What so bad about ‘Killer Robots’?

Dr Alex Leveringhaus,
Oxford Institute for Ethics, Law and Armed Conflict
Department of Politics and IR
Oxford Martin School
University of Oxford
alexander.leveringhaus@politics.ox.ac.uk
What is a robot?
Adapted from Winfield, 2012

• Robots have certain functions:
  1. Sensing
  2. Signalling
  3. Moving
  4. Energy
  5. Intelligence

⇒ Functions are usually integrated into a **body**.

• Def I: Artificial device that can *sense* its environment and *purposefully* act on or in that environment.

• Def II: An embodied artificial intelligence.
Definitions, definitions

• No clear-cut definition available.
  – software robots (e.g. Computer viruses) lack a body yet they can still move around and interact with their environment.

• Modern weapons systems may incorporate/interact with robotic elements, but they may not be robots themselves.
  – Computer system that analyses flight patterns of hostile missiles and then passes on the info to a robotic weapons platform.
Why do we want robots?

- Robots are often used for tasks that are:
  1. DULL
  2. DIRTY
  3. DANGEROUS
  4. DODGY

   and that humans are probably not very good at...
The Dragonrunner Robot

- Developed by Carnegie Mellon University and Automatik, Inc.
- For use in urban environment.
- Can fit into a backpack.
- Used by the British Army in Afghanistan for the detection of Improvised Explosive Devices and bomb disposal.
The Alpha Dog

• Big Dog/Alpha Dog robot is developed by Boston Dynamics, NASA, Foster Miller, and Harvard.
• Robotic ‘pack mule’. Legged Squad Support System (L3S).
• Accompany soldiers in difficult terrain.
The Predator Drone (MQ-1)

• Uninhabited aerial vehicle. Manufactured by General Atomics.
• Deployed by US military and intelligence services (CIA).
• Used in the War on Terror for ‘targeted killings’.
• Can also be used for reconnaissance missions.
• Remotely operated.
Taranis (in study)

- Unmanned aerial vehicle manufactured by BAE systems.
- Stealth plane.
- Can track and destroy radar stations without assistance from an operator.
Iron Dome

- Used by Israel in order to detect and destroy hostile rockets.
- Analyses the threat.
- Issues command to the operator. Operator decides on whether to launch counterattack.
- Command issued to missile firing unit.
- Command can also be issued to a UAV...
• Sentry Robot-manufactured by Samsung.
• Deployed in DMZ between N and S Korea.
• Stationary robot.
• Two machine guns and one gun that fires rubber bullets.
• Can track and identify targets that are up to 2.5 kilometres away.
• Has a microphone and speaker system.
Putting the ‘Killer’ into the Robot

• Non-lethal military robots
  – Alpha Dog.

• Non-lethal robots with lethal side-effects
  – Electronic countermeasure systems.

• Robotic and computer-based targeting systems (RTS/CBTS).
  – Designed in order to apply force to a target.
  – Application of force is intentional.

=> Sentry Robot, Taranis
In-The-Loop-RTS

- Remote controlled RTS are also known as tele-operated robots:
  1. Sensors.
  2. Transmit information/images via a video-satellite link.
  3. Information relayed on video screen to operator (or group of operators).
  4. Targeting decision made via chain of command.
  5. Enactment of decision by operator via remote control.

=> In-the-loop-systems (for definitions, see DoD report)
Out-of/on-the loop-RTS

• Machine acts independently of the operator.
• On the loop:
  – The human operator is on stand by and can intervene in what the system does.
  – Question of effectiveness of stand-by option.
  – Transparency of system.
• Out of the loop:
  – The human operator is not on stand-by. Machine acts independently once it has been launched.
Autonomous robots???
Moral and Operational Autonomy

• Moral autonomy:
  – Moral autonomy: act for reasons that we give ourselves (Kant, Rousseau, Rawls).

• Operational autonomy:
  – Carry out tasks independently from an operator
    • Technological Capacity: Machine (M) can take care of itself. M does not require an operator to carry out a specific task.
    • Self-Direction: M is allowed to act within a certain domain. Restriction are lifted.
The targeting process

• RTS/CBTSauton may be autonomous with regard to many tasks.
  – Flying.
  – Landing.
  – Collecting data.
• However, what critics of autonomous weapons seem to object to is possibility that the decision-making loop underlying the targeting process is automated:
  1. Acquire information about a potential target.
  2. Analyse information about the potential target.
  3. Decide whether to engage the target.
  4. Enact the decision.
The Yuk Factor

• ‘Robots must not kill humans.’
• ‘Machines must not make decisions about life and death.’
• ‘War becomes riskless.’

(See Sharkey)
How ‘yucky’ are Killer Robots?

- **Does it really make a difference if a robot or a human makes the decision to apply force?**
  - High altitude bombing during Kosovo War. Michael Ignatieff ‘Virtual War’.

- **Do machines really make decisions about life and death?**
  - Big and small decision-making loops. Will stand within chain of command. Pre-programmed.

- **Riskless War or Riskless Killing?**
  - Must we assume the equal risk of being killed in order to be allowed to kill?
  - Michael Walzer’s ‘Moral Equality of Soldiers’.
  - Agent-centric vs. Target centric approaches. Long-standing debate within the ethics of self-defence and the ethics of killing in war (see Jeff McMahan).
The Non-Compliance Objection

• RTSauton cannot comply with key ethical principles.
  – Discrimination. Distinguish between combatants and non-combatants.
  – Necessity and proportionality.

• Discrimination:
  – Integrate barcode into uniform.

• Proportionality and necessity:
  – Assessments of proportionality and necessity do not merely involve number-crunching. They involve ongoing value judgements (see Thomas Hurka).
  – An ordinary soldiers will find it difficult to make this judgement. Chain of Command.
The Abdication of Responsibility Objection (Andreas Mathias)

• Many machines already have some level of artificial intelligence and, in order to optimise their behaviour, are equipped with learning mechanisms.

• Machines will exhibit emergent behaviour.

• Depending on what they have learnt, machines may act in ways that are not entirely predictable.

• Because the operator cannot foresee what the machine might do, he is not responsible

⇒ Responsibility gaps, in which no one is responsible.
The Abdication of Responsibility
Objection
(Robert Sparrow)

• Thesis: the use of military force is (prima facie) permissible if, and only if, someone can be held responsible for it.

1. RTSauton with high levels of AI may develop reasons for killing that originate within the system, rather than the operator.

2. We cannot hold the operator responsible for the decision the machine makes, nor the machine itself.

3. No one is responsible for killing.

⇒ The use of RTSauton (with high AI) is impermissible.
Rebellious Roberto

Rebellious Roberto, the commander of a powerful rebel faction operating in the jungles of South America, orders a group of drugged child soldiers to attack government forces who are stationed in a village. The child soldiers kill the soldiers. Once they are done with the government soldiers, they massacre the entire civilian population of the village.
A plea for excuses?

• Would Rebellious Roberto only be responsible for the killing of the government soldiers (whose death he explicitly ordered)?
• Is Rebellious Roberto exculpated from wrongdoing?
  – RR had no, or limited, control over what the children did, once they had killed the government soldiers.
  – RR did not know that the children would also massacre the villagers.
• Neither of the two defences offers sound exculpating reasons.
  – RR provided the children with weapons and drugged them.
  – RR knew that it was unsafe to ask the children to carry out the mission. He should expect them to get out of control.
    (On these and related issues, see Lepora and Goodin).
Is responsibility really the issue?

• Operators of RTSauton (with high AI) can sometimes be held responsible for what the machine does even though
  – they did not have direct control over what the machine did; and
  – they did not know that the machine would engage in a particular wrongdoing.

• Operators, as well as the military in general, know that AI is never 100% safe. The use of Killer Robots (with high AI) always entails the imposition of risks on combatants and non-combatants.

• Warring parties are responsible for the imposition of these risks.
  – Michael Walzer: Due care.
Risk, responsibility and negligence

• The *real question* is whether the imposition of risks is justified (that is, does not violate any moral duties) or negligent.

• The use of an unsafe weapon means that the operator is, indeed, culpable.
  – This may even be true if nothing bad happens. For some legal philosophers, the imposition of an unjustifiable risks is already a form of harm, even if the risk does not materialise (see Finkelstein).

• The justified use of RTSauton means that the operator is still responsible, though not culpable.
The Neo-Classical Objection

• Neo-classical or revisionist just war theorists, such as Jeff McMahan or Cecile Fabre, assume that killing in war is, in principle, permissible if it is done to further a just cause.
  – Combatants who kill in pursuit of an unjust cause violate the rights of the combatants they kill.
• Program RTSauton with high levels of AI in order to assess the overall justness of the conflict they fight in?
  – Even if this was possible it is hardly in the interest of governments that use these systems to program a conscientious objection function into these weapons.
• Human combatants may at least conscientiously object to serving in a campaign if they are convinced it is unjust.
• Robotic combatants, on the other hand, may not be capable of doing so. Neo-classicists could thus reject the development of robotic combatants.
What so bad about Killer Robots, then?

• Not clear that Killer Robots raise morally distinctive issues that have not already been raised by other systems.

• The two most convincing arguments:
  1. Can these systems be designed in such ways that their use is non-negligent? If not, then it may be impermissible to develop and deploy Killer Robots.
  2. Neo-classicists may object to the creation of artificial agents that, though equipped with high levels of AI, are incapable of distinguishing between just and unjust wars.

• At the moment, the use of robotic systems appears rather limited.
Thank you!