

Week 6 SGTA Task

1. Asaro (2020) identifies several moral risks posed by the deployment of autonomous weapons, including: the harming of civil lives, arms races, unpredictability, increase of political conflicts and cyber attacks, and the rise of a new kind of weapon of mass destruction. Out of these, choose two risks that you find most relevant and urgent. How can these risks be characterised? How do they threaten to undermine human rights and values?

Asaro (2020) describes autonomous weapons as capable of identifying, selecting, and engaging a target with the intention to kill, all without meaningful human input or intervention. Asaro (2020) identifies the significant threat to civilian lives and infrastructure as a result of this lack of human oversight. An imperfect AI system could potentially fail at differentiating combatants and civilians, value the importance of a military objective over the potential civilian casualties, and lack the compassion or mercy that a human might exhibit in a military engagement. Asaro (2020) discusses how these systems fundamentally undermine rights and values from a deontological perspective. The use of autonomous weaponry, much like traditional weaponry, fundamentally deprives its victims of the right to life, bodily integrity and dignity. The use of automated weapons draw into question the validity and justification for the inflicted violence, making it difficult to determine what is intended and unintended as described by deontological ethics.

This concern ties in directly with the unpredictability of AI systems as they are currently implemented. An autonomous system without direct human oversight may escalate conflicts unnecessarily and behave unpredictably in circumstances for which it was not explicitly trained. Asaro (2020) identifies that as the time frame and geographical area over which these systems operate increases, it will become more difficult to predict how that system will behave. This issue is compounded when autonomous systems interact with unpredictable opposing autonomous systems, as there is no practical way of determining what the end result will be. This is significant in analysing the effect of these automated systems with regards to civilian harm. It is also a concern with regards to responsibility in the event of unpredictable behaviour and how that may influence the lives of combatants and non-combatants on both sides of a conflict. The unpredictable nature of autonomous systems interaction makes it difficult to assign or determine fault, and lack the human ability to recognise the potential impact of a decision and seek feedback from those higher up. Both the threat to civilians and the unpredictable nature of AI systems constitute significant and immediate moral and practical risks in the use of autonomous weaponry.

2. What is the agency without duties challenge posed by the deployment of autonomous weapons, according to Eggert (2025)? Which response to this challenge do you find most plausible and convincing and why?

The "Agency without Duties" challenge as described by Eggert (2025) describes a foundational disconnect between our moral framework as humans and how autonomous systems interact with others. In traditional moral theory, the right to defend ones self against harm assumes a right to not be harmed and a potential attacker's corresponding duty to do no harm to you. Eggert (2025) makes the distinction that autonomous systems are fundamentally different to human moral agents, as they lack sentience and are unbound by moral duties to others in the same way humans are. Eggert (2025) makes the argument that if an autonomous system cannot possess duties, is it possible for us to say that victims of harm perpetrated by autonomous systems are also victims of rights violations.

Eggert (2025) also outlines a number of responses to this approach, including the 'Bad Luck Approach', the 'Thomsonian Approach' and the 'Moral Room Approach'. Of these, I find the Thomsonian approach the most convincing. The Thomsonian approach as described by Eggert (2025) rejects the concept that rights violations must be caused by a corresponding breach of duty. Although it requires us to adjust our traditional view of moral agency, it maintains the integrity of the rights framework in the specific context of autonomised harm. It allows the victims of autonomised harm to be seen as having their rights violated, even in the absence of a traditionally morally responsible agent. While imperfect, this approach is much more practical than the unproductive 'Bad Luck' approach, and the 'Moral Room' approach which supposes AI should 'act' as if it possesses moral duties.