LAWS and military robots may significantly decrease the public acceptance of robots and lead to rejection of robotic technology as a whole.

What we may need is a new approach towards anthropomorphic terminology in LAWS. This involves using the existing terms or developing a new terminology that would allow us to both embrace the human tendency to anthropomorphise robots and express the difference between the human and humanlike. Therefore, just as we use such terms as "Artificial Intelligence" or "Artificial Moral Agents", we should use the language that clearly denotes the machine-like rather than human-like nature of LAWS. Examples include "Artificial

Autonomy", "robotic autonomy", "quasi-autonomy" or "autonomous-like". The use of such terminology would allow us to develop conceptual rather than merely linguistic clarity, and achieve the following goals:

- *Accurately describe the current state of robotics technology:* While robotics systems are complementary to human skills in many areas, they are far from achieving the human levels of performance, in particular the human level of autonomy, intelligence and agency. Given the radically different nature of humans and machines, any future work will continue to develop the humanlike rather than human artificial systems. Using terminology which expresses the difference between the human and humanlike would facilitate discussion on LAWS and autonomous robots in relation to not only the human engagement and accountability but also other types of weapons.

- *Develop a human-centred approach towards accountability:* While it is unnecessary to view robotics as "all-or-nothing" technology, we shall never compromise on the anthropocentric perspective towards the responsibility for the creation and use of weapons. This is in the line with IHL, where the only subject accountable for the use of weapons is the human being. In order to achieve such a goal, we should avoid confusing humans with humanlike systems.

- *Increase public acceptance:* The use of autonomous military robots has raised significant concerns worldwide, with robots' humanlike qualities being among the main issues. The CCW debate on LAWS constitutes a chance to bring clarity to the public debate on LAWS and autonomous robots, with accurate terminology playing a crucial role in decreasing public concerns. In other words, it is the understanding of robotics rather than the increasing use of robots in and outside the military domain that may increase the public acceptance of robotic technology.

Conclusions

Anthropomorphisation is the universal human tendency, inherent to all areas of the human world, which can be hardly avoided or eliminated. While we project human traits onto LAWS and autonomous robots, however, we shall be very careful not to take our projections literally. Anthropomorphism relies on the illusion of human characteristics rather than the actual degree of humanness in the robotics systems. In order to avoid ascribing the human levels of accountability to the systems which only appear to be human, we should be clear about the difference between the human and humanlike. This is why it is important to contain such a difference in terminology we use to describe LAWS. The ultimate goal is to maintain the anthropomorphic perspective in the debate on LAWS, where the only subject responsible for the creation and use of weapons is the human being.

References

[1]   *Report on Activities. Convention on Conventional Weapons Informal Meeting of Experts on Lethal Autonomous Weapons Systems*, Campaign to Stop Killer Robots. 2014.
[2]   Caporael, L.R., *Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine.* Computers in Human Behavior 1986. **2**(3): p. 215-234.
[3]   Duffy, B.R., *Anthropomorphism and The Social Robot.* Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems, 2003. **42**(3-4): p. 177-190.
[4]   Heider, F. and M. Simmel, *An Experimental Study of Apparent Behavior.* The American Journal of Psychology, 1944. **57**(2): p. 243-259.
[5]   Kanwar, V., *Post-Human Humanitarian Law: The Law of War in the Age of Robotic Warfare.* Harvard Journal of National Security, 2011. **2**: p. 616.
[6]   Sharkey, N. and L. Suchman, *Wishful Mnemonics and Autonomous Killing Machines.* AISB Quarterly, 2013. **136**: p. 14-22**.**