Advanced Artificial Intelligence

Ethics problem

Artificial Intelligence in warfare

Björn Bengtsson
Nicklas Branding
Hiwot Lassa
Welemhret Baraki

24/10-2020

1. AI Introduction

The general definition of Artificial Intelligence is behaviour that feels humanlike (M. Cummings, 2017). One clear example is programs that mimic human decision making like looking at a photo and deciding what it is (Image recognition). There are different types of AIs, there are machine learning that apply what it has learned from its former experience. In Machine Learning the programing often learns by reading a big amount of data and learning from it like photo recognition. There is also AI that doesn't learn from its former experience like a house thermostat (M. Cummings, 2017) or monitoring your blood pressure.

A big Ethical problem that occurs at the moment around AI is automation. The problem is that robots are starting to replace human jobs. A clear example is an automated cashier. Finland estimates that over a third of the Finnish people are at risk to lose their job to a computer. With more complicated AI, more jobs will be at risk (Pajarinen, 2014).

2. Introduce AI Warfare

Recent advancement and in artificial intelligence have a lot of ethical issues provided in military warfare and Cyber security warfare. The main challenges of artificial intelligence in warfare are meaningful human control (Hallaq, Somer, Osula, Ngo, & Mtichener-Nissen, 2017; "Killer Robots and the Concept of Meaningful Human Control | Human Rights Watch," 2016)

Fully Autonomous AI weapons will without the human intervention and control and make decisions, identify targets and take subsequent actions that humans can do ("Killer Robots and the Concept of Meaningful Human Control | Human Rights Watch," 2016). Still many countries are discussing how to use autonomous weapons and human control is the burning issue and focal point for experts as well as countries ("Killer Robots and the Concept of Meaningful Human Control | Human Rights Watch," 2016).

In the present day there are no international treaties prohibiting the development and deployment of AI in war, which has given room for numerous countries to develop novel AI Lethal Autonomous weapons. The countries that are actively pursuing development of LAWs are China, USA, Russia and the United Kingdom. These countries raise a significant probability of LAWs being used in the future, which makes understanding the ethical and legal issues concerning AI in warfare (Umbrello, Torres, & De Bellis, 2020).

3. AI Ethical Dilemmas

One huge benefit with using AI in warfare is the lowering cost of human life. This benefit is already shown to improve the life expectancy of the human life with the use of AI robots such as a bomb diffuser called iRobot Packbot seen in figure 1 (Sharkey, 2008), search and rescue and precision targeting with satellites are widely used in the US Army (Hallaq et al., 2017). Moreover, the US Army is developing and testing the use of AI robot soldiers in the battlefield seen in figure 2, which will later replace the human soldiers leading to lowering the cost of human life on the battlefield ("Killer Robots and the Concept of Meaningful Human Control | Human Rights Watch," 2016).

Figure 1 iRobot PackBot. PackBot performs dangerous missions like bomb diffusion. (Noel Sharkey, 2008, *Cassandra or false prophet of Doom: AI Robots and war*)

Furthermore, the physiological benefit for human soldiers that have taken a life on the battlefield will drastically be reduced with the use of AI soldiers. Returning soldiers of war usually suffer from post-traumatic stress disorder, with the help of AI soldiers this psychological problem will be gone (Umbrello et al., 2020).

A big disadvantage is that starting a war becomes easier. Because human cost is lowered, you don't really have to think much about the human lost in the army (Sharkey, 2008). If a country handles a war poorly and loses humans, people will react, but if they just lose robots that they can easily replace, no one will bat an eye.

Figure 2 The Modular Advanced Armed Robotic System (Maars). Maars is an AI robot with a semi-auto rifle attached to its body (Noel Sharkey, 2008, *Cassandra or false prophet of Doom: AI Robots and war*)

The Ai does not question orders, which makes it easier for the one controlling to do more horrors easier. They can easily be ordered to kill all the people in an area and ethical cleansing. A human can do the same thing as well, but the difference is that a human can think through the order, question it and say no. The robot will not have this capability.

Another case is that even if the order is clear, the robot will do something opposite or handle it poorly. An example would be shooting all bystanders on the way to their target. The reason this can happen is because of lack of experience and bugs.

(M. Cummings, 2017) Claim that the AI in the military sector does not get as much training and testing such as AI in the commercial sector. The commercials have more data they can access, they can create it from other applications or buy it, and they can test the AI directly on the users most of the time. If something wrong happens it will mostly be an annoyance. But the military will have a hard time getting data, they have to create it for themselves and test them in controlled environments.

The problem is that in the training they cannot train the AI for all scenarios, so how will it react to a scenario it has not trained on? Also they can miss important data in their data training and that can lead to big consequences. It happens in the commercial sector for example the AI for cropping on twitter. Twitter trained an AI to crop images so that it focuses on human faces, the problem is that it will crop out black people and only focus on the white people (Hern, 2020). The same thing will probably happen to AI warfare, they will probably miss cases and the end behaviour for the cases can be brutal. Another problem is also warfare and has less development and testing compared to the commercial sector. The problem is that the commercial sector also has AI that makes poor choices and behaves wrong. If the military spends less time and resources, it's more likely that the AI will have more errors and behave wrong. Which can lead to more casualties.

4. Issues in AI-Powered Autonomous Weapons

A risk with more military being controlled by computers and AI is that people with programming knowledge can hack the robots and misuse them. For instance terror networks, governments or a private person. Which is why a set of rules for how the AI system is built is required (Umbrello et al., 2020).

A risk with AI systems in warfare is that the system cannot distinguish between combatants and non-combatants if it has not been trained on this sort of data. Which is an issue because there is not enough diversity in the datasets currently used. Another complexity is to distinguish between civilians, non-civilians and wounded civilians and combatants (Hallaq et al., 2017; "Killer Robots and the Concept of Meaningful Human Control | Human Rights Watch," 2016). How should the AI distinguish and draw the line between military and civilians?

The international humanitarian law (IHL) in 1949 built precautions to spare the civilian population, individual civilians and civilian objects in wartimes, with this also that no attack should necessarily give the combatant any prolonged suffering, meaning that the AI must be trained to make complete accurate precisions for when to launch an attack on a combatant to completely reduce the prolonged

suffering of combatants (Hallaq et al., 2017). Another significant legal and ethical issue is given in the laws of war (LoW) which is outlined in the Geneva and Hague convention stating that it is prohibited to kill civilians and surrendering soldiers except under strict circumstances where it is justified to possibly prevent further loss of life (Umbrello et al., 2020). How will the robot be able to make that decision and will the AI understand when a person starts killing civilians?

5. **Philosophical Aspects of AI Ethics in warfare**

When the robot does an action which is inhuman or punishable or an error. Who is responsible? Is the one giving the command, is it the programmers? The higher ups giving command? It can be hard to tell because the AI has an intelligence of its own but it's still made by humans and given order by them. You can't put the blame on the AI.

It is even harder to pinpoint responsibility when the AI does something wrong that was not ordered. Is it the programmers that skipped the case and did not test enough, was it the higher ups that used the AI even if the programmers said it was not ready. Or the one giving orders directly if he knew about the flaws.

In normal acts of war, returning humans are usually affected by post-traumatic stress, given by explosions, killed teammates and killing of enemies (Umbrello et al., 2020).This is a normal part of our consciousness to reflect back on the things we humans do. An AI will not reflect on its actions. Can they be called intelligent if they cannot reflect on their past actions?

Can a machine act intelligently? Can it solve any problem that a person would solve by thinking? We think that a robot can follow orders without a problem. But when you ask the AI to make choices for themselves it is not sure that it will make the right decision. Can they make tactical decisions on all the abstract data and context?

6. Future of AI in warfare and Possible Solutions

AI will be the next treat in the future if not controlled in a legalized and institutionalized way. Autonomous weapons are excellent in engaging the targets without human intervention. Artificial Intelligence technology has extended to a point where the installation of AI is practicable and the damage is getting high. Switching human soldiers to machines in AI is a good thing because it reduces human fatality and war but it has a side effect for lowering the threshold for going to battle (Hallaq et al., 2017).

Currently, there is a global issue or prediction about the development of the AI arms race. If this technology goes ahead with AI weapon development, the endpoint of this technological path will be autonomous. Autonomous weapons are perfect for groups who needs to deal great damage, with these autonomous weapons, the groups can assassinate, destabilize nations, reduce populations and selectively kill a particular ethnic group. Therefore, the solution is not to start the development of such weapons, because they are not good for humans.

The possible solutions to the problems is to introduce governing rules explicitly stating governments and individuals not to misuse Autonomous weapons. Autonomous weapons systems should have sufficient datasets making it possible for the AI to get enough training in order to be used in the

battlefield. With sufficient datasets developing AI models used on the battlefield leads to easier learning and smarter decisions on identifying targets. The autonomous weapons detection capabilities is limited to a very small amount of data attributes, for instance targets can be identified with their faces, military clothes and so on. For the future the Artificial General Intelligence methods with reinforcement learning will be the future solution to alleviate the problems of Autonomous weapons in identifying targets. Another solution to identify combatants is that they require to wear a tag that the robots can read. It will make it easier for an AI to distinguish civilians from combatants. The problem is that the majority of people would not want to wear a big bulletin board that says that the robot is allowed to kill them.

Human Control over Autonomous weapons is critical for the whole world, governments and individuals. So as to minimize the risk and mitigate the problems of AI in warfare human intervention and control is crucial. Preparing standard rules that the countries must follow to use AI powered weapons. A good solution would be that some set of rules are made in the UN, because all the countries in the UN must follow them. The problem is that countries having veto, like the USA, China or Russia are not interested, one solution is to have the UN or the Ethic committee agree upon a standard for what the AI can do and not do. Autonomous Weapons will affect the Environmental condition, the solution is to develop power efficient devices and governments should have to minimize wars and reconcile to peaceful negotiations.

Instead of having humans controlling the air and water borders, AI will be monitoring the air and water border. Moreover, the human cost is lowered and it can end like nukes, nearly everybody has it but do not use it because it is too effective

Autonomous weapons can be hacked by governments and hackers to misuse wrongly, the possible solution is to enforce strong security standards. The autonomous weapons preprogramed target parameters, and learning algorithm is expected to meet their Goals and not easy to change. The input parameters for building a model should have to follow the international laws and standards and able to succeed their intended goal.

The countries with a lot of resources will easily be able to build big armies. It is because they will be able to mass produce the AI soldiers when it is cheap enough. They also do not need to train all the robots, they just train some of the same model, and copy what the trained robots learned to the others. That is much cheaper and takes less time to train a human soldier. Each human soldier needs to be trained and that takes a lot of time and costs a lot of resources. Also when you lose a big amount of human soldiers, you cannot easily replace them, you need to train new ones. This does not apply to robots, if they lose many, they can just build new ones. The consequences of this is that countries that will have developed AI robots first, will be able to defeat the other countries.

References

Hallaq, B., Somer, T., Osula, A.-M., Ngo, K., & Mtichener-Nissen, T. (2017). *Artificial Intelligence Within the Military Domain and Cyber Warfare*. Dublin, Ireland. Retrieved from http://wrap.warwick.ac.uk/94297http://www.academic-bookshop.com/ourshop/

Hern, A. (2020, September 21). Twitter apologises for "racist" image-cropping algorithm | Technology | The Guardian. Retrieved October 17, 2020, from https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm

Killer Robots and the Concept of Meaningful Human Control | Human Rights Watch. (2016, April 11). Retrieved October 9, 2020, from https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control

M. Cummings. (2017). *Artificial Intelligence and the Future of Warfare*.

Pajarinen, M. (2014). *Computerization Threatens One Third of Finnish Employment*. Retrieved from http://v.gd/grjSKN

Sharkey, N. (2008). Cassandra or false prophet of doom: AI robots and war. In *IEEE Intelligent Systems* (Vol. 23, pp. 14–17). https://doi.org/10.1109/MIS.2008.60

Umbrello, S., Torres, P., & De Bellis, A. F. (2020). The future of war: could lethal autonomous weapons make conflict more ethical? *AI and Society*, *35*(1), 273–282. https://doi.org/10.1007/s00146-019-00879-x