When AI goes to war: youth opinion, fictional reality and autonomous weapons

By Elvira Perez Vallejos, Rob Wortham and Eugene Miakinkov

This paper relates the results of deliberation of youth juries about the use of autonomous weapons systems (AWS). The discourse that emerged from the juries centered on several key issues. The jurors expressed the importance of keeping the humans in the decision-making process when it comes to militarizing artificial intelligence, and that only humans are capable of moral agency. They discussed the perennial issue of control over AWS and possibility of something going wrong, either with software or hardware. Concerns over proliferation of AWS and possible arms race also entered the discussion and the jurors were skeptical about the possibility of regulation and compliance once AWS enter military arsenals. We conclude that the juries were very apprehensive and hostile to the introduction of autonomous weapons systems into military conflicts.

Keywords: Artificial Intelligence; Youth Public Opinion; Youth Juries; Ethics; Autonomous Weapons; Suspension of Disbelieve; Intelligence Systems in Military Conflict
Categories: *Human and societal aspects of security and privacy; Computer systems organization -> Robotic autonomy; Social and professional topics -> Governmental regulations*
Corresponding Author: *Elvira Perez Vallejos*
Email: *Elvira.perez@nottingham.ac.uk*

Introduction

Weaponization of artificial intelligence (AI) presents one of the greatest ethical and technological challenges in the 21st century. A consortium of AI and robotics specialists have warned about the potential danger of using AI in war in an open letter at the 2015 International Joint Conference of Artificial Intelligence in which autonomous weapons have been described as the "third revolution in warfare, after the invention of gunpowder and nuclear weapons".[1] Many authors have highlighted the need for negotiating the trajectory of technological development on autonomous military robots, ideally in the early stages of development, among relevant social groups and actors including human-rights activists, researchers, developers, engineers, philosophers, policy-makers, military

[1] Future of Life Institute – http://futureoflife.org/open-letter-autonomous-weapons– Accessed 5/01/2017.

authorities, lawyers, journalists and the public.[2] [3] Despite the vital importance of this development for modern society, legal and ethical practices, and technological turning point, there is little systematic study of public opinion on this critical issue. This interdisciplinary project addresses this gap. Our objective is to analyse what factors determine public attitudes towards the use of fully autonomous weapons. To do this, we put the public at the centre of the policy debate, starting with youth engagement in political and decision-making processes.

On the one hand, the international community is concerned that instead of limiting conflict, using autonomous weapons in war will proliferate it.[4] On the other hand, defense departments and the technology sector point to many benefits of using autonomous weapons, which range from limiting military conflict to saving human lives.[5] [6] Instead of taking sides in the debate, our research will contextualize it by inviting young adults (16-17 years old) to become part of a youth jury. The aim of the youth juries is not simply to find out what young adults think and feel about fully autonomous weapons, but to discover what shapes their thinking; how they came to define certain scenarios as problematic; how they attempt to work together to think through solutions to these problems; the extent to which they are prepared to change their minds in response to discussion with peers or exposure to new information; and how they translate their ideas into practical policy recommendations.

This approach is inspired by the wave of deliberative experiments and initiatives that have been conducted in recent years on topics ranging from healthcare reform and nuclear power to local town plans and community policing. The theoretical assumption behind deliberation is that people are able to change their moral, political or behavioral preferences when they encounter compelling reasons and evidence to do so. When it works well, deliberation gives fluidity to democracy and reduces the narrow meanness that is so often associated with the sordid politics of 'winners' and 'losers'. It opens up a space for people to think about the future they want, and how they might act collectively in ways that take all actors into account.

[2] The social construction of technological systems: New directions in the sociology and history of technology. Wiebe Bijker, Thomas Hughes, and Trevor Pinch. MIT Press, 1987.
[3] Negotiating autonomy and responsibility in military robots. Merel Noorman and Deborah Johnson. Ethics Inf Technol, 16, 51-62, 2014.
[4] Killer robots. Robert Sparrow. Journal of Applied Philosophy, 24(1), 62-77, 2007.
[5] Governing Lethal Behaviour: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture, U.S. Army Research Office Technical Report GIT-GVU-07-11. Ronald Arkin. http://www.cc.gatech.edu./ai/robot-lab/online-publications/formalizationv35.pdf – Accessed 5/01/2017.
[6] Autonomous military robotics: Risks, ethics, and design, funded by US Department of Defense/Office of Naval Research. Patrick Lin, George Bekey, and Keith Abney. http://ethics.calpoly.edu/ONR_report.pdf - Accessed 5/01/2017.

While there is now a considerable research literature on the normative, epistemic and pragmatic value of public deliberation[7] [8] [9] [10] [11] , hardly any systematic research has been conducted on the ways in which young adults deliberate. Valuable observational studies have explored how young adults talk about political issues [12] [13] [14] [15] , but they have not addressed the deliberative questions of autonomy and responsibility in military robots. This is not only a gap in the literature but a missed opportunity to generate discussion and reflection as well as to learn about the ways in which practical reasoning occurs within the next generation of thinkers, often dismissed as lacking sufficient maturity to contribute to public policy. When examining young people's attitudes toward politics in Britain, research shows[16] that today's youth generation is deeply critical of political parties and professional politicians, however, they are interested in political affairs and feel that politicians could do more to connect with young people and listen to their concerns.

This paper will focus on the deliberation process and discourse around moral responsibility for autonomous robots with capacity for decisions that affect morally significant outcomes. Opinion formation is messy, often framed by competing and even inconsistent values. Supporting young adults to think through this messiness is a major aim of the youth jury process. The youth juries are structured with a view to encourage an atmosphere in which unconstrained deliberation can flourish. It is important for the juries to be noisy and discursive and that jurors become aware that they are engaged in a process of collective judgment, one that calls for both candour and compromise. From the outset, the idea of being a member of a jury is emphasized and participants know that they are expected not only to offer ideas about the ethical dilemmas intrinsic to fully autonomous weapons (e.g., responsibility and accountability), but to work as a group to think through a set of recommendations that adults in general, and policy-makers and the robotics/AI industry in particular, would feel compelled to take seriously.

The evidence presented to the jury was a combination of multimedia news (see Appendix) that showed case plausible -but fictitious- scenarios that triggered discussions and elicited reflective responses. The jury was asked to suspend their disbelief and immerse themselves

---

[7] Deliberative Democracy: Essays on Reason and Politics. James Bohman and William Rehg. MIT Press, 1998.

[8]Special issue: democracy in theory and practice. Stephen Elstub. Routledge, 2010.

[9] Deliberative Systems: Deliberative Democracy at the Large Scale. John Parkinson and Jane Mansbridge. Cambridge University Press, 2012.

[10] The Foundations of Deliberative Democracy: Empirical Research and Normative Implications. Jürg Steiner. Cambridge University Press, 2012.

[11] Deliberation and Democracy: Innovative Processes and Institutions. Stephen Coleman, Anna Przybylska and Yves Sintomer. Peter Lang, 2015.

[12] Uninterested Youth? Young People's Attitudes Towards Party Politics in Britain. Matt Henn, Mark Weinstein and Sarah Forrest. Political Studies, 53(3), 556-578, 2005.

[13] Hidden ethnography: Crossing emotional borders in qualitative accounts of young people's lives. Shane Blackman. Sociology 41(4): 699-716, 2007

[14] Family Talk, Peer Talk and Young People's Civic Orientation. Mats Ekstrom and Johan Ostman. European Journal of Communication, 28(3), 294-308, 2013.

[15] Facing an uncertain reception: young citizens and political interaction on Facebook. Kjerstin Thorson, Information, Communication & Society 17(2), 203-216, 2014.

[16] Young people and politics in Britain. Matt Henn and Nick Foard. Sociology Review, 23 (4), 18-22, 2014.

in a series of sketches of fictional scenarios (i.e. short news report videos and newspaper headlines) that initiated the process of deliberation. The jury considered both problems and future recommendations about the role of AI in military conflict. The scenarios featured two specific contexts; ISIS deploying fully autonomous drones which could choose their own targets (e.g. mobile anti-aircraft battery) within a predefined area to fight back allied forces at Aleppo (Syria). These drones operated without any human involvement, limiting the ability to abort any mission. If civilians were used as a human shield, the weapon simply ignored them and targeted anyway. This scenario highlighted the dangers of proliferation and quick replication of autonomous weapons. Unlike nuclear weapons, a piece of code for AI could be obtained on the black market and endlessly replicated at little cost, and the hardware for this type of weapon does not require costly or hard-to-obtain components and materials.

A second scenario featured the Ukrainian army using humanoid robots to fight pro-Russian separatist forces and the suspicion that the United States were supplying these robots. This scenario highlighted the lack of legal status for autonomous robots within armed conflict or internationally agreed laws of war, and also the worries about escalation of tensions between Russia and the U.S. Reports from the Ukrainian army focused on the efficacy, accuracy and highly effective overall results of this type of weapon alongside relieving suffering and distress among civilians and other non-combat Ukrainians.

These scenarios illustrated what happens when metaphorical claims about machine autonomy are taken literally. They triggered discussions and debate among jury members who were confronted with the possibility of fully autonomous robots equipped to make their own decisions, given their pre-defined goals, internal state, and sensory input. The jury facilitator (E.P.V.) introduced dilemmas and plausible risks including drones being uncontrollable in real-word environments, subject to design failure as well as hacking, spoofing and manipulation by adversaries. Jury members were confronted with questions such as; who is responsible if one of these drones doesn't function as planned? The developers of the guidance systems? The programmers? The person/entity that launches it? The manufacturer?

Methods

The youth jury methodology is fully described in Coleman, et. al[17] and Pérez et. al.[18] Ethical approval was granted by the Nottingham University Research Committee at the School of Computer Sciences. In total we ran two juries with 15 participants each. A total of 30 jurors (14 females) contributed to this report. Jury sessions were audio recorded and transcribed for thematic analysis independently by the authors. The audio transcripts were first read several times and then double coded for themes independently by one of the

[17] The Internet On Our Own Terms: How Children and Young People Deliberated about their Digital Rights. Stephen Coleman, Kruakae Pothong, Elvira Perez Vallejos, Ansgar Koene. (2017). Available at http://casma.wp.horizon.ac.uk/casma-projects/irights-youth-juries/the-internet-on-our-own-terms/. Accessed 01/02/2017

[18] Juries: Acting Out Digital Dilemmas to Promote Digital Reflections. Elvira Pérez, Ansgar Koene, Chris Carter, Ramona Statache, et al. ACM SIGCAS Computers and Society, 45(3), 84-90, 2016.

authors (E.P.V.) and also by an independent researcher with experience in qualitative analysis. Any disagreements in coding were addressed in discussion. Coding consisted of searching for sought themes and emergent themes in the transcript. We used NVivo v10 and Microsoft Office Word 2010.

Results

Firstly, it is important to note that in this paper we do not seek to correct the assertions made by the Jurors about the capabilities of robots, nor the existing or proposed regulations and laws relating to military robots.

The discourse around responsibility and autonomous robots was rich and complex. In general it was agreed that robots, as intelligent and autonomous agents, were the most promising emerging military technologies. The more intelligent they become the more useful and effective they could be, especially if they were able to save human lives. However, ethical qualities and an international agreement of what is 'good' and 'bad' were requirements that had to be embedded in the design of the robot to ensure robots do not turn out unscrupulous, destructive and a risk for humankind. The discussion around control emerged earlier in the conversation and it turned out to be a key challenge, identified in both juries. Jurors were concerned about losing control and perceived the robots as the 'real enemy'. While some jurors argued that we should never allow robots to be fully autonomous because of the inherited risks, others argued that robots could be more ethical than humans and could save human lives if designed according to the Laws of War, and the Hague and Geneva Conventions, among other regulations from military war guidelines. As one juror related:

> I found it [the scenario] kind of scary, because if they [military] make such intelligent technology, what if this intelligence increases and we become to lose control over them? Something really bad could happen… and how could we [humankind] stop them [robots]?
>
> Because the machines learn you cannot put a limit to what happens, they are going to be changing and adapting [and therefore be uncontrollable].

It became immediately clear that having ultimate control over the robot was a mandatory prerequisite if we were to legitimize research and development of autonomous weapons. According to another juror:

> I think any military organization has backups, so they are not just let this things happen without any safety measures, a kind of switch that could turn them all [the robots] off.
>
> Like a phone you can turn on and off. Machines need power and eventually they would break down. I think the real risks are very small.

This proposal resonates with the keeping humans 'In the Loop' or 'ethical governor' argument[19], in which an authorisation process requiring communication between the human controller or ethical governor and the robot is always required. However, as illustrated by the next argument, fully autonomous weapons could operate without human input and this was seen as a warning against placing too much trust in a robot that potentially could turn uncontrollable, even though some jurors were skeptical about this possibility:

> A: But that is nor fair, because you cannot stop what happens to the coding if it is a learning algorithm, no matter what happens it will learn something new and you cannot control what happens.
>
> B: Then you can just turn it [the robot] off.
>
> C: But what if it learns to switch itself back on? And they can stop us from turning them off?
>
> B: That is impossible! If they are switched off, how are they going to turn themselves back on? Technology is not that sophisticated!
>
> C: But that is what artificial intelligence is all about, they learn and they can turn themselves on. As soon as you turn them on they are going to learn something new.
>
> B: Once they are turned off we can take the weapons off
>
> D: What if they do not turn themselves off? And they shoot you!
>
> B: But the switch does not necessary has to be in the actual machine, it can be elsewhere, in a control room…
>
> E: Sometimes human controlling, instead of the actual machine, can be worse. You never know, someone in power could come in, be very negative, be a dictator like and take advantage, so it may be better if the robot has control of itself.
>
> B: You cannot replace [human] decision, you cannot replace a General with robots. There is always going to be human life involved in war even if it is robots vs. robots.

This last comment exposes the dilemmas of authority and trust in the context of human-autonomous weapon interaction. Interestingly, jurors did not comment on the possibility of human error or unreliability when working under pressure or controlling multiple units at one time due to cognitive overload. Jurors were more concerned about the personal

[19] Just say "no!" To lethal autonomous robotic weapons. William Fleischman. Journal of Information, Communication and Ethics in Society, 13(3-4), 299-313, 2015.

qualities of the controllers, their ethical stands, intentions and hidden political agendas. On the other hand, robots were perceived as lacking moral agency and therefore prone to act unethically, while some other jurors felt that robots could be fairer that humans.

However, the theme of control expressed within the concept 'what if something goes wrong?' keep emerging as a recurring concern:

> I would not say they can be more fair. At the end of the day robots do not have emotions or anything… so they are going to be more ruthless, they do not know when the right time to stop is, they are just going to carry on and carry on [killing people]…
>
> Even if you program the robot with a set of rules that describe what is good and what is bad, we go back to the question; what is something goes wrong? And they go back to be just what they are, just a robot [without moral agency].

Another topic that concerned our jurors was the issue of proliferation and accessibility. A solution around the concept of proportionality, meaning that each country could have a cap to control and limit the amount of autonomous weapons was presented:

> It says in the article [scenario #2] that it is relatively easy to weaponize a civilian drone. If you have people like ISIS that they are not in governmental power but that they get a lot of money from the oil they have been supplied, they can get hold of these things and is so easy to weaponize them for the wrong reasons, it can be extremely dangerous.
>
> Robots and weapons and drones, none of it is ethical, but the fact that they now exist and people know how to make them, we need to change the way we are thinking 'Ok people know how to make them, we need to start defending ourselves from it', so sometimes in order to defend we need to start using them but the ideal thing would be that each country has proportionate amounts of each weaponry, but this is not going to happen because people in each country develop in a different way. I do not think weapons are ethical but they exist and we have to have some control and it has to have equal distribution.
>
> A: If a country starts using these robots, other countries will do the same, and try to create a better ones.
>
> B: There is going to be competition
>
> C: All [the countries] trying to do something better… or capture those robots, or someone hacks them [as counterattack]
>
> D: They [the robots] are not going to be doing what you are telling them anymore.

The deliberation process moved to the recurring topic 'how to control the robots', in case they become fully autonomous and attack humankind or targeted countries under human orders. When jurors were prompted with the issue about who (or which institutional body) should regulate all the military industry responsible for creating new weapons, their deliberation process started by pointing out that different parties could have conflicting interests and therefore, an approach were responsibility was shared could be the most appropriate solution, for example, by appointing an international organization like the UN or an inclusive and varied group of stakeholders.

> A: That is a hard one because you can put it in anyone's hands, but not everyone has the right views on who should has it and who should not have certain weapons. Because [on the one hand] we could say 'we should not give weapons to this country' but that could be completely biased because should one country or one set of people [be] deciding? Who is one person to say who should and should not have weapons?
>
> B: Regulation should not just come from one person or a group of people because if you look like through history there is always has been a problem with a one sort of person to ruin everything. So it should be lots of people that has to regulate it and approve its use. Yes a lot of people, government has a say, the military has a say and the people that made it has a say, all make sure that they are not put in the hands of those that should not have them.
>
> C: I think the UN should be the ones to be involved because they are literally every country in the world and it means that every country will have a say in it. So it is not just America or Russia saying I want to do this and this and this. It has to go through every country in the world, so there is a big majority vote. It is going to be the last bias. It is the way the more people are happy about it.

Jurors were also skeptical about regulation compliance, and also about reaching an ethical consensus among different countries. These concerns were expressed with statements like:

> It does not really matter who has control. Whoever [country] makes the rules are going to make them for their own advantage.
>
> There are people like ISIS that you cannot regulate anyway. The rules will be broken and people that should not get hold of this weapons will get them.
>
> There is no point in bringing ethics into this. The two parties fighting are going to have different ethics, different perspectives.

In one of the juries, young people reflected upon a hypothetical situation in which robots fought against robots. This scenario puzzled many jurors, especially female jurors. One feature of deliberation is the generation of empathetic reasoning. As the deliberation

process went on, the initial confusion began to get resolved. The more the subject was discussed, the more likely were jurors to relate to a situation in which others were suffering by stressing the importance of saving human lives:

> A: I may not get it, but what is the point of robots fighting against robots?
>
> B: Territory
>
> A: It can just go on forever if you keep replacing robots nobody is going to lose.
>
> C: But fighting people with robots is just worse and unfair.
>
> D: But these robots are robots with autonomous coding so they are constantly learning and adapting so even the people that invent them will be changing them as well, so no one will be fighting the same exact robot, they would be different types of robots doing different things.
>
> A: Yes, but what is the point of robots fighting robots?
>
> B: For territory, for power… it is exact the same reason [than humans against humans], it is just taking humans out of the equation and replacing them with robots
>
> A: But humans die and some of the countries are actually affected, but if a robot fights another robot they can just be replaced.
>
> E: So do you think that the ideology of human not dying is a bad idea?
>
> A: No, but it is pointless. We do not see the point of fighting robots against robots.
>
> B: A country's economy is not unlimited, at one point it will run out of money and it will not be able to replace any more robots.
>
> E: If there is any way we could reduce the death of humans by replacing them with robots, even if it is a constant battle of robots with robots, I think it would be more beneficial because we are saving human lives.

Jurors also deliberated about the fairness of war:

> A: it should be a way to make war fair
>
> B: But why should war be fair? It is war and you go there to win.

C: Someone in power with lots of money could buy the biggest army in the world and there is no real way to stop them, they could just buy and buy and buy [robots] and because robots have no morality, they could not decide if what they are doing is fair or not, they just follow your orders, so they could do anything.

D: There are check lists that you have to complete before you go to war to ensure is fair.

B: But nobody would go to war if it is equal, you go to war because it is not equal, and because they know they will win. With 50-50 chance nobody is going to risk that.

Jurors were also concerned about the vulnerabilities of any software and possible technological problems. A juror pointed out that some systems could be impossible to hack, however, jurors arrived at the consensus that any software was vulnerable to hacking or being 'switched off':

A: The robot is just a code, very easy for the people who bought the robot to change the code and simply ignore these conventions of war [e.g., Geneva Convention], because once you have bought them, they are yours and you can change anything you like, literally deleting a few lines in the code and then you can do anything you like with that robot basically.

B: But that cannot happen without the manufactures permission [pointing to a legal problem] because the code is made even before the robot itself is made.

C: They can do it without their permission

B: So why do they get hack all the time?

D: But don't you think there is firewall and firewall after firewall to intense amount of complex code before you can hack something. If the USA army decided to start making them the amount of code would be un-hackable. Not to anybody unless Russia or some technological advance country just blindly would put lots of billions to try to hack it, and even then, they can just change the code and sort that issue [the hacking] in a couple of seconds.

E: But then every country would try to by un-hackable and nobody is going to allow their country be hacked.

F: There is always faulty software. So there is always a kinking software, a little way where someone can get through the files… at least it is an army-based technology that's got lines and lines of encryptions and it needs billions of pounds to get into it, so most companies always got a little kinking software you can get into it.

J: Also going back to what you said about being hacked, that is like… Commons cannot be hacked, that is the reason we have secret services, MI5, FBI and CIA if we did not need to gain information we were not allowed to access.

A: Could not anyone could turn them [the robots] off? If you are using robots, if that is the only thing [weapon] that you are using… Couldn't the opposition turned them off? And then you have nothing… that does not work. Essentially, anyone could just turn them off, it is not complicated.

When prompted with the first scenario (i.e., Ukrainian army is using robots to fight separatists forces in Donbas) jurors again expressed concerns about the possibility of 'what if something goes wrong', a technological error and its fatal consequences and the need to destruct those lethal uncontrollable autonomous weapons. A juror suggested a way to control these robots and the need to apply the utilitarian concept of 'the Greatest Good for the Greatest Number':

A: But there are weapons against electronics too. I do not know if someone has heard about electromagnetic pulse, it just shuts off any single electronic, such as wi-fi, any connection… you could just stop them [the robots] with a simple thing like that.

B: But it cannot affect all the electronics and devices around us…

A: It would be only in a certain radius, like this room.

C: But if the robots are working in a city and you have to stop these robots and you apply this electronic thing… what is it called? Electromagnetic pulse, you will affect all the city, including traffic lights…

A: If these robots are killing people, looking for the opposite army, you will have to use it [the electromagnetic pulses] anyway. Humans can think for themselves on the spot. A robot cannot suddenly… so if you shut all the electronics, humans can realize what is going on, a robot will not be able to realize if it has been turned off and realistically, you have to think in the 'Greater Good', so yeah more people may die if the traffic lights gets turned off, but if those robots were going to take up the whole city, which one would you chose?

C: That is why you should not put these robots there in the first place…

It looked for a while as if the discussion might have been heading for an impasse, with two incompatible moral positions in conflict with one another about who should have the

absolute control over autonomous weapons. In searching for a way to define the problem, the juries managed to sum up the discussion that lead to the two conflicting arguments:

> If I could talk to the regulators I would say 'do not use them'. The benefits do not out-way the possible disadvantages of it going wrong. Even phones they thought they were fine and release them and they started blowing up, so if cannot get a phone right, why should [they] get it with technology to kill people. It is not worth it.
>
> I think you cannot say 'do not use them at all' because it is a very logical thing replacing a robot that is very replaceable with a human because once that human gets killed… a robot you can replace it but that person's life is over. I think it is quite logical using them in replacement for humans. Sometimes [it] is justifiable, yeah.

When it came to the question of who is responsible for the behaviour of autonomous robots, especially fully autonomous robots, a variety of positions were articulated and no differences between scenarios were observed.

In line with arguments supported by scholars like Matthias[20] and Sparrow[21], some participants argued that it would not be possible to hold humans responsible for the behaviour of autonomous robots, especially when deep learning is guiding decision-making resulting in unwanted tendencies. Even though jurors tended to resist the fictitious possibility of fully autonomous robots, once they suspended their disbelief, many jurors reflected on the epistemological nature of being a robot. Interestingly, the suspension of disbelieve was soon boycotted by one of the juror who reminded others that such technology 'does not exist' yet and also that currently it would be illegal to develop it. It seems jurors felt deeply uncomfortable when believing such robots were 'real'. Jurors concerns about losing control over the robots was illustrated with expressions of fear about the impossibility of regulation and the dangers of terrorism:

> C: But what if a robot can choose? A really sophisticated and intelligent robot. But then is that even a robot? If it can choose?
>
> D: It would be a robot? Because there are lots of robots that can make choices, but I think that when they start having emotions that is when that barrier starts getting broken because it is extremely difficult to replicate that.
>
> C: What do you class as a robot? If the robot thinks… but if it thinks it is not a robot anymore. It cannot think by itself and be a machine.

[20] The responsibility gap: Ascribing responsibility for the actions of learning automata. Andreas Matthias. Ethics and Information Technology, 6(3), 175-183, 2004.
[21] Killer robots. Robert Sparrow. Journal of Applied Philosophy, 24(1), 62-77, 2007.

> D: If it has it owns thought processes and emotions then it is not a robot. It is not a machine any more.
>
> A: But there is no such thing as that. That does not exists. There is not yet a 100% AI military offensive weapon, it is not easy. It goes against the Geneva Convention, so they cannot legally. If one country does it and the rest finds out then all the other countries will get on their backs.
>
> E: But people break the law all the time.
>
> F: Terrorists. People that are ready to kill people, I do not think they care about law.

Only one juror entertained the idea that autonomous robots might someday be held responsible in some narrow sense for their own behaviour[22] [23] [24] [25], especially when those robots are capable of performing acts involving life and death with some kind of moral framework previously embedded in the design ethics based control systems of the technology. This type of reasoning transfers human capabilities to the robot:

> The [fully autonomous] robot would hold responsibility, but you cannot exactly take a robot to court.

Some argued, similarly to Crnkovic & Çürüklü[26], that responsibility should be shared between highly sophisticated robots and the human actors involved, especially when robots have been designed with ethical capabilities. This position allows functional responsibilities within a network of distributed responsibilities in a socio-technological system:

> If the robot has been programmed to do something and suddenly start making mistakes, why is it making mistakes? Is it because it has been programmed badly or is it because has a mind on its own? […] you can then blame the robot or the person behind.

Most participants argued that only humans should be considered to be capable of moral agency and argued that humans should always be responsible for the behaviour of robots.[27]

---

[22] Robots and responsibility from a legal perspective. Peter Asaro. Proceedings of the IEEE Conference on Robotics and Automation, Workshop on Roboethics, Rome, 2007.
[23] On the moral responsibility of military robots. Thomas Hellstrom. Ethics and Information Technology, 15(2), 99-107, 2013.
[24] Terminating the terminator: What to do about autonomous weapons. Wendell Wallach. Institute for Ethics and Emerging Technologies - http://ieet.org/index.php/IEET/more/wallach20130129 - Accessed 6/01/2017.
[25] Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). Mady Delvaux, Committee on Legal Affairs, European Parliament, 2016.
[26] Robots- Ethical by design. Gordana Dodig Crnkovic and Baran Çürüklü. Ethics and Information Technology, 14(1), 61-71, 2012.
[27] A legal theory for autonomous artificial agents. Samir Chopra and Laurence F. White. University of Michigan Press, 2011.

[28] [29] Participants that took this position argued that there was nothing new about autonomous robots in the sense that the legal and moral concepts currently applied to other complex technologies such as medical equipment or autonomous cars. Ultimately, the engineers were seen as mainly responsible for building machines that are potentially dangerous for the society:

> We cannot compare humans and machines. Machines have no needs or desires, they do what we [humans] tell them to do.
>
> The person that made the robot should be responsible if something goes wrong. The person that designed that particular piece, the code that made the robot go wrong.
>
> The people that invent them should not shift the blame.
>
> The coder are the responsible, the ones that created [the robot] in the first place.
>
> The governments and the people that deploy them should be held responsible.
>
> You cannot blame a robot because the robot is doing a job.
>
> A robot is doing what it has been told to do. It has no choice really. It cannot opt out on doing it. It is the person's whose code it decision.

In general jurors were apprehensive about the idea of co-existing with this type of advance lethal technology and kept highlighting the possibility of robots endangering the safety of humankind:

> Once the robots win against the other [enemy] robots they are going to come after the humans. I would not use robots because they can go wrong.
>
> They will go after the humans, you cannot stop that.
>
> The robots may be able to recognize each other. If they are programmed and aim to take the power of the country they will not stop until they get the power of the country, which will involve in getting rid of the humans, so it is always end up in humans getting kill.
>
> It is just technically robots against humans.
>
> If machines gain consciousness they will erase the human race.

[28] Learning robots and human responsibility. Dante Marino and Guglielmo Tamburrini. International Review of Information Ethics, 6, 46-51, 2006.
[29] Ethical regulations on robotics in Europe. Michael Nagenbour, Rafael Capurro, Jutta Weber, and Christoph Pingel. AI and Society, 22, 349-366, 2008.

Discussion

Our jurors deliberated about the possibility deploying fully autonomous robotic weapons. Their verdict was clear; even if engineers are able to create the perfect war robot, able to follow all the articles of the Geneva Conventions, the laws of war, act more morally than the human soldier, not suffer the psychological and emotional stress that human combatants suffer, and be constrained to act 'ethically in war', the risks outweigh the benefits. Losing control over autonomous weapons systems was the major reason not to participate in the development of such technology.

It is significant to note that a recurring theme in the jury sessions is the idea that robots would get 'out of control', and in some way seek to destroy mankind. This notion is not presented in either scenario, nor prompted by the investigators. This is strong evidence that young peoples' models of robots are based on cultural folk ideas, and more specifically, that these pre-existing ideas played a significant part in their search for moral consensus. Throughout Western history we have created such stories, from the Greek Pygmalion myth, via Shelley's Frankenstein to the recent film Ex Machina. This repeated narrative of our creations turning on us with this intention to either enslave or destroy humanity fuels our mistrust of AI and autonomous robotics[30].

It is the authors view that these existential fears can cloud our judgement. They effectively prevent us from clearly recognizing the significant impact that robot autonomy based on the current capabilities of machine intelligence and robotic technologies may have on human culture. They provide an opportunity for those wishing to promote military robots to simply focus on explanations of how the robots cannot become conscious in the folk sense, or cannot operate beyond their pre-defined goals or objectives, as a means to justify their safety, efficacy and moral neutrality in warfare. We must draw attention to this sleight-of-hand approach. It is essential to focus on what exists, and what its dangers might be for human culture in the near future, rather than spend our resources hypothesizing about future imponderables.

The consensus that the Engineers should be held responsible for building machines that could be dangerous for society, shows the lack of understanding that in fact this responsibility is shared between technology professionals, the corporations for which they work, and governments who define the laws and regulations under which corporations operate.

Some exceptions were made, however, when contemplating the possibility of some robots being stronger, faster and smarter than humans with the potential to save more lives than actual humans.

The fact that there was no firm moral consensus about the dilemmas presented within these fictitious scenarios among the juror members, suggests that the scientific community must rationalize a set of norms and then 'give them teeth' through regulation and law, so that

[30] In Our Own Image. George Zarkadakis. Random House, 2015.

they become widely accepted societal norms over time. Scientists have an increasing responsibility to set evidence based on acceptable norms, and there is strong precedent here – climate change, smoking, drug misuse would all be good examples. We already have precedents in law today where a human must take responsibility for the consequences of actions of their subordinates, even when those subordinates are acting autonomously within a broader set of goals. Corporate manslaughter is just one example from company law, and there are precedents from martial law as well, from the Nuremberg Trials[31] to the Abu Ghraib scandal[32] where the defence of 'just following orders' were deemed insufficient. It is currently an open question about how these societal norms should be applied or adapted to deal with autonomous robots, rather than autonomous humans (soldiers).

Conclusions and Further Work

Our jurors primarily felt that the risks of autonomous robots in warfare outweigh the benefits, although the primary risk identified was existential – robots getting out of control and destroying humanity. There was some recognition that as robot performance exceeds human capability robots have the potential to save lives. There was however, no firm moral consensus about the dilemmas presented in the scenarios.

In searching for these societal norm, after working with young adults, we will expand our project to include a wider demographic sample including veterans, non-veterans, and active military staff. We will then begin to examine how the use of robotics in war is changing public perceptions of military conflict.

This study is unique because young adults are often undermined and excluded from public debate and the development of societal norms. The value of this research lies simultaneously in its contribution to the emerging field of fully autonomous weapons and in generating recommendations that can influence government policy-makers, industry chiefs, and public discourse. This study is vital for a critical understanding of young adult's perceptions of AI in armed conflicts and its implications for the future policy and industry decisions.

We aim to provide industry stakeholders with a roadmap of factors that determine public opinion about autonomous weapons and help frame their research and position their products. Finally, our research will inform the general public as well as bringing young adult's opinion into the debate about AI and military conflict.

The study is being funded with £5K by The Digital Economy Crucible 2016, an EPSRC funded leadership programmed, organized by Cherish-DE at Swansea University.

[31] Interrogations: The Nazi Elite in Allied Hands, 1945. Richard Overy. Viking, 2001.
[32] The Trials of Abu Ghraib: An Expert Witness Account of Shame and Honor. Stjepan Mestrovic. Paradigm Publishers, 2007.

Appendix 1

Scenario 1: ISIS AUTONOMOUS DRONES FIGHT BACK ALLIED FORCES AT ALEPPO

www.theguardian.uk.co Monday, 9 January 2016

ALLEPO - In an unprecedented move, ISIS began to deploy autonomous drones to help them restore the broken front in North Western Syria. As ISIS has been pushed back around the beleaguered city of Aleppo in the past month, its website has revealed that it will be using civilian drones equipped with guns and bombs against "enemies of Islam". ISIS has confirmed that the drones are operating autonomously without any human involvement. The drones have been seen to attack both civilian and military targets, work in groups that exhibited highly intelligent behavior never seen before, which include performing complicated tactical maneuvers, resupplying themselves, and strategically selecting targets for attack.

As Artificial Intelligence and Robotics expert, Prof Rob Wortham, from the University of Bath, explains "it's relatively easy to weaponize a civilian drone that can be purchased off the shelf, or on line, relatively cheaply." The bigger question surrounding the attack is where ISIS obtained access to software allowing this consumer-orientated technology to achieve such high levels of autonomy.

While it is not the first time that ISIS has used drones, the battle east of Aleppo presents the first evidence of autonomous drones using sophisticated artificial intelligence in war.
A consortium of AI and robotics specialists have warned about the potential danger of using artificial intelligence in war in an open letter in 2015. Announced at the International Joint Conference on Artificial Intelligence, July 2015, the open letter warns that "autonomous weapons have been described as the third revolution in warfare, after gunpowder and nuclear arms."

One of the dangers with autonomous weapons, warn researchers, is the specter of a proliferation. Unlike nuclear weapons, a piece of code for AI can be endlessly replicated at little cost, and the hardware for autonomous weapons does not require costly or hard to obtain components and materials. AI software can be bought on the black market, which is where ISIS most likely obtained the software that powers their drones.

The big ethical question facing governments around the world now is how to prevent, contain, and combat proliferation of this new technology.

There are currently no laws in Syria, or at the international level, to codify the use of autonomous weapons.

As the special committee of the UN Security Council gathers this morning for an emergency meeting, it must address several important ethical questions: How do we hold humans accountable for the actions of autonomous robot systems? How is justice served when the killer is essentially a computer? Should a ban on the use of autonomous weapons be enacted, and if so, how could it be enforced? If ISIS is using such weapons, should the

West supply the opposing forces with similar technology, potentially increasing proliferation of this new military technology?

Both military and technology observers warn that the replacement of human soldiers with machines could "start a global AI arms race". Governments and non-state actors may well aim to get the upper hand to maintain a strategic AI advantage.

Scenario 2: UKRAINIAN ARMY IS USING ROBOTS TO FIGHT SEPARATIST FORCES IN DONBASS
www.nytimes.com Monday, 9 January 2016

KIEV – Residents in the southern Ukrainian town of Marinka, near the city of Donetsk, have reported sightings of humanoid robots engaging in fire fights with the Russian-backed insurgents of the Donetsk People's Republic. It is the first such documented case in the history of war and robotics.

The conflict in Ukraine has been raging since 2014 when anti-government protests toppled the pro-Russian government in Ukraine. In response, Russia annexed the Crimea and supported separatist forces in the Donetsk and Luhansk regions of Southern Ukraine. Since February 2015, a cease-fire has been agreed on by the leaders of Ukraine, Russia, and the EU. But in recent months there has been heavy fighting, despite a ceasefire agreement, as the separatists continue to advance. The Ukrainian army responded to the violations of cease-fire by the separatist forces by deploying over 200 autonomous weapon systems alongside Ukrainian army soldiers.

Russian President, Vladimir Putin, has accused the Ukrainian Prime Minister, Petro Poroshenko, of using war robots, or autonomous weapon systems, which currently have no legal status in armed conflict. Mr. Putin has also accused the United States of supplying the weapons.

In a statement to associated press, the Minister of Defence of Ukraine, Stepan Poltorak, confirmed the limited use of autonomous weapons systems, called Auxiliary Robotic Units (ARU) but refused to confirm that these weapons were supplied by the United States.

They currently have no legal status and this is one of the grey areas in the internationally agreed laws of war. Poltorak added that "the machines are equipped with state-of-the-art Artificial Intelligence making them truly autonomous…and not requiring a human operator to control them remotely." He said the Ukrainian army has been secretly working on the development of autonomous weapons since 2002.

According to one local witness, Maria Kuliakova, the ARUs "kicked out the rebels from Marinka, helped to evacuate civilians, provided medical assistance, and delivered supplies." It was something that the Ukrainian army was unable to do for almost a year, she added.

According to the Ukrainian Ministry of Defence, the operation was a success and the Ukrainian army sustained virtually no casualties, while the separatists suffered over 40 people dead or wounded, with another 100 taken prisoner. Andriy Bohatenko, a private in the Ukrainian army who participated in the offensive, said that an ARU "saved my life…they are faster than the rest of us, they also do not get hungry or get tired…we can sleep at night now, knowing that they are watching out for us." The surprising offensive represents a major reversal of the Donetsk People's Republic and its Russian ally. The Ukrainian army is now poised to retake Donetsk, the stronghold at the heart of the rebel-held territory.

While the Ukrainian forces have suffered virtually no casualties in this new, surprising offensive, international observers worry about escalation of tensions. The fear is that Moscow may start sending more military equipment and even troops to Ukraine to help the separatists. But experts say Russian forces may suffer a fate similar to the separatists and it is not clear what Moscow can do in this situation.