Ethics of Robotics, Autonomous Systems & Artificial Intelligence

Video transcripts for the Introduction to ethical robotics, autonomous systems and artificial intelligence in Defence and pragmatic tools to manage these risks.

Trusted Autonomous Systems for the

Centre for Defence Leadership & Ethics, Australian Defence College

Contact
Dr Kate Devitt, Chief Scientist
Trusted Autonomous Systems
Kate.devitt@tasdcrc.com.au
M: 0413 977 917

Summary

"Military ethics should be considered as a core competency that needs to be updated and refreshed if it is to be maintained" Inspector General ADF, 2020, p.508[1]

Robotic, autonomous systems and artificial intelligence (RAS-AI) can help protect Australia and project our strength RAS-AI are a force multiplier accelerating the tempo and quality of decisions to enhance capability. Humans are responsible for technologies they employ. The decisions made by human-machine teams must be justified and aligned with human values, intent and expectations. RAS-AI can remove humans from high-threat environments; reduce capability costs and achieve asymmetric advantage. They can be consistent; immune to stress, hunger and fatigue. RAS-AI can improve ethical decision-making across contexts in Defence; acting as an ethical safety catch; and we must intervene when humans or technologies act contrary to Australian values and intent are unreliable or put humans at risk of unintended harm. We must build technologies to minimise ethical risks. The Defence Science & Technology report 'A Method for Ethical AI in Defence' helps Defence operators, commanders, testers or designers ask five key questions about the technologies they're working with.

Responsibility – who is responsible for AI?
Governance – how is AI controlled?
Trust – how can AI be trusted?
Law – how can AI be used lawfully?
Traceability – how are the actions of AI recorded?

There are four tools that may assist in identifying, managing and mitigating ethical risks in Defence AI systems. The 'Data Ethics Canvas' by the Open Data Institute encourages you to ask important questions about projects that use data and reflect on the responses. Such as the security and privacy of data collected and used, who could be negatively affected and how to minimise negative impacts. The AI Ethics Checklist ensures AI developers know: the military context the AI is for, the sorts of decisions being made, how to create the right scenarios, and how to employ the appropriate subject-matter experts, to evaluate, verify and validate the AI. The Ethical AI Risk Matrix is a project risk management tool to identify and describe identified risks and proposed treatment. The matrix assigns individuals and groups to be responsible for reducing ethical risk through concrete actions on an agreed timeline and review schedule. The LEAPP is a Data Item Descriptor (DID) for contractors to Defence to develop a formal Legal and Ethical Program Plan to be included in project documentation for major programs that had a significant AI component. The checklist, risk matrix and LEAPP are all available in 'A Method for Ethical AI Defence Report' at https://www.dst.defence.gov.au/publication/ethical-ai

[1] Inspector General ADF, (2020), Afghanistan Enquiry Report (The Brereton Report), Commonwealth of Australia, ISSN 2207-6069 (Online)

TRUSTED AUTONOMOUS SYSTEMS
DEFENCE CRC

Table of Contents

Video 1 – Introduction to the ethics of RAS-AI in Defence

Introduction

GANOS Hi my name is Wing Commander Michael Gan from Air Force Plan Jericho.

I'm in the Australian War Memorial and I can't think of a better place to talk about the interaction of ethics and technology than where we are right now, which is under the V1 flying bomb from World War II. Robotic and autonomous Systems are systems that can do part of their job (or all of their job) without human intervention; and artificial intelligence is a really huge field about machines that can think a little bit like humans. So two really large fields and we've seen a lot of those come out in the last couple of years.

See here behind us at the V1 is one of the very first applications of those robotic autonomous systems in the military; which was used to send weapons across from Germany to England without risking the life of German pilots. It was one very early application to reduce the cost of war in that way. But it also had a whole heap of other issues that came out in the way of ethics because the weapons were not guided, they were imprecise, they were indiscriminate and had a lot of ethical issues that we would have to deal with today.

Robotics and autonomous systems have a great deal of utility: They can reduce casualties, reduce risk, they can be operated in areas that may be radioactive or unsafe for personnel. They can also use their capabilities to go through large amounts of data and be effective or respond very quickly to rapidly emerging threats. However, because a lot of this is using some sort of autonomous reasoning to make decisions, we have to make sure that we have a connection with the decisions that are being made, whether it is in the building phase, whether it is in the training phase, whether it is in the data, which underpins the artificial intelligence, robotic autonomous Systems.

So, we have an obligation, whether we're designing, whether we're setting the rules, or whether we're operating robotic and autonomous systems, to be fully aware of their limitations and their uses and how we're going to use them. But also the tasks that we're going to give them and the environment which they will be used in. So definitely the task and context are critical for us to understand in bringing these new technologies, which have a great deal of potential.

MICK Hi I'm Major General Mick Ryan, I'm the Commander of the Australian Defence College, I'm also a General in the Australian Army.

How do we use artificial intelligence to supplement some human cognitive functions so that we can make better decisions potentially more quickly (but not always more quickly--sometimes we might want to make them more slowly) but how do we make better decisions that are more connected across a more integrated organisation and then test those decisions before we implement them? I think this is a real area where we can make some advances as to how we run war games more quickly, with potentially tens of thousands of iterations, very

quickly, to support senior leader and mid-level leader decision-making. How do we use it to help us prioritize things like logistics, medical support, and these kinds of things?

Potentially, how do we use it better to support what we build internally and what we buy from overseas, possibly? So, that's the thinking piece. And there's any number of other functions we might be able to use in training and education. Then the doing phase is how do we use robotic systems to supplement human physical capacity. AI supplementing cognitive capacity, robotics is about supplementing physical capacity.

Now we got a lot of experience in this, right? We've been supplementing human capacity ever since people developed wheels and tools to till the ground in the agricultural revolution thousands of years ago. So we know how to use tools. But when these tools have a level of cleverness, not intelligence, but a level of cleverness, that's a bit different and we need to continue evolving our ideas and concepts for using them and how we prepare our people to work as a team member for these things, not just as a tool-user for these things and I think that is a mindset change that we have not yet progressed towards in our organisation.

My first real exposure to some of these things was in Iraq in 2005, working with the Americans and then Afghanistan in 2006 where we had an American detachment that had a tiger shark unmanned aerial vehicle, unarmed, of course. But we worked very closely with them and use them for a lot of intelligence gathering and surveillance before missions, during missions, but also for deception as well. I didn't have the ethical dilemma of autonomous systems using lethal weapons. But sometimes you would make the decision to use it for deception rather than for the mission itself and you had to decide what was the better course of action. When I was Brigade commander in the first brigade, we had a few more of these systems, including having the infantry and engineers trial robotic systems for clearance activities. And my view is, if you can put a robot between something that blows up and a human being, that is a really good thing.

LAUREN My name is Lauren Sanders and I'm a Doctor of International Criminal Law. I've spent 20 years in the military as both a signals officer and a legal officer and my area of expertise is largely operational law and the application of International Humanitarian Law to ADF operations. I am also the Managing Director of a small legal firm called 'International Weapons Review' or IWR, and we focus on providing industry advice on how they can operationalise their capability focussing on legal compliance.

I think a great example of the use of autonomous systems in Defence was the introduction and the development of the way that Defence used UAS in an ISR capability in the targeting cycle. So a great example from recent operations would be the use of ISR to augment the targeting capability or the visual range of an Apache helicopter in operations in Iraq in 2017. Then MUM-T, but now, HUM-T, so human-machine teaming, which was using the capability of the Reaper to effectively act as a forward location to visualise what the pilots were looking to target.

Interestingly, because that system was one that was being used or tested in operations, the level of trust by the Command wasn't there yet to actually properly rely on that system to cut

that visual feed to the pilot, to trust them to take the target as a consequence of what feeds they were getting from that UAS. What was actually happening instead is that those systems were being used to assist in target verification only after the targets had been verified through the traditional deliberate targeting process. So, hopefully, a mature version of that system will be using that HUM-T process to actually speed up and extend the range of those capabilities without having to come back to a Command decision headquarters to actually go through that deliberate targeting process, but use it more as a dynamic targeting system.

Capability Advantage of RAS-AI [07:05]

MICK: Technology is never a silver bullet in any military institution. It really is about the ideas, the new organisations, and the people who use these new technologies. So, how do we team these things with people in a way that makes us more likely to be successful in our missions, from warfighting through to humanitarian assistance, disaster relief; and do it in a way that accords with the values of our country first, and foremost, and then our institutional values as well. But that's kind of the framework I use when I think about these systems.

I guess the thing I think about now is we still don't have a lot of them, we might have a few thousand autonomous systems. But I can foresee a day where instead of having one autonomous system for ten or a hundred people in the ADF will have a ratio, that's the opposite. We might have a hundred or a thousand for every person in the ADF. We don't have a training system, we do not have an organisation and we certainly don't have an ethical framework at the moment that would cope with that kind of density of massed autonomous systems; and that doesn't include autonomous systems that might be able to make the decision to kill a human being.

That occupies my thinking: how are we leaning into the environment in what might it be in 10 years? How do we enable our people to not just cope in this environment but maintain a level of decision-making capacity and awareness of what's going on, so these systems don't get away from us.

JASON: Professor Jason Scholz I'm the Chief Executive Office for Trusted Autonomous Systems which is a Defence Cooperative Research Centre.

I think there's a huge capability advantage to come from RAS-AI for the Australian Defence Force. We have a vision in our Centre for what we term the smart, small, and the many. This is a contrast to a platform's view of the past, which are complex, large, few and often manned systems. Those manned systems we really can't get enough of and we can't afford to get enough of. So ,we see this vision for smart, small, and many to be a transformative one, and a disruptive one.

I think it will mean ways to develop and achieve what's termed mosaic warfare in the US which is the ability to set up multiple simultaneous kill chain options. This can be a huge deterrent to an adversary and provide many more options in defence and offense. I think also reflection: If you look at the conflict between Azerbaijan and Armenia late last year, is an

idea of what the shape of things are like today using these disruptive technologies, much less the things that we're developing and are being developed.

Technology, people, and systems together can make for better decision-making in a military context. In order to do that, we should recognize some things: people behave differently when they're accountable and perhaps they know they're being watched. People watch the technology during operational use, but it can work in another way as well, because not only might those robotic systems be watched, but the robotic systems can watch one-another and they can watch what people do. So, an aberrant behaviour might be due to a failure of a machine. For example, might be picked up by other machines and communicated and it might signal an attack or a signal that something's gone wrong with an algorithm and so on. So, we have the potential to have a much more systemic perspective on accountability.

STEPHEN: I'm Stephen Bornstein, I'm the CEO of Cyborg Dynamics Engineering and I'm the Managing Director of Athena Artificial Intelligence, which is a spinout AI Company, initially created through the Justified Autonomous UAS Project within the TASDCRC. So the ADF is quite a small force in terms of its numbers, but within Australia, we have a lot of technology and particularly with Australian industry content, we're now seeing the ability that we can produce these small, semi-disposable, or attritable systems at a low cost that can achieve high payoff based on the level of autonomy and AI and how we actually network to achieve a combined effect with a small number of operators commanding them.

What makes an ethical decision-maker? [11:26]

MICK When I was in East Timor in 2000, I was serving with the 6th RAR Battalion group. I had the engineers and we were in charge of keeping the roads open, especially during the wet season. One particular day, the road was closed between Balibo and Maliana on a hillside and there were landslides and I had to make a judgment about whether it was safe to send our sappers out to clear the hillside; clear where it had gone over the road, so we could reopen the roads and that was potentially putting them in significant harm. It wasn't about being shot by someone—it was about could another landslide see sappers and their equipment being washed down a hillside and hurt or killed. That was a real dilemma.

Now we ended up going ahead, but it took us a lot of research and a lot of work beforehand and risk assessment with myself and other people to move ahead. That's just one of the kind of ethical dilemmas you have not just during everyday life, but in operations as well. Well, the first thing I do is who else has faced these kinds of dilemmas previously. How have they dealt with them and then I think about what's the purpose of doing or not doing certain actions, and then I think about primarily the people, not just their physical welfare but also their moral welfare and what might be the impact on them of doing something that either won't accord with our values or won't meet the mission in a way that we should or might just have a short-term outcome that will compromise longer-term strategic objectives. But most importantly, I think, is that you need to think about the moral welfare of those who are involved.

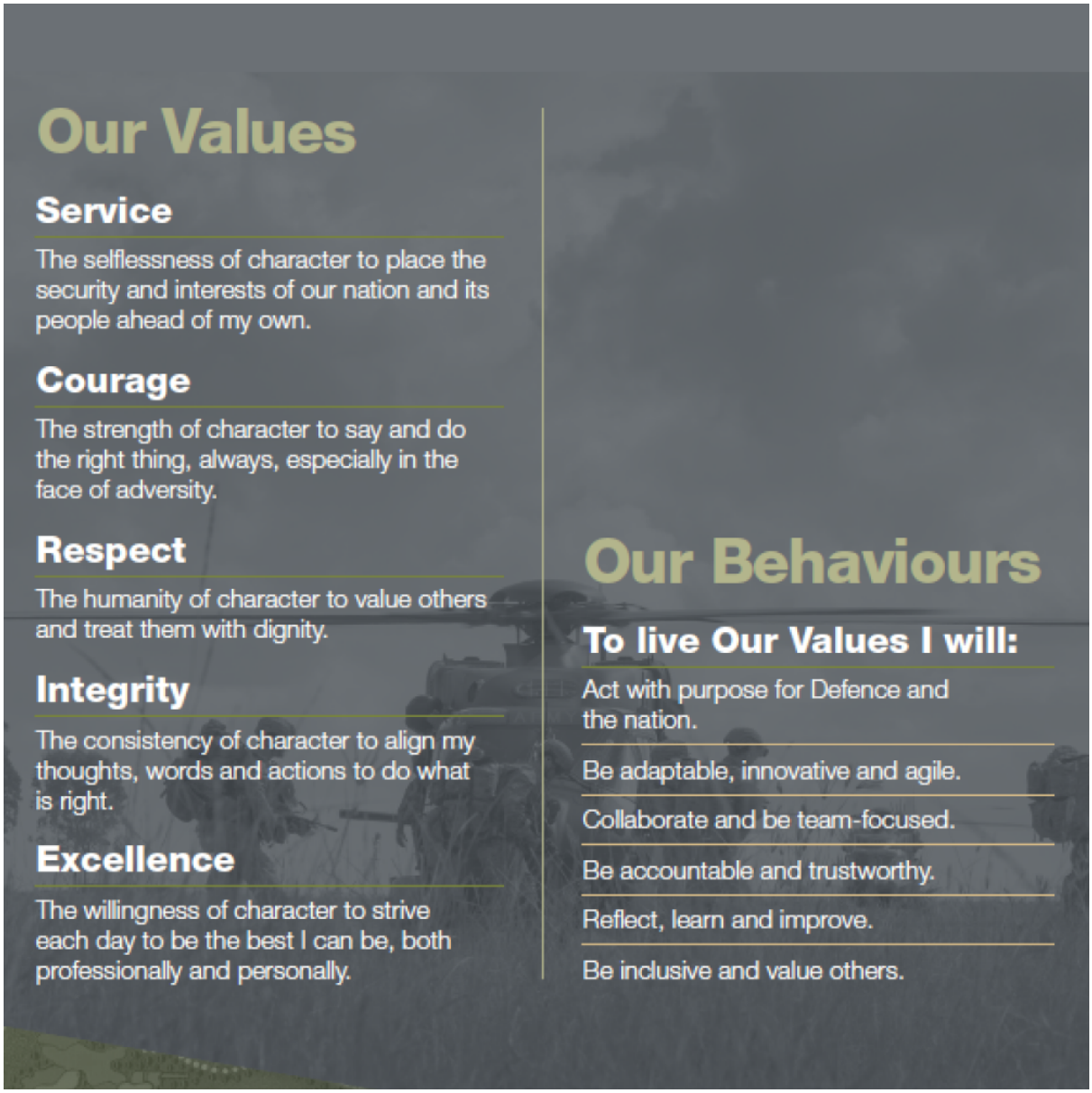
I am by nature someone who likes to move forward, which can make me impatient, but it can also sometimes mean you don't take all the advice you need to take. So you need to check yourself and make sure you are exposing yourself to a diverse range of views. Sometimes you don't have time to seek the full range of views or opinions that you might need. Sometimes you just need to make a decision.

Every situation is different, but you get better at it the more decisions you make and really good decision-making is about making as many decisions as you can and learning about what works for you, in both analytical and kind of hasty decision-making over many years and indeed, decades. Well, firstly, you've got to understand your own country that you're serving. You can't make ethical decisions without having a contextual understanding of the values of your country. What it stands for, what it believes in, and what it'll fight for. So, that's the first and probably the most important thing.

Secondly, you've got to understand the values of your institution, and the ADF, over the last couple of years have gone through, developed, and implemented a new set of institutional values, which have been very important, I think. You've got to ensure people understand what those values are and then train constantly and drill people in what is the ethical way of staying with those values, in what are pretty austere, ambiguous circumstances and operations. You can't just give people a lesson on these values, you need to drill people on the different kinds of rules of opening fire, the rules of engagement, that kind of thing. It needs to be almost muscle memory, not something that's academic and maybe useful. This has to be something that's integral into how people think and how they behave. And it's not just leaders, it's everyone.

The rules of armed conflict and rules of Engagement we run people through drills before they go on operations to try and build that muscle memory of good ethical decision-making when they go on operations. So, I think in the ‘thinking’ piece, you can use artificial intelligence and clever algorithms, and simulations to expose people to lots and lots of different scenarios. It doesn't provide a template for them; it doesn't prepare them for everything. But it gives them a mental model of how they might think about these things when circumstances turn bad.

Once again, the physical stuff. We need to prepare people to team with robotic systems almost from the day they join the military, not from when they go to a unit, but these things need to be teamed with people right from the start. There's no point arguing how many these things are going to be or if they're going to be; they are going to be part of the environment; they're going to be part of society and we're going to see more robotic systems doing more things. We need to team our people early and we need them to understand that they're going to be an ever-present part of their life in a military institution.



https://cove.army.gov.au/article/unified-defence-values-and-behaviours

JASON: Certainly, character ethic is reflected very strongly in defence values, which are part of our ethical viewpoint. So, those include:

> **Service**, and placing others above self;
> **Courage**; strength of character, to say and do the right things;
> **Respect**, which includes the humanity of character and the valuing of others;
> **Integrity**, and consistency of actions in character to align thoughts, actions, and words in doing what's right; and
> **Excellence**, which is that willingness of character to strive to be the best you can be to live the values,

I also note Defence has a set of behaviours about acting with purpose for defence and the nation; being adaptable, innovative, and agile; be collaborative and team-focused; and to be accountable and trustworthy to reflect, learn and improve; and to be inclusive and value others. All of these values and behaviours are included whether we are a 'robotic and autonomous systems' augmented force, or not.

Now, ethical decision-makers are human, but they are also machine decision-makers and machine decision-aids. We embed human conceptualisation in the machines so they are our representatives in battle, in effect. By this, I'm not meaning that machines may have moral autonomy (or not yet, at least anyway). Yet, if machine learning means we train machines; are we not imparting our values through that process? So perhaps then, machines might comply, in some sense, with defence values in a rather unique way.

This may be a little dangerous, but some food for thought--perhaps those values of service for a machine can mean machines operating as tasked under mission command.

Maybe courage would mean machines operating within the ADF system of control, as put forward by the Chief of the Defence Force, and that includes the courage to intervene if the machine believes, if you like, a human has made an error. For example, a weapon to prosecute a target that has been identified locally by the weapon system to be a red cross through machine learning and might defer that action or respond back to the operator requesting the operator to override it, or override it prior to the action, if necessary.

Respect is another one of those values. Machines accept human guidance and they log it for the record.

Integrity with machines processing, reporting, and acting consistently within commission, command, constraints; and

Excellence, machines optimizing to be the best they can be, in support of human command.

So not only is there a question about what it is that makes an ethical decision-maker, but who it is that makes them. So, what are the principles involved in the research, design, development, and use are also relevant to that question? How will you deal with, in the future, the potential for the machine to say no to you, that the machine may object on legal or ethical grounds for an action that you're requesting?

What are ethical risks of robotics, autonomous systems and artificial intelligence? [19:26]

MICK Humans must own the decision-making process. They can augment their cognitive and physical functions—I don't see a problem with that. But when machines start making decisions about human life, there's a real issue there. I know that these kinds of research say machines often make better decisions, autonomous cars have less accidents than humans, and autonomous aircraft have less accidents than humans; and I understand that, and I'm not

disputing any of that research. But the military is the only element of society that can lawfully take a human life as its first option and human beings and military institutions are subject to an unlimited liability. That imposes a profound responsibility on commanders to make good decisions about taking the lives of others or risking the lives of their own people. We cannot hand that over to an algorithm. They do not understand the context of why human life is important. They do not understand the context of why protecting it instead of destroying it is always our first option.

JASON There are a lot of concerns by the public and defence people alike that our ethical standards may be quite different from our potential adversaries and this kind of ethical asymmetry, in a sense, might create concerns that we are ethically conservative and put our defence force at a disadvantage. I don't think we should lose the ADF's advantage If we get the systems right and the processes and policies. For example, in Autonomous Warrior 18 we developed and demonstrated a policy-based control system for robotic autonomous systems and AI. This allowed management of time, space, spectrum, but might also manage targeting related processes by changing in effect, rules of engagement and electronic policy, which can be implemented throughout the entire network, or 'Internet of robotic autonomous systems Things'

A Method for Ethical AI in Defence [21:37]

KATE I'm Kate Devitt the Chief Scientist of Trusted Autonomous Systems Defence Cooperative Research Centre.

In 2019, Plan Jericho Air Force wanted to investigate the possibility of AI ethics principles for the Australian Defence Force. They worked with Defence Science and Technology Group and the Trusted Autonomous Systems Defence Cooperative Research Centre, to have a workshop to figure it out. At that workshop, over a hundred people from 45 organizations thought about what would be good AI ethics principles to have in an Australian context, supporting Australian values, with evidence from experts across many different subject area domains, from military ethics, human factors, engineering, artificial intelligence, a military context, legal scholars, etc…. So we brought them all together to find out what we might want to do in an Australian context.

www.dst.defence.gov.au/ethicalai

From that workshop came 'A Method for Ethical AI in Defence' DSTG report. That now provides the first step of a best practice framework for ethical AI development, testing and use in the Australian Defence Force and Australian Defence more broadly. That framework recommends five facets of AI

1. Responsibility,
2. Governance,
3. Trust,
4. Law, and
5. Traceability.

These facets encourage us to ask questions rather than provide specific answers. Under responsibility: Who is responsible for the AI that you're developing or going to deploy? Is it a commander? Is it an operator? What is the system of control in the system of responsibility?

And how are that responsibility and accountability managed inside Defence?

Governance: governance covers a huge range of mechanisms that might enable and constrain the use of robotics, autonomous, systems and artificial intelligence in Australian Defence. Governance can cover laws, it can cover culture, it can cover operational guidelines, or safety manuals. So there are many systems of governance that can be worked in together in interlocking ways to ensure safe and effective use of these systems in Defence.

Trust is a really interesting aspect of the facets of ethical AI for defence. Trust is multi-faceted. It is certainly about competence. You want to trust something that's reliable, that's got enough experience, enough skills in order to get the job done in a way you anticipated to be done appropriate for the context of its deployment. But it's not just about competence. So, for example, you might trust Google Maps to get you from A to B, because you know it's a really reliable system. But you might also be a bit worried and say, “hmm. What is Google doing with my data? Am I agreeing to the use of my location by Google in some future way of interacting in the world?” You might ask that question. In that case, you're not questioning the reliability of Google, but you might be questioning its integrity.

So, integrity is the other aspect of trust worth focusing on. We don't just want competent systems. We want those that have a high integrity level. Well, what is integrity? What for human's probably got to do with character, motivation, and honesty—and how to transplant that into a machine is an interesting question and maybe there is some asymmetry there. But certainly, we can make machines honest in the sense that they can be suitably transparent—within security considerations—with humans, so that humans can understand what they're doing and that humans can explain and justify how they're integrating with those systems and machines.

You can certainly have good integrity in the design parameters of robotics, autonomous systems and artificial intelligences and you can test for the use of those systems with humans and with human decision-making in an evaluation context in order to make sure that the integrity of those systems is maintained when they're deployed.

Law, it is a big topic. We've included it in the method for ethical AI for defence framework, because abidance with international legal obligations is a key part of what Australia commits to. So Australia follows International humanitarian law and all new means and methods of warfare need to go through Article 36 reviews and so we have included law as a key part of what it is to be ethical using robotics autonomous systems and artificial intelligence.

Traceability is the fifth and final facet of ethical AI for defence. Traceability means that the decisions that are made by these systems are traceable to humans in post-event reviews and in the analysis of how we're making decisions, what went into those decisions, who was there, when did they occur, and what were the consequences of those decisions. This enables us to grow and learn, both in simulation and in the unfortunate case of a deployment in a conflict, enables us to consider how they're used and to improve the way these systems are used into the future.

Video 2 – Pragmatic tools for considering and managing ethical risks in AI for Defence

Introduction

VO There are many benefits to increasing AI and autonomous systems capabilities in Defence, including removing humans from high-threat environments, reducing sustainment costs, achieving greater mass on the battlefield, exploiting asymmetric advantage, accelerating capability development timelines and capitalising on advances made in the civil sector. An effective and practical methodology to reduce ethical risks will best support Defence and industry in developing AI systems. Developers already must produce risk documentation for technical issues. Similar documentation for ethical risks ensures developers identify, acknowledge and attempt to mitigate ethical risks early in the design process and throughout test and evaluation.

This video introduces four tools. The first is 'Data ethics canvas' by the Open Data Institute. The remaining three tools have been developed specifically to assist Defence and industry in developing AI systems for Defence.

1. An AI Checklist for the development of ethical AI systems
2. An Ethical AI Risk Matrix to describe identified risks and proposed treatment.
3. A Legal and Ethical Assurance Program Plan (LEAPP) to be included in project documentation.

Data Ethics Canvas (1:32)

https://theodi.org/article/data-ethics-canvas/

KATE: The Data Ethics Canvas by The Open Data Institute, this is an open-access canvas, and it enables you to really interrogate why are you collecting the data you're collecting, what problem were you trying to solve with this data, what sort of security protocols and privacy constraints need to be put on the access to that data in the way it might be utilized, who might be harmed by the collection or use of this data in an AI project, and what might you be doing to mitigate the risks of the use of that data. Because at the end of the day, AI is powered by data. So if you don't have a good grip of what data you're using and why and making sure that it's secure end-to-end, then you're not being ethical with your AI project.

STEPHEN: The Data Ethics Canvas for us, was an initial framework to allow engineering teams as well as non-engineering teams, so the military personnel who are advising the project to consider a variety of ethical aspects of our data, how we curate the data, preventing biases in our data, ensuring that we can trace our data back to its original source for our machine learning models. That gave us an initial framework and almost just a brainstorming session of things that we need to consider for our specific project from an ethical standpoint. The Data Ethics Canvas helps us consider the end-user and who might be associated with the results of the AI by actually asking us those sorts of questions of what happens if something goes wrong. What if there's unjust bias, and how do we mitigate against those things? And then, our Engineers can go away from that and work out how do we produce an unbiased network: What happens if our AI labels something in a way that it shouldn't label; how do we change our confidence levels to be more sure about one thing, but we don't have to be as sure about another thing. So, we flag things from an ethical standpoint if we're really unsure about something and provide that information to the user to prevent their cognitive bias, that they make a better decision.

Ethical AI Checklist [3:28]

A	Describe the military context the AI is for	E.g. Force Application, Force Protection, Force Sustainment, Situational Understanding, Personnel, Enterprise Logistics, Business Process Improv.
B	Explain the sort of decisions AI helps with	E.g. Is it a single decision-maker, multi-decision maker; once-off decisions vs. sequential decisions
C	Explain how the AI integrates with human operators to ensure effectiveness and ethical decision making in the anticipated context of use and countermeasures to protect against potential misuse	E.g. What are the human factors and system factors and what are your scenarios and T&E process?
D	Explain framework/s to be used	E.g. Method for Ethical AI in Defence, safety frameworks, human factors and legal frameworks suitable to the context etc…
E	Employ subject matter experts to guide AI development	E.g. If team lacks the expertise to undertake one or more of steps A-D, then they should onboard the skills gaps through hiring, consulting and oversight.
F	Employ appropriate verification and validation techniques to ensure compliance	E.g. Auditability, accountability, explainability and lawful abidance must be demonstrable

VO: The components of the AI ethics checklist are:

A. Describe the military context in which the AI will be employed. For example, is your system used for force application, force protection, situational understanding, sustainment, enterprise functions and so on
B. Explain the types of decisions supported by the AI. Are these one of decisions by a single decision maker? Sequential decisions? Cooperative decisions or decisions under conflict?
C. Explain how the AI integrates with human operators to ensure effectiveness and ethical decision making in the anticipated context of use and countermeasures to protect against potential misuse. Create scenarios and show how systems operate in these scenarios. Test using gaming and simulation and experimentation. Demonstrate in trials and exercises.
D. Explain framework/s to be used, for example regulation, ISO standards, code of practices, or AI Ethics tools such as the data ethics canvas, LEAPP and AI Ethics principles
E. Employ subject matter experts to guide AI development, for example, data and AI specialists, end users, decision scientists, military ethicists, legal experts, systems engineers and so on
F. Employ appropriate verification and validation techniques to reduce risk. Including providing a requirements analysis and verifying these; providing a logical analysis and validating it.

KATE: The AI ethics checklist is a set of checks and balances to make sure that your AI project is considering ethical risk in an appropriate way. The main idea is pretty simple, you need to imagine how your technology might be deployed in its anticipated context of use. So,

some AI Technologies are going to be really useful, for example, in an HR setting and will never be used out on the battlefield and vice versa. So, what's the context in Defence that your AI system is going to be deployed for? (see ANNEX A) Now, come up with scenarios under that deployment, which involve ethical risk.

Now, you may not know in your project team, what those ethical risks might be, so you need to bring in those who do. So, that's the third part, make sure the subject matter experts that you have inside your team or outside of your team are brought into workshops in order to understand and to identify ethical risks in the first instance, and then to sort of workshop mitigation strategies for those risks that you can put into your risk matrix.

Who are the people you need? Well, you probably need people who are good at decision science. It's actually understanding the kinds of decisions that are going to be made by the artificial intelligence and the humans involved with the technology. You probably need to bring in military philosophers and other kinds of philosophers or ethicists who have backgrounds in ethical theories, ethical reasoning and decision-making. You want end-users in the military—so, who are the Defence personnel who are going to be asked to use this equipment? How do they understand its intended use in the intended context because they have the deep expertise in the actual decisions that need to be made on the ground. So, you want end-users and you want those in senior positions who might be the ones authorizing the use of those technologies as well. You want people who have human factors background experience, those who are really well-versed in how to bring humans and machines together and the sort of information interfaces that will enable and optimise human decision-making and those mixed teams to flourish under deployment. That's just some of the expertise that you might need to bring to bear. You're going to need people with legal backgrounds to manage the legal considerations. You're going to need data privacy and security experts if you're dealing with particular issues around those matters. So, each part of these systems will need experts who are able to provide best practice, apply legal frameworks, apply regulatory requirements, safety requirements, and other types of considerations to ensure that these systems are trusted when they're fully developed.

Finally, once you've got those scenarios and you know what the risks are, you're going to need to build the systems in order to mitigate the risks and then test those systems. You're going to need to test and evaluate the AI systems with human operators to make sure that they work as intended. And that you can verify and validate the algorithms and the decisions that they're being made before you're hoping to get those technologies to be acquired by Defence.

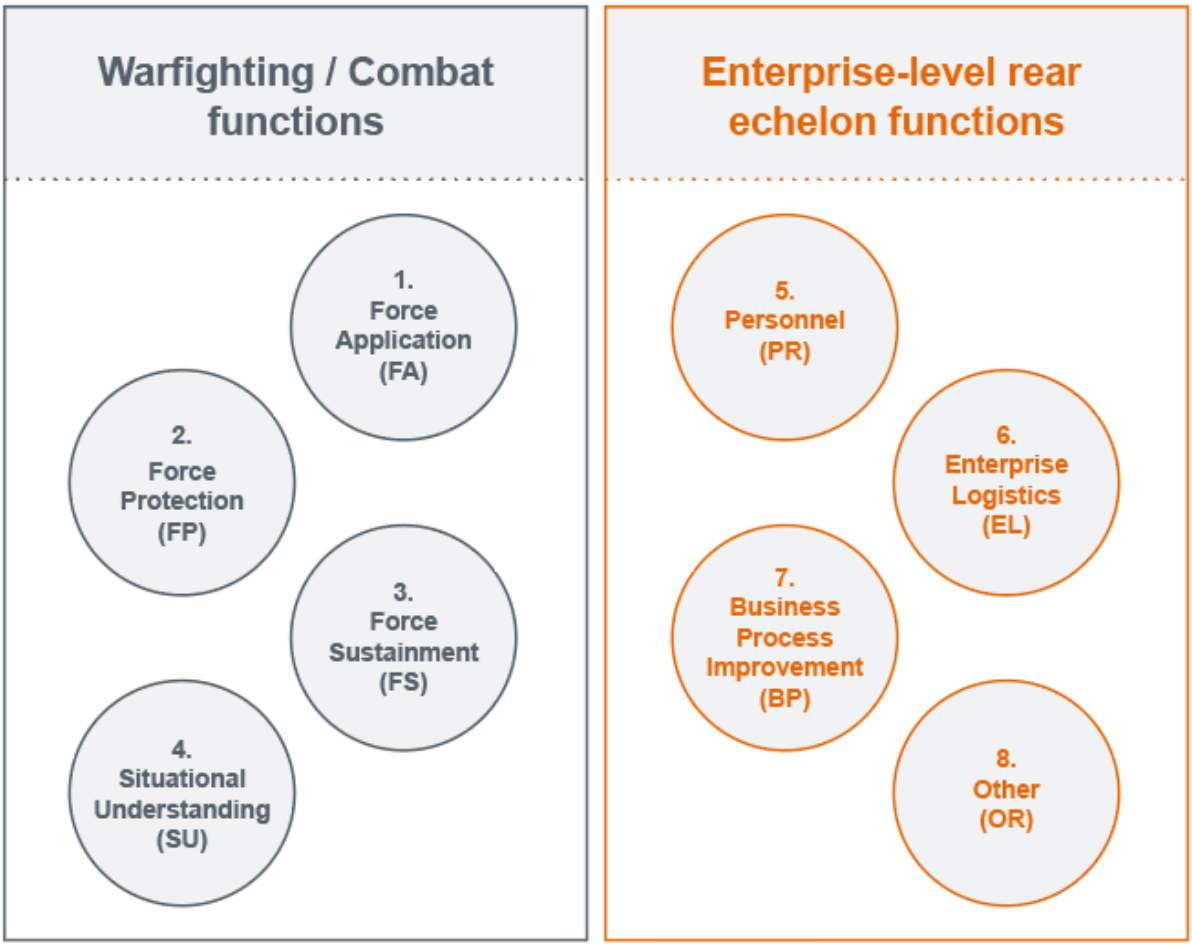
When we think of the sort of decisions that robotics autonomous systems and artificial intelligence can help us with, it's important to realize that decision making and decisions themselves can have different structures depending on what an agent, a human, a robot, or an artificial intelligence are thinking about and what their obligations are. For example, I'll give you four different kinds of decisions that might change the ethics of how you might deploy these systems (see ANNEX B).

The simplest case is when you have a one-off decision by a single decision-maker. That might be your decision to go to sleep at 9 p.m. or go to sleep at 9:30 p.m. That's just one decision, and the consequences of that decision, you'll feel the next morning if you needed a little extra sleep. We might think of a robotic situation, a drone might be flying autonomously, for example, and it might have a set of parameters where suppose communications go down, and that drone has a decision point, and it says, when my comms go down, I need to go home. So, that drone might have a one-off decision to go home back to base, based on its programming.

Contrast that with a sequential decision-making algorithm. If you've got an autonomous flying drone and it is manoeuvring around and navigating, it might be making hundreds of decisions about how it's actually navigating through the air and responding to the environment that it's flying in. That one decision depends on the previous decisions that it was making. Those sequential decisions are going to be very influenced by the previous decisions that that system has made and often don't have an intervention on them. So, sequential decisions have consequences based on previous decisions and ongoing dynamic changes. One-off decisions are things that happen and then there is an action that occurs.

Another sort of decision is what happens in collective decision-making contexts. So, if you've got a number of assets, for example, in a search-and-rescue operation, you have different kinds of human groups and different sort of robotics and autonomous systems assets, they need to work collectively and cohesively together in order to make decisions about how they need to save a person or persons from a harm situation. That's really different from a single decision-maker and will have different ethical risks associated with it.

The final example is decisions under conflict and when you have different interests at stake. The way you might have to make a decision when you're facing an adversary who might be seeking to harm you or those that you're caring about, again, provides a different set of considerations and those considerations are relevant regardless of whether you have humans or robotics, autonomous systems and artificial intelligence as part of making those decisions.



GANOS For example, you could use in Force Application or Force Protection an autonomous vehicle (either for attack or defence) or even carry ordinance, cargo, or personnel. So, that's within the autonomous systems. You could use artificial intelligence in Command and Control. So, one of the applications we're looking at is how you Commanders make better decisions by taking recommendations of artificial intelligence systems that can gather enormous amounts of data and bring out really important things for the Commander.

You could also use artificial intelligence for force protection, such as for medical purposes. One of the real effective uses of artificial intelligence is to help doctors identify and diagnose issues by taking a large amount of information from their patients, such as scans, and identifying things like lung problems, cancers, tumours, and similar things like that. In the area of Situational Understanding, artificial intelligence is also being used and can be used to take in a whole lot of digital imagery and identify things of interest and classify things within that imagery.

In the areas of Force Generation & Sustainment, you actually have the greatest number of artificial intelligence applications, including in logistics, optimisation, prediction, and because Force Generation & Sustainment, even though it's the least combat of these

functions, it is closest to the commercial and the civil area and there's a great deal of artificial intelligence applications in that area as well.

So, when we choose to use these applications in the military context and in conflict, or in the grey zone or close to conflict, we do run into a lot of ethical issues, and there's a lot of things that we have to consider that they even considered back in World War II with the new weapons, as technology was brought in in each of those applications. I'll give you some examples from those contexts that I mentioned before.

For example, in the Command and Control situation, the artificial intelligence, which is supporting the Commander in his decision support mustn't be biased; it must know what information the Commander wants and doesn't want. But most importantly, the artificial intelligence should have value alignment. So, it should have the same values as the Commander, or the government, or the force which it's working for--that's a key thing of importance. Another example of an ethical situation will come from the medical example I gave you earlier for Force Protection. We would envisage in the very near future the ability to monitor lots of medical parameters from our combatants and our non-combatants to make sure that they stay safe, to see if they're wounded, and to ensure that they operate at their maximum capacity.

This brings up a whole lot of issues. For example, where does privacy come into it? How many things can you monitor of a person in the services or outside the services without causing an ethical issue with the privacy? And one of the other issues which came from a workshop that we ran in 2019 is even more complicated. What happens if you have a female combatant who is monitored medically to make sure that she's okay, about to go into a combat zone, and the monitor (the AI) detects that she is pregnant. What do you do then? Do you inform the combatant—she may not know she's pregnant. What would you do with that data? Should you not keep? Should you throw that data away? What would you do with the combat? So, it's a huge area that you can come up with, and even with that non-combatant use, it's not in direct combatant like the use of a flying bomb. There's a lot of critical areas where you have to consider the application of ethics with the use of robotics and autonomous systems in the military and in conflict zones.

JASON Robots can certainly also be in a position to fight other robots. I don't think there are any ethics in that aspect in the sense of unmanned systems. And we'll need to be successful in that so that if there is a fight that requires our robotic systems to take out a first wave, then we have to make sure there aren't leakers. So, although there may be a robot-on-robot battle, that will be important to minimize the effect on people afterwards.

In terms of what risks we might perceive for humans involved in ethical decision-making in this kind of future. In one way, we might want to see the human in a sense as the hero, but there's an issue because those people that are the hero can also be the hazard. The machine can play a role in helping to ameliorate the hazard. So, if you like to maximize the chance of heroic acts and ethical acts and to minimize those that are not. Civilian deaths are an example of that. In the Middle East, area of operations, reports that are public, have expressed that in terms of civilian deaths, around half of those civilian deaths were collateral damage, but the

other half were simply not lawful targets. So, there's a role here for the machine and the system, not only to monitor but also to help intervene, to give pause for Commanders, to instil tactical patience; to identify symbols of protection on buildings, on people, on things. These are all part of what can be done.

It could be that there are times in that human as hero/hazard balance that the machine takes the initiative rather than the human or vice versa. That balance and that mix needs to probably be determined empirically. Furthermore, decision-making simply by people or only by machines at each end of the spectrum might be a vulnerability in a range of ways. So, I would see in the future that we would want to see decision-making more fully integrated between humans and machines in order for the strengths of one to offset the weaknesses of the other, be that a human strength in a machine weakness or vice versa, as the case may be.

There's also a question about humans and their role in this. I would view that potentially human has this role of Command, and that is the expression of intent to another, it includes authority, responsibility, competency and a role for humans and machines in control to realize that intent to express a capability for action and shared awareness in order to bind those together to act. However, the notion of that Command and Control being hierarchical in the future may not be the most efficient way for the ADF to operate. The segregation of assets, done through hierarchical Command and Control, could be made more efficient potentially in the future as we build greater trust in these systems to share and more dynamically by and components of control together to affect a greater sense of manoeuvre warfare in the way they operate.

STEPHEN The scenarios that we considered, we wanted to trigger scenarios where humans would trust the AI and reduce the cognitive workload of the human in essentially high-tempo scenarios where there might have been a number of objects moving in and out of collateral damage zones as well as things that a human might not be able to pick up. For instance, uniforms and small objects on a smartphone were another examples where we can pick up on AI because we get the full resolution of the video. And then, similarly, we wanted to look at scenarios where the human shouldn't trust the AI or the human has to overrule the AI because of various reasons. For instance, they might be using a protected symbol to try and avoid being targeted and the humans have got to justify a decision to override the AI or where the AI simply unsure about the decision and a human needs to make an ultimate decision on that matter.

The scenarios designed have been set up and have helped inform us in how we actually integrate the rules of engagement, laws of armed conflict, and ultimately the human factors aspect of integrating the AI with the human. Whether those are warning messages or whether we actually put the AI in the loop or the AI on the loop, the scenarios help inform that because in some instances, we simply want a warning and we want to carry out with the human because we might not want to biases. In other instances, where you may be breaching rules of engagement, we may have a requirement to actually justify through self-defense or loss of protection or any number of these reasons. But the scenarios helped inform that from a design perspective.

Some of the technology features we've added to reduce ethical risk is really at the training of the AI level, for instance. We trace every single neural network back to every single piece of data that was used that went into it, as well as every data it was validated on. And then we push that through out as a model, and we've got videos and plots. And then, we do performance based on the accuracy of the resulting network against the validation dataset. All of that for us is about establishing the level of trust in the AI, and you'll never have perfect hundred percent accuracy—anyone who says you can do that is lying. It's about the user and the person who makes the decision on whether to deploy this, actually understanding what level of trust they should have in it. And therefore, the fact that they do need to cross-check what's being provided to them.

The legal and ethical workshops were done almost from project inception. We initially engaged with a former military legal officer from the Australian Defence Force who deployed on operations. He gave us a number of scenarios where there were certainly difficult decisions that had to be made based on limited intelligence, as well as things that we had to consider from the standpoint of the Six Steps of Targeting. And that helped establish the initial framework for our user interface perspective. From there, we invited a variety of moral philosophers, ethicists, lawyers, human factors experts, and we put everyone together, and we considered what does AI supporting laws of armed conflict (LOAC) and rules of engagement look like and how do we implement that in from a coding standpoint?

What came out of that was about a 70 page legal and ethical framework based on a variety of ethical principles as well as the ADF's six steps of targeting and the laws of armed conflict being the Hague Convention and the Geneva Convention and how we would create that framework and put it into ultimately Athena Artificial Intelligence.

Ethical Risk Matrix [22:08]

Activity description	Ethical issue(s), principle(s)	Risk to the project objectives if ethical issue is not addressed	Actions/ Outcomes	Timeline	Person/s responsible	Status
Ensure operators act ethically under uncertainty	*Governance confidence*	*Operator misunderstands the accuracy and reliability of outputs or recommendations of the AI classifier*	*Experiments for implicit and explicit understanding of AI outputs by operator in ethical decision making*	*Q4 2020*	*Bob Cook*	*Pending*

VO Create an Ethical AI Risk Matrix with detail for each project activity:

- Define the activity you are undertaking
- Indicate the ethical facet and topic the activity is intended to address.
- Estimate the risk to the project objectives if issue is not addressed?
- Define specific actions you will undertake to support the activity
- Provide a timeline for the activity
- Define action and activity outcomes
- Identify the responsible party(ies)
- Provide the status of the activity.

KATE The Ethical AI Risk Matrix, this is a project management tool that you can use in order to identify ethical risks in your AI project, Assign those ethical risks to an individual who needs to take responsibility for mitigating those risks with a limited amount of time, but it has to have a schedule, and it has to have some obligation to resolve them. But it enables a project to keep track of these ethical risks and to have accountability for how they are managed when that technology does go into service or operations.

STEPHEN In regards to how we manage risk throughout the project we use an Ethical AI Risk Matrix. What that does is that allows us to look at what are the ethical risk for the project: What are the actions we need to do to mitigate those risks; who's responsible for it; and at what time are we going to do it? At every Milestone of the project, as well as periodically, we actually review these ethical risks and they've either been mitigated, they've been closed out through the design, or they're ongoing and we have to continue to monitor them. That's really the way that we manage those ethical risks for the product as it goes through high technology readiness levels.

Legal and Ethical Assurance Program Plan (LEAPP) [23:48]

GANOS The LEAPP is the abbreviation we've given for the Legal and Ethical Assurance Program Plan, which is quite a big mouthful. But it's something that we thought was important, particularly for those projects which are very complex, have a lot of AI in them, heavily enabled by AI, and possibly have some serious consequences as a result of the AI. So, it's the highest level, most complex, and most resource-intensive tool that we have, and so, you really only want to use it on major projects with artificial intelligence components.

What the LEAPP is, it's part of a whole suite of tools which are plans, which are used by the government, used by the Commonwealth to make sure that the provider of whatever service or software you're acquiring is going to meet our requirements in many ways. So, there's a software safety plan, (there are many safety management plans), human factors plans. In this case, we want to make sure that the Commonwealth gets what it wants, in what the contractor is going to deliver regarding the legal and ethical aspects of artificial intelligence it has purchased.

It comes in a form of a plan which you give to the contractor in the early phases of contract negotiation. There is normally a basic plan which you modify depending on the project that you have and the requirements of the government, but also the capability of the contractor. The contractor and the government will negotiate the outcome of the plan, which is a framework, which will tell the government, which will tell the Commonwealth, what the contractor is going to do regarding legal and ethical assurance within the program it's buying. And that could be a buying a platform, could be buying a software system, it could be buying a whole service of which earning a part out of this artificial intelligence, so there's quite a range of things that it has.

LAUREN The Legal-Ethical Assurance Program Plan is an iterative discussion between the developers and Defence to enable some of those legal and ethical issues to be contemplated and incorporated during the design and development phase. We know at the introduction to

service of Capabilities, we know at different gates along the way, there are certain legal compliance documents that need to be provided to support the application or to support the acquisition of those particular capabilities.

The purpose of the LEAPP is to have those considerations built-in to the design and development process. So, once the capability gets the introduction into the service point, those AI systems don't have to be unwound to have those considerations incorporated into them earlier on. For example, the requirement for Distinction is a general International Humanitarian Law (IHL) requirement that means that there is a legal obligation for the ADF, when they're using Force, to distinguish between combatants and civilians.
So, how does the Defense Force know when they're buying a RAS-AI capability that is purporting to be able to do that, to identify and then prosecute targets that it will be doing so to accord with that basic principle of International Humanitarian Law? There will be questions asked about things like, what's the pixelation of the particular feed that the system is relying on? Is there a presumption that anybody who can't be clearly identified would be deemed to be a civilian—which also accords with some of our legal obligations.

What indicia are being used to identify combatants, and is that specific to a particular conflict? And can that be adjusted to a particular enemy set depending on where the system is anticipated to be used? Or is it more general than that? And in those considerations that will also assist also in identifying how that system can be used when it is introduced into service. Because if some of those answers are that earliest age demonstrates, it doesn't have the fidelity to properly identify a combatant from a non-combatant, then the use case for that system is not going to be in a targeting context, it may only be for ISR to support other decision-making and intelligence tools.

GANOS: The main aim of the LEAPP is the contractor will provide to us, to the government, an explanation of what it's going to do. Once you agree on the plan, a plan is delivered to us. You check it as the project goes on to make sure the contractor is doing exactly what they promised they'd do and what you'd agreed upon. How does this plan interact with the other plans? The test and evaluation, how are you going to verify that it will occur? What sort of people are you going to involve? What sort of frameworks are you going to set up to check that the contractor, that the service provider, will meet our requirements?

How will the LEAPP help Defence? [28:19]

LAUREN: Defence should expect a good LEAPP will provide an opportunity to iteratively develop a capability so that when it is introduced into service it meets the legal and ethical needs of Defence when that capability is being used. It should reduce the time lag required to conduct that certification process, but also it should then enable Defence to have a better understanding of how that capability can be legally and ethically incorporated into its actual operational design. So that systems aren't being promised as a panacea to a particular problem when they actually can't deliver on what was being anticipated at the start point of delivery. But also so that there can be a concurrent building of trust in the assurance of that system, trust in how those decisions are being made. So when those systems are actually introduced into service, there's rigorous testing that can be demonstrated to show those people who are

relying on these systems for their protection and for decision-making in respect of the use of force that they'll be meeting their legal obligations, when using those particular systems.

GANOS The LEAPP will help Defence by giving Defence a clear understanding of the contractor's expectations of how much work and effort will go into the legal and ethical assurance. But also to ensure that Defence meets its own requirements, both from the legal framework, from the stakeholder framework, to make sure that whatever is being acquired is fit for purpose. But also the requirements placed upon us by our own governance but also our social responsibility. The LEAPP can be used for any sort of program, not just a platform or software. It could be software for human resources, for medical reasons, or for logistics. In any of those, we would like to have an understanding of what effort is going in to ensure that the legal and ethical components of the program and the artificial intelligence within the program have been dealt with.

What should AI developers know about the LEAPP? [30:32]

GANOS One of the most important parts for an AI developer is to understand what framework Defence is trying to follow up and to understand what our requirements are and the aspects which you are really focusing on. Each program, each project will have a different LEAPP because you're building a different service or technology, but also it will be used in different environments—this will change the legal and ethical issues associated with the service. Which means that the Contractor has to understand what Defence is looking for in the different areas and in the different contexts that we use the LEAPP for. The other part of the Contractor is to use the LEAPP for is to articulate to the Defence what they're thinking. Is it a large program? Is it a small program? Does it cover many other aspects, or are other aspects captured in other areas in the project, in safety or software management or human factors?

What should Defence personnel know about the LEAPP? [31:33]

LAUREN Defence Personnel should think about who to talk to when they're developing a LEAPP with industry, universities, and within Defence capability about what their particular area of expertise is going to be to assist in understanding the legal and ethical issues associated with that particular Capability. It doesn't necessarily just mean lawyers. I mean, I should say it just means lawyers because lawyers write themselves into everything. We were at ourselves into the Geneva Conventions, so we were definitely there, but in respect of engaging those people who are doing the design, engaging the experts who can articulate clearly, how is the algorithm being developed? What does that algorithm then mean when it's having certain data inputs put into it? Is there bias in that data that needs to be addressed or considered? The process itself needs to be iterative, and it really does need to be a discussion between the Capability Project Industry, and Defence with the use of SMEs like lawyers, ethicists, and those designers, researchers, and experts in AI that can translate effectively the geek-to-grunt is probably a great way to describe it, with those lawyers and ethicists shoved in on the side to make sure that information actually aligns with the Defence's needs. But the industry and the project are obviously going to have business efficacy, and I'm not going to

waste their time and money in developing a system that's not ultimately going to be picked up by Defence.

GANOS Defence personnel have to understand what their requirements are. They have to fully understand the system which they're procuring, and the issues associated with that. So from the LEAPP, they should be able to understand where the issues might be, how they're being addressed by the Contractor, and then how it's going to be verified. All part of the plan is to make sure that yes, the work has been done, but we must be able to verify the activities within the LEAPP, which we find important, particularly in the test and evaluation and the verification.

Managing ethical risks in AI for Defence [33:31]

STEPHEN we don't want ~~it~~ to get a product to an acquisition standpoint. And then, we go through a weapons review, where our AI combines with a weapon system to form a means of warfare and we find that our product was not defined effectively from day one to be used in an operational scenario because that would be a waste of effort, and we wouldn't end up producing the Capability that we set out to do. By considering it from day one, we can actually set up a framework that will meet the requirements of a weapons review upon the time of acquisition and we can ensure that the hard work we're doing now will result in a Capability when it comes time to deploy this in theatre. When it comes to RAS AI in Defence and ethics associated with them, I think it's very important to consider how has a given company or a given AI supplier establishing trust in that RAS AI product and understanding that up skilling yourself in how do we establish trust in these products, how do we provide assurance, what sort of testing was done to those products to establish that level of assurance? And ultimately, that should be the most important thing before we start giving this to your soldiers, seamen, or aircrew.

JASON I'd also recommend you talk about ethics with your colleagues. Talk about it with the general public. The general public always has a view on this and they're really worth listening to because that's who we in Defence and outside of Defence are there to represent. Have an opinion, build your opinion, and build your character in the process.

MICK I'd say to people, educate yourself. We are wonderful at training and educating people and Military institutions, but you cannot spend your life in the schoolhouse. The vast majority of your time in the Australian military is going to be spent in a unit and on operations or on task somewhere outside of the schoolhouse. You need to dedicate yourself to learning about these systems. Do some professional reading, maybe do online learning courses. But you need to understand these technologies and how they impact on individuals and on organisations. Technological literacy by soldiers, sailors, and aviators and by our most strategic leaders is a key capability for our organization moving forward and it will underpin not just how we do things on the battlefield or in a humanitarian assistance operation but will underpin strategy making and policy-making by our strategic leaders.

ANNEX A Contexts of AI in Defence drawn from Defence Science & Technology Group (DSTG) Technical Report DSTG-TR-3786 'A Method for Ethical AI in Defence', Appendix E https://www.dst.defence.gov.au/publication/ethical-ai

	Force Application (FA)
Description	The conduct of military missions to achieve decisive effects through kinetic and non-kinetic offensive means.
AI examples	Autonomous weapons (AWs) and autonomous/semi-autonomous combat vehicles and subsystems
	AI used to support strategic, operational and tactical planning, including optimisation and deployment of major systems
	AI used in modelling and simulation used for planning and mission rehearsal
	AI used in support of the targeting cycle including for collateral damage estimation
	AI used for Information Warfare such as a Generative Adversarial Network (GAN-) generated announcement or strategic communication
	AI used to identify potential vulnerabilities in an adversary force to attack
	AI used for discrimination of combatants and non-combatants

	Force Protection (FP)
Description	All measures to counter threats and hazards to, and to minimise vulnerabilities of, the joint force in order to preserve freedom of action and operational effectiveness
AI examples	Autonomous defensive systems (i.e. Close in Weapons Systems)
	AI used for Cyber Network Defence
	AI used to develop and employ camouflage and defensive deception systems and techniques
	Autonomous decoys and physical, electro-optic or radio frequency countermeasures
	AI to identify potential vulnerabilities in a friendly force that requires protection
	AI used to simulate potential threats for modelling and simulation or rehearsal activities
	Autonomous Medical Evacuation/Joint Personnel Recovery systems

	Force Sustainment (FS)
Description	Activities conducted to sustain fielded forces, and to establish and maintain expeditionary bases. Force sustainment includes the provision of personnel, logistic and any other form of support required to maintain and prolong operations until accomplishment of the mission.

AI examples	Autonomous combat logistics and resupply vehicles
	Automated combat inventory management
	Predictive algorithms for the expenditure of resources such as fuel, spares and munitions
	Medical AI systems used in combat environments and expeditionary bases
	Predictive algorithms for casualty rates for personnel and equipment
	Algorithms to optimise supply chains and the recovery, repair and maintenance of equipment
	Algorithms to support the provision of information on climate, environment and topography
	AI used for battle damage repair and front-line maintenance

Situational Understanding (SU)

Description	The accurate interpretation of a situation and the likely actions of groups and individuals within it. Situational Understanding enables timely and accurate decision making.
AI examples	AI that enables or supports Intelligence, Surveillance and Reconnaissance (ISR) activities including:
	object recognition and categorisation of still and full motion video
	removal of unwanted sensor data
	identification of enemy deception activities
	anomaly detection and alerts
	monitoring of social media and other open-source media channels
	optimisation of collection assets
	AI that fuses data and disseminates intelligence to strategic, operational and tactical decision makers
	Decision support tools
	Battle Management Systems
	AI that supports Command and Control functions
	Algorithms used to predict likely actions of groups and individuals
	AI used to assess individual and collective behaviour and attitudes

Personnel (PR)

Description	All activities that support the Raising, Training and Sustaining (RTS) of personnel.

AI examples	AI used for **Human Resource Management** including:
	record keeping
	posting and promotion
	disciplinary and performance management
	recruitment and retention
	modelling of future personnel requirements
	prediction of HR supply and demand events and anomalies
	AI used in individual and collective **training and education** including modelling and simulation
	AI used for testing and certification of personnel
	AI used to model the capability and preparedness of permanent and reserve personnel

Enterprise Logistics (EL)

Description	Activities that support rear-echelon enterprise-level logistics functions including support of permanent military facilities
AI examples	Autonomous rear-echelon supply vehicles and warehouses
	AI used for optimisation of rear-echelon supply chains and inventory management
	AI used in depot-level and intermediate maintenance, including:
	Digital twinning
	Predictive maintenance
	Global supply chain analysis, prediction and optimisation
	Enterprise-level analysis and prediction for resource demand and supply (i.e. national/strategic fuel requirements)
	AI used in the day-to-day operation of permanent military facilities

Business Process Improvement (BP)

Description	Activities that support rear-echelon administrative business processes that are not related to personnel or logistics.
AI examples	AI used for Information Management and record-keeping
	Informational assistants such as policy chatbots
	AI that supports management of policy and procedures
	AI used to optimise business and administrative processes, including modelling and simulation tools

ANNEX B A taxonomy of decision problems

Content contributed by Tristan Perez to Defence Science & Technology Group (DSTG) Technical Report DSTG-TR-3786 'A Method for Ethical AI in Defence' https://www.dst.defence.gov.au/publication/ethical-ai , see also French, S., Maule, J., & Papamichail, N. (2009). Decision Behaviour, Analysis and Support. Cambridge: Cambridge University Press

Decision-maker/s	**Type of Decision**	**Example/s**
Single decision-maker	*Single-stage* once-off decisions	A decision as to whether continue with current mission objectives or consider alternatives given changes in the operational conditions.
		A decision about deploying a particular type of weapon towards a hostile asset
	Multi-stage sequential decisions in time	Management of a supply chain to support a replenishment of supplies for a mission over number of days or months
		Motion control of a network of autonomous systems to deliver un-interruptible communications for C2
		Missile guidance towards a fixed target
Multi-decision maker	*Decisions under conflict* Games	Once-off games, e.g.
	Cooperative vs. non-cooperative	Two governments negotiating over a contested land or sea area
		Sequential games, e.g.
	iterated vs. non-iterated	Two aircraft/marine craft in a pursue and evade situation
	Zero sum vs non-zero sum	Multiple autonomous systems avoiding collisions while seeking to attain individual mission goals
	Two vs N players	Managing a network of military assets during engagement
	Consensus decisions social choice	A resolution of the UN Security Council
		A number of countries developing guidelines for the conduct of trials of autonomous systems at the International Maritime Organisation Meeting
		A group of manned assets and group of AS deciding how to engage with a hostile asset
		A jury deciding for guilt or innocence

Prime minister and council decision to escalate war

DEFENCE SCIENCE & TECHNOLOGY GROUP, 'FACETS OF ETHICAL AI FOR DEFENCE' (2021)	US DEPARTMENT OF DEFENSE ETHICAL AI PRINCIPLES (2020)	AI ETHICS PRINCIPLES AUSTRALIAN GOVERNMENT (DEPARTMENT OF INDUSTRY INNOVATION AND SCIENCE, 2019)	IEEE ETHICALLY ALIGNED DESIGN PRINCIPLES (IEEE GLOBAL INITIATIVE ON ETHICS OF AUTONOMOUS AND INTELLIGENT SYSTEMS, 2019)	PRINCIPLED ARTIFICIAL INTELLIGENCE: A MAP OF ETHICAL AND RIGHTS BASED APPROACHES (FJELD ET AL., 2019)	THE GLOBAL LANDSCAPE OF AI ETHICS GUIDELINES (JOBIN ET AL., 2019).
Responsibility *Who is responsible for AI?*	Responsible	Human, social and environmental wellbeing Human-centred values	Human rights Wellbeing	Promotion of human values Professional responsibility	Responsibility
Governance *How is AI controlled?*	Governable	Transparency and explainability	Transparency	Human control of Technology Transparency Explainability	
Trust *How can AI be trusted?*	Reliable Equitable	Reliability & Safety Privacy protection Security Fairness Contestability	Effectiveness Competence Data agency Awareness of misuse	Safety and Security Privacy Fairness and non-discrimination	Privacy Justice and fairness Non-maleficence
Law *How can AI be used lawfully?*	-	-	-	-	-
Traceability *How are the actions of AI recorded?*	Traceable	Accountability	Accountability	Accountability	

Further Reading