# SafeGroud Presents the series ‘Stay in Command’ :Lizzie Silver on the Tech Perspective

## Sep 2020

Welcome to SafeGround, the small organisation with big ideas working in disarmament, human security, climate change and refugees. I’m Matilda Byrne.

Thank you for tuning in to our series Stay in Command where we talk about lethal autonomous weapons, the Australian context and why we mustn’t delegate decision making from humans to machines.

This episode we’re looking at the “Tech Perspective”. We are going to discuss the technological concerns of lethal autonomous weapons and their implications on the tech industry.

And so with me today I have a great guest with me today in Dr Lizzie Silver. Lizzie is a Senior Data Scientist at Silverpond which is an AI company based in Melbourne, which is also where I am coming to you from - so welcome Lizzie, thanks so much for joining us today

**Lizzie Silver**[00:00:52] Thanks for having me

**Matilda Byrne:** Before we jump in, I’m just going to talk a bit about the definition of killer robots in case any of our listers are unfamiliar with exactly what it is we’re talking about.

So killer robots or fully autonomous weapons are weapons that have no human control over the decision making. So when they select a target and engage the target so decide to deploy lethal force on that target, there is not a human involved in that process and it is just based on AI and algorithms. So with these fully autonomous weapons there are lots of concerns that span a whole of areas that span a number of different areas - today we are going to go into technological concerns in particular because we have Lizzie and her expertise, but there's also things like moral, ethical, legal global security - a whole host of concerns really.

## What is the most concerning thing about killer robots?[00:01:49]

**Matilda Byrne:** And what I’m interested in Lizzie, is, just to start off with if you could tell us what is it about fully autonomous weapons that you find the most worrying, so what about them makes you driven to oppose their development.

**Lizzie Silver:** It’s really a fundamental issue with these issues is you can't give a guarantee on how they’re going to behave. WIth humans we can’t give a guarantee on how they're going to behave but that’s why we have all these mechanisms for holding a human accountable. Now you can’t hold an algorithm accountable in any meaningful way. So what you would like to do is find a way to characterise how it’s going to behave in every situation, but the thing is a conflict situation is just too complex. There are too many potential inputs and outputs, different scenarios that could confront the AI. You’re never going to get through all of them. You’re never going to be able to fully characterise the space. So what

you'd like to say is say 'Ok, on this sample of situations we've seen this kind of behaviour, so we expect that in the future, in similar situations we expect the system is going to behave similarly'. The trouble with a conflict situation is that things don't stay similar- it's an adversarial context so we can't give any guarantee that the system is going to perform the way we expect and want it to, and they're changing by design, you know the enemy that you're fighting is trying to change things in ways that will through you up. It's just a fundamental characteristic of warfare.

[03:30] **Matilda Byrne:** For sure and I think you know, warfare is one of the most unpredictable and chaotic environments that we could possibly think of to put any kind of machine into which is really troubling.

## Could you speak to 'explainability' as one of the characteristics of AI raised by the international community?[00:03:49]

**Matilda Byrne:** and I think you also hit on some of the characteristics of AI that have been flagged by the international community as concerns with these weapons, so things like unpredictability, reliability, obviously the huge concerns around accountability and another one that has come up is explainability - I wonder if you could speak to that at all and sort of the black box phenomena

**Lizzie Silver:** Yeah, I mean explainability is really what we are trying to use to get some sense of comfort with these algorithms. We want some way to characterise their behaviour. But there isn't single kind of explainability, and something may be explainable and turn out to be unpredictable in ways we wouldn't expect. We keep coming up with new kinds of explainability, new explanations for how these complex systems work. Every time, we think to test them out on new benchmarks - an example of that is a paper that came out fairly recently on texture versus shape, that fund that AI systems really pay attention to the texture of objects and not the shape of them because there's more information in texture and it's easier for them to use and if they can solve the problem just using texture they will. And on major image benchmarks you can actually do really well just using texture, you're not forced to use shape information. And the way the research has demonstrated this is by creating a new dataset where the objects had conflicting shapes and textures. So you might have a picture of a cat and overlay it with elephant skin and a human looks at that and says it's a cat, bt a human looks like that and says it's an elephant - so it's behaving in a way thats different to how humans behave. And it's been trained to do this because paying attention to textures has been enough to recognise elephants in the past, it never needed to learn the shape of an elephant. But you think about how that behaviour is going to play out in a warfare situation where the enemy is inventing new kinds of camouflage and you're not necessarily going to be able to predict how the system is going to behave when the enemy changes its outfits and looks like civilians. And particularly when if you're retraining the algorithm to improve its performance constantly, you could have a system that avoids civilians up to a point, and then learns to target them.

(06:30) **Matilda Byrne:** Yeah, for sure. I think it's really interesting how really simple things can sort of, confuse the Ai. SO some of the proponents of fully autonomous weapons have stated that they will be

able to make sure it can identify civilian objects, things like a red cross or red crescent symbol in a conflict zone so it wouldn't attack those things. And then that for me was quite puzzling because it would be so easy for an enemy actor or non state actor to then put a red cross symbol on the side of their vehicle and then not be attacked, so I think these are really complex and nuanced things.

## Facial recognition, bias and distinction[00:07:11]

**Matilda Byrne:** You were talking about texture and shape, as an example of something that AI is trained to recognise. WHat about facial recognition and where we are in terms of that because obviously in terms of targeting specific people in  a conflict scenario that's going to be a big part of sensory input, we can assume.

**Lizzie Silver**: Yeah, so it turns out that the AI reflects the input data so AIs that are developed in mostly white countries tend to be a lot better at recognising white faces, AIs that are developed in China are much better at recognising Asian faces. And they also reflect the biases of the people who develop them, it seems likely that if we develop these they're not going to be as good necessarily at recognising the people that we are in conflict with - they are likely to look different to us on average.

**Matilda Byrne:**: For sure, and I think that algorithmic bias is something that has been flagged as really problematic in terms of how it impacts certain races more than others and disproportionately impacting certain people from parts of the world which is also a great concern.

**Lizzie Silver**: Yeah, and the fact that it performed worse on women than on men, particularly black women more than black men indicates that there's a bias based on the date set potentially the creators haven't focused so much on performance in that sub set and that you know, can't be because of population difference - there's just as many women as men, slightly more actually.

**Matilda Byrne:** That's really interesting, and also that it's perhaps biases in the people that are training in the first place that are perhaps more inclined to then train it on more men, why is that?, and then those become embedded in what the software is able to do, and then I think when you extend that to a warfare situation and think of those biases that are being embedded by a state that is waging war, and is in conflict what it's going to reflect is just going to exacerbate, potentially the harms that occur in conflict.

**Lizzie Silver**[9:31]: Right, None of these are intended issues with the AI. They're all just emergent properties based on how they're trained and the data sets that they're trained on, and the way people put them together without specifically setting out to mitigate these problems. Now there's a whole lot of unintended situations in warfare, there's a lot of unknown unknowns and if we can't even get right equal performance across races and genders in a situation where things are not changing very rapidly, it seems really overly optimistic to me to think that we can distinguish civilians and non-combatants from combattants. It seems incredibly optimistic to me to say that the AI is going to be able to make some kind of ethical judgement in the moment and say 'this person was attacking v this person was surrendering', those actions can look really similar'.

(10:37) **Matilda Byrne:** For sure, and I think there's a lot of importance of having that human evaluation and what youre talking about with this distinction between enemy combatants [with civilians] is getting into the territory of international humanitarian law. And I think there's a important part to that decision making that a human can read certain signals and cues and evaluate them when theres something thats not necessarily concrete, they can understand the wider context and use their human judgement really which is what is essential, on top of this other information to then make a decision and this is why when we talk about meaningful human control or human involvement it's so important that we do have that added layer, I guess, and it's ot purely the algorithms decision making based on what it can attempt to understand and compute.

## What benefits, from a technical point of view, might these weapons have for militaries?[11:36]

**Matilda Byrne:** I think it's probably worth also acknowledging why these weapons are desirable for the military because they are under development in a handful of countries; so the US, Russia Israel, UK and also Australia [as well as China and South Korea] and so they wouldn't be trying to develop them if they didn't think there were benefits. Some of them are obviously around I guess strategic things so not having to have as many boots on the ground is obviously desirable for militaries and there's other ethical questions around whether that's good or not, but what about from a technical point of view, why do you see these weapons as potentially being appealing?

**Lizzie Silver:** So there are some situations where I can really understand the appeal. Firstly, machines can react a lot faster than humans so there are already automated systems like C-RAM and FALENX that are intended to react to incoming missiles that are coming in faster than a human could overt them. And those are supposed to defensive systems, but they do sometimes, screw up, you know there have been cases where they've shot down planes  Now with the i you have a similar situation, you ma want the AI to take control of a fighter jet to perform evasive maneuvers faster than a human can but then when you are performing aggressive maneuvers faster than a human can then you have a problem right? Because the AI then needs to make all of these complex ethical judgements. The other situation in which you might want it, is if you have developed these systems where you have maintained human control, he AI is maybe helping to plot a drone but it is not deciding who to strike with the drone and then you go into an area where your communications are cut off with that drone, so the idea here is that you should be able to keep fighting in the same capacity even though communications are cut off. This I'm less sympathetic to because I don't see the need to continue to use lethal force if you can't control the weapon, right. You can always just have that drone turn into a reconnaissance drone while it is cut off from you. Those are some of the situations.

## On lowering the threshold of war

And then there's an argument I'm really unsympathetic to that I've heard from some military people - that they would like to just not expose their people to the sole crushing situation where they have to kill another human being, and while i do understand that that does psychologically damage people, i think

removing us from that damage just makes it so much easier to get into conflict situations, so much easier to create civilian casualties and i don't think we should be in a position to remove ourselves from the harm that we are doing.

**Matilda Byrne:**[14:40] For sure I think that idea of lowering the threshold of war because now our own people won't be exposed to this, doesnt mean theres not other people and civilians that are now being exposed, and we need that barrier to war in the fist place so that we don't escalate unnecessarily or are over zealous in starting wars. I think its really important.

## On delegating the kill decision to a machine

[15:01]**Matilda Byrne:** And I think the other moral component when we're talking about machines and this decision making, so this idea that you mentioned that if they lose communications should they go on and just continue. And especially around deploying lethal force, this idea of handing over a kill decision to a machine and so are we ok as society, as humanity, to allow machines, these algorithms to make decisions over life and death, I think alone is very problematic and a precedent that is troubling to set.

[15:32] **Lizzie Silver:** Yeah, I would not be comfortable trusting my life to a computer vision system that was supposed to recognise me as a non-combatant. I mean I work with these systems everyday and they fail in the most surprising ways. And it is just too complex and difficult a situation: warfare. Adversarial situations are ones where you can't really give guarantees on performance, every guarantee that we have in statistics is based on some assumption that the data stays similar to a certain degree. Even transfer learning guarantees where they talk about "we are transferring to a new situation, the data is fundamentally going to change" but t guarantee that the system is still going to work we have to assume that something is still going to stay the same, and I don't know what you can assume stays the same in war, except basic physics.

## What about hacking?

**Matilda Byrne:**[16:31] Sure, and I think also in terms of guarantees in these systems is safeguarding them from hacking and that being a huge risk.

**Lizzie Silver:** [16:39] Yeah, absolutely. That is a risk in all military systems currently, but it gets a lot worse when you can take something that's already autonomous. I mean, it's a risk with remote piloted systems too. Right? Cause if you can hack into them, then you can create havoc, right? You don't need to be present to turn them on the side that created them, similarly with autonomous weapons.

## What are some positive applications of AI?[00:17:10

**Matilda Byrne:** Mm, I think, we've talked a lot about, I guess, some of the limits of AI and how they're particularly problematic in the application of weaponizing AI, having them involved in fully autonomous weapons, but it's really important not to forget the positive applications of AI and even in certain

contexts within defense, like you mentioned earlier, evasive maneuvers in pilots, there's some really great stuff happening in mine clearance with using robotics and AI. But even more broadly than that, I think in society in general, it's really important to step back and look at the good that it's bringing. So I wondered if you could speak to maybe some of the other domains where AI is being used in a really positive way for social good.

**Lizzie Silver:** [00:17:53] Yeah. There's a ton of stuff out there. I mean, one of the things that my company has developed is a decision support tool for clinicians, for ophthalmologists. It recognizes fluid in the retina, which is one of the things that you get with macular degeneration, which causes blindness. So, the AI picks up these pockets of fluid, some so small that they might be missed by a clinician and helps to quantify how bad that is. And that's another system that's under human control. Right? It's just making recommendations to the clinician who then makes the actual decisions about treatment. There's loads of uses in industry as well. We're working with power companies to [00:02:00] help take inventory of their assets across the power network, which can help with making sure everything's up to code for the bushfire risks.

## Funding for AI in defence versus social-good applications[00:18:48]

**Lizzie Silver:**A lot can be done with these tools. And I'd like to see more investment in peaceful uses of AI. What was the investment that you said the Australian government, uh, put into. Developing an ethical military AI?

**Matilda Byrne:** [00:19:02] Yeah so just the one project that was about researching how to embed ethics into fully autonomous weapons specifically that the Australian government funded was worth roughly $9 million.And that's part of a huge amount of much, much, much larger funding that goes into things like the research centers that innovate on autonomous systems for the military. So trusted autonomous systems when it was launched, received $50 million. And just [00:03:00] earlier this year, the Royal Australian Air Force allotted $40 million to Boeing to innovate a new autonomous combat aircraft also. So these huge scale amounts of money that we're talking about.

**Lizzie Silver:** [00:19:51] Interesting. I was recently part of a medical research future fund grant application, unsuccessful alas, that was for a pool of funding that I think was seven and a half million dollars to be split across several small projects. And that is smaller than the grant that was made just for an ethical military AI. I think there's so many applications of AI in medicine. Why are we the funding this pie in the sky idea of an AI that's going to make ethical judgments. Why not put that into medical AI development that's what we could really use?

**Matilda Byrne:** For sure. I think divesting of even just a fraction of funding, it's amazing to think about what could be achieved.

## On the challenge of dual-use [00:20:45]

One big challenge though. I think perhaps when you have both of these things happening simultaneously, so these developments for social good, as well as autonomous capabilities within the military is this idea of dual use and how different components and there's this complexity where it can easily be repurposed. If you could speak a little bit to that, I'd be really interested to hear your thoughts on how this is a problem and how it could be mitigated.

**Lizzie Silver:** [00:21:10] Yeah, so it's difficult because everything that goes into developing that aircraft that can autonomously strike enemy combatants also goes into an aircraft that autonomously does search and rescue operations, right? It's still got to navigate, it's still got to find a target. The only difference is the deployment of force. So everything that you build can be used for multiple purposes and that has some consequences. You know, some people will not work on computer vision at all because they're, so you're worried about the problems with facial recognition. So we're losing talent because there are people removing themselves from the industry. They're so worried about these bad uses. I think if there were a treaty banning, some of these bad uses people would feel better about working in the area. They wouldn't worry that their boss might tell them to work on a military project that they just can't countenance.

## The importance of a regulation or ban for industry

**Matilda Byrne:** [00:22:16] For sure. I think that really speaks to the heart of why regulation is particularly important, especially for the industry; people knowing that there is a line of "this is what we may be building and can innovate, but also this is what will never be built and I can, as a worker, with confidence know that my work will not be used in this way" because there is a ban in place and there is stigmatization of these weapons and they are considered unacceptable. I think as well, the delineation in law makes that a lot stronger. So one part about having a ban in an international scale is it can then be legislated and it's understood internationally that we have these particular lines in place, the delineation.

## Individual companies taking action/forming policies [00:22:59]

**Matilda Byrne:** But I guess that on the other side of the coin, there's also things like companies making their own policies saying that they won't do contracts with defense on these sorts of things. What do you think the place is for action like that within the tech industry?

**Lizzie Silver:** [00:23:13] I think that's really important. Um, but I also think it's really hard to get right and hard to sort of set the level for a particular company. There is always some project where, you know, most of the people in the room are gonna say that's fine and other people are gonna have a problem with them. For example, a vegan employee is unlikely to want anything to do with animal agriculture. Whereas other people will be fine with it. We got offered a project at Silverpond years ago, before I joined the company, actually- this goes to the dual-use point as well. So the idea was to create

an AI that could target and kill feral animals in rural areas. And we turned that down because it was just too close to an automated killing machine, but feral animals in rural areas are a real problem and very hard to control. They're really bad for the environment. So people can have different ideas about which projects are okay. In which aren't. If we knew that that tool for controlling feral animals were never going to be used to kill humans then maybe we would have been okay with it depends on the person. So I think each company needs to set their own line and we've taken the approach [00:08:00] that if any employee's not okay with a project, they don't have to work on it. I think that's also really important.

## How can individual tech workers take action?[00:24:44]

**Matilda Byrne:** Mmm. I think that's really good, yeah, giving individuals a bit of ownership and decision making so that it can promote responsible innovation, ethical thinking by employees. Obviously AI ethics in general is becoming a really big field. And I guess more on perhaps the individual worker, if you are listening, there's some great things that you could do. So Future of Life have a open letter that's for roboticists and kind of AI people in these kind of technological areas. And it's basically an open letter pledging to not be involved in the development of fully autonomous weapons and calling for a ban. So if you've been listening to this and you're really compelled, I would really call on you to go and check that out and sign it if you're concerned about this issue and want to see action.

[00:25:37] Do you think Lizzie as an individual there's anything else beyond kind of, I guess, conscious decisions that you think that people could do maybe within their company?

**Lizzie Silver:** [00:25:46] Um, I think becoming aware of the campaign is really important. Eventually this is a political process, right? We have to convince politicians to sign onto a treaty. If we could get an industry wide stance, so people could talk about this and come to a position on it, then we'd have a lot more power to influence the government on this. I think right now, most people in tech are just not aware that the Campaign to Stop Killer Robots is a thing, so spreading the word is important. And just thinking about how it could affect your work. Maybe it doesn't now, but think about what you would say if your boss asked you to work on a military project where your model was [00:10:00] going to be used to target people. What would you say?

**Matilda Byrne:** [00:26:40] For sure. I think there's some really great concluding thoughts from you, Lizzie, so thank you so much for joining us.

**Lizzie Silver:** [00:26:46] Thanks for inviting me.

**Matilda Byrne:** [00:26:49] My pleasure. It was great to have you on. If you do want to find out more and get educated, more awareness; stopkillerrobots.org  is the global website where you can find information or you could check out a new report Australia Out of the Loop available on the SafeGround website.I hope you enjoyed today's episode on the tech perspective.

If you want to know more, look for us on Facebook, Twitter, and Instagram Australia Campaign to Stop Killer Robots or use the hashtag #ausbankillerrobots.  Become part of the movement so we Stay in Command.

Thank you for listening, and please share with your friends. For access to this and other episodes along with full transcription and relevant links and information, head to safeground.org.edu/podcasts

[00:29:39] Our podcasts come to you from all around Australia. And we would like to acknowledge the traditional owners throughout and their continuing connection to country, land waters and culture.

Stock audio provided by videvo downloaded from www.videvo.net -
Thank you for listening to Safe Ground