Ethics and Autonomous Weapons Systems

University of Bristol, May 25th 2016

At the invitation of Bristol Global Insecurities Centre and British Student Young Pugwash, Dr Joanna Bryson and Dr Alexander Leveringhaus addressed the ethical and legal issues pertaining to the emergence of autonomous weapons systems. The panel touched upon how emerging weapons technologies, such as Unmanned Aerial Vehicles (drones), remote controlled tanks, or underwater vehicles, all present us with unprecedented ethical and legal issues. These issues included responsibility and culpability, regulation of these new technologies, concerns over moral agency, and the future role of humans in warfare.

Doctor Joanna Bryson focused primarily on ethics with regard to Artificial Intelligence (AI) in general. Dr Bryson became particularly interested in the concept of culpability during the Iraq war, especially in light of George Bush's demands for more ethical robots. She differentiated between moral agents who are responsible for their actions and moral patients that need to be looked after. Dr Bryson believes that robots cannot be assigned culpability because they cannot be regarded as moral agents. She argued that there is both a desire by some to treat AI as human but also to assign blame to them in order to avoid responsibility for our actions. Joanna highlighted that we fully author AI; we decide what it is and is not capable of. In her opinion we should retain moral agency as there is no logical reason for us to give this to AI.

Dr Bryson was involved in the creation of the principles of robotics for the UK, one of only three countries to have any. The five principles are:

- Robots are multi-use tools
- Humans, not robots, are responsible agents
- Robots are products. They should be designed using processes which assure their safety and security
- Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users
- The person with legal responsibility for a robot should be attributed

Joanna stated that she wants most is sustainability, a lack of suffering and an end to conflict, which in the case of AI can be achieved through better regulation. However, she also said that was she most fears is that regulation will reduce individual differentiation and hamper learning; both what she most wants and fears increase as intelligence is increased. In closing, Joanna made the following recommendations:

- That we should avoid making humanoid robots to avoid assigning them blame
- That we should think about our data like we think about our homes
- That although AI is in principle not much more dangerous than humans, it may be able to develop faster than our regulatory systems and that it solves the principal-agent problem too well

As a philosopher, working with engineers and artificial intelligence experts, Dr Leveringhaus started out emphasising the necessity for more interdisciplinary work in the field of emerging weapons technologies, to be able to address the multitude of ethical and moral implications that arise alongside the emergence of new weapons technology.

Leveringhaus addressed three key issues in the debates surrounding the emergence of automated weapons technologies:

- Rather than focus on ‘responsibility’, we need to ask: what should be the level of accepted ‘risk’ under conditions of unpredictability?
- That we should be cautious of accepting humanitarian arguments for automated weapons
- That warfare should remain humane

Reiterating a key point of Dr Bryson’s talk, Dr Leveringhaus firstly emphasised the importance of not attributing moral agency to robots. Addressing the ‘responsibility gaps’ that occur with the use of automated technologies, he suggested we shift conceptually to consider responsibility with regard to risk, rather than responsibility. He argued that the level of unpredictability ought to be considered against the risk level of their use; if these types of autonomous weapons are so unpredictable that it is far too risky to use them, it would be negligent to do so.

Moreover, Dr Leveringhaus addressed the argument that not all degrees of automation is necessarily morally problematic, and the prominent humanitarian argument that high levels of automaton reduces the potential for war crimes, making warfare more humane and armed conflict more civilized. He pointed to a particularly problematic aspect of machine autonomy: the functions that are related to a targeting process and the implications of removing human operators from the actual application of force. From this, Dr Leveringhaus made a powerful philosophical argument for the importance of human operators in warfare. What’s distinctive about humans is that they have the ability *not* to ‘pull the trigger’ in certain circumstances, even if ordered to do so. In this regard, the main problem with killer robots, is not the legal responsibility gaps, but the fact that they push back human operators that have the ability to be ‘conscientious objectors’.