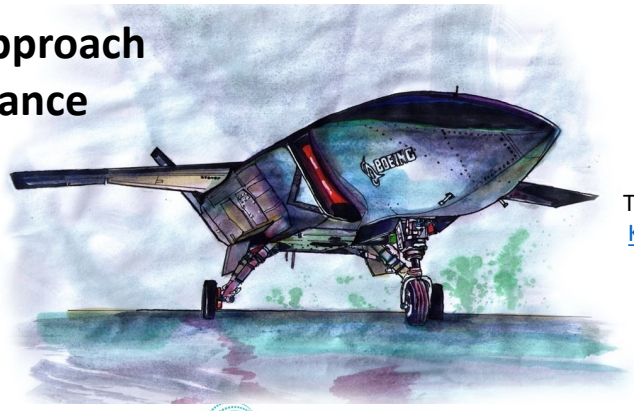
TRUSTED AUTONOMOUS SYSTEMS
DEFENCE CRC

Australia's approach to AI Governance for Defence



Dr S.Kate Devitt
Chief Scientist
Trusted Autonomous Systems
Kate.devitt@tasdcrc.com.au

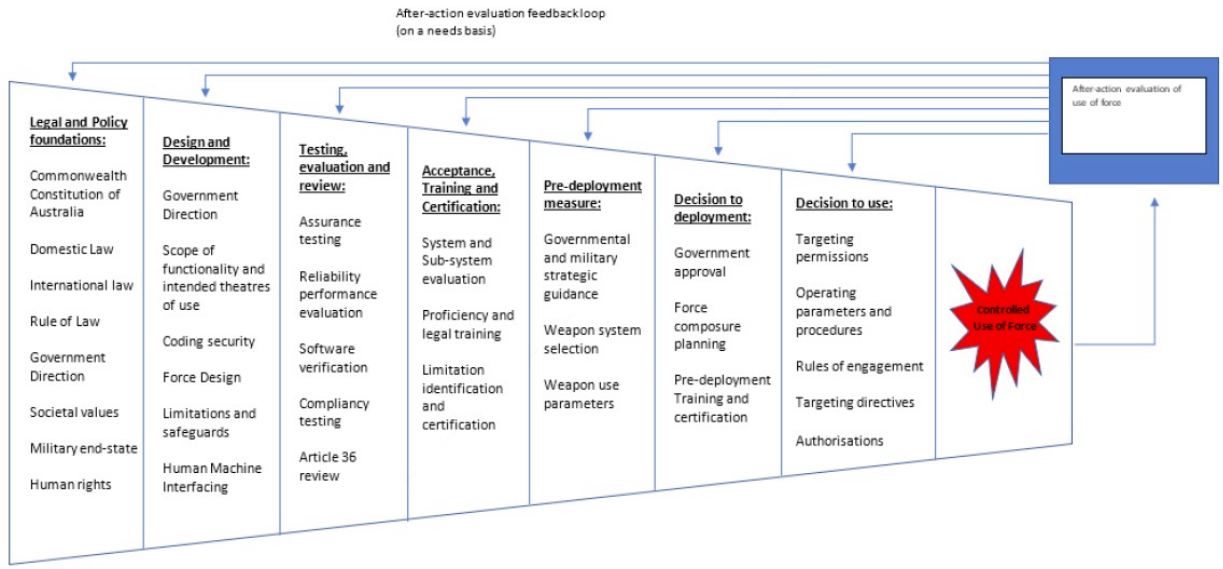
Australian Government
Department of Defence

NEXT GENERATION TECHNOLOGIES FUND

Queensland Government

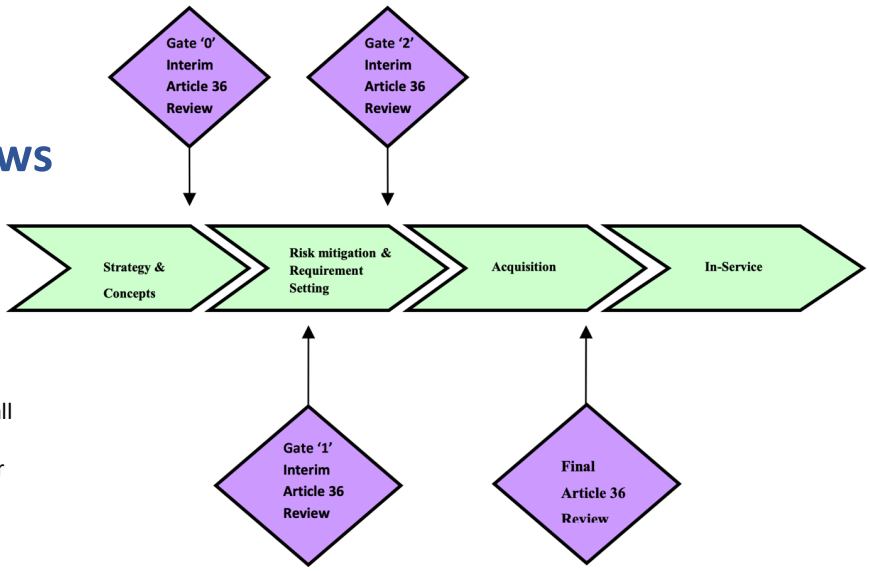
System of Control

Australia's System of Control and applications for Autonomous Weapon Systems (2019)
https://meetings.unoda.org/section/ccw-gge-2019-documents/



Australia's commitment to Article 36 Reviews

Australia's Chief of the Defence Force (CDF) has mandated that the actions of the Australian Defence Force (ADF) with respect to the development and procurement of weapons and their intended use in armed conflict are to be consistent with Australia's obligations under all applicable treaties and customary international law (CIL), in particular under the law of armed conflict (LOAC).



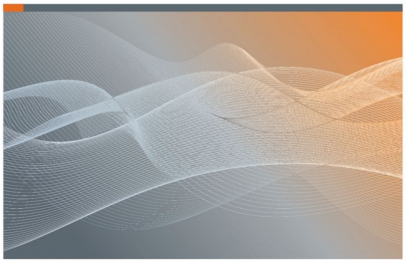
See also 'The Australian Article 36 Review Process' (2018) https://meetings.unoda.org/section/ccw-gge-2018-documents/

TRUSTED AUTONOMOUS SYSTEMS

OFFICIAL

Australian Government
Department of Defence

A Method for Ethical AI in Defence



Defence Science and Technology Group

DSTG-TR-3786

This is a report on the outcomes of a workshop only and does not represent an official position of Defence. It represents views expressed by participants and stakeholders of the workshop.

Defending Australia and its National Interests
www.defence.gov.au

OFFICIAL

Contexts for Ethical AI in Defence

- *Method for Ethical AI in Defence* recommends a risk-based methodology for ensuring Legal and Ethical compliance
- Risk assessment based on:
  - Nature of AI and training data
  - Function
  - Situation
  - Assurance testing
- Risk assessment already a mature concept in ADF acquisition and operations

AI Ethics Checklist

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

AI Ethics Checklist

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

Contexts for AI in Defence: Functions

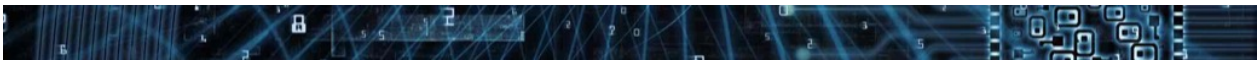
How do you define the Defence contexts that AI will be used in?

ADF Warfighting Functions
Command
Force Application
Force Protection
Situational Understanding
Force Projection
Force Generation and Sustainment

Method for Ethical AI in Defence Contexts
Force Application
Force Protection
Situational Understanding
Force Sustainment
Personnel
Enterprise Logistics
Business Process Improvement

AI Ethics Checklist

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



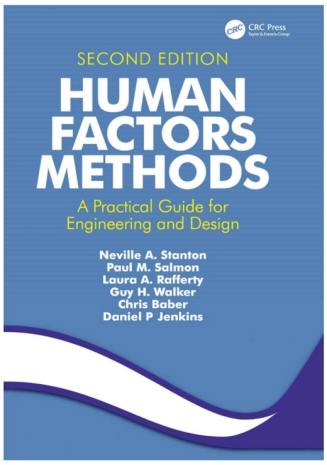
Decision-maker/s	Type of Decision	Example/s
Single decision-maker	Single-stage once-off decisions	A decision as to whether continue with current mission objectives or consider alternatives given changes in the operational conditions.
	Multi-stage sequential decisions in time	Management of a supply chain 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 Cooperative vs. non-cooperative iterated vs. non-iterated Zero sum vs non-zero sum Two vs N players	Once-off games, e.g. Two governments negotiating over a contested land or sea area Sequential games, e.g. Two aircraft/marine craft in a pursue and evade situation Multiple autonomous systems avoiding collisions while seeking to attain individual mission goals 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

TRUSTED AUTONOMOUS SYSTEMS

AI Ethics Checklist

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

Critical Decision Method



Goal specification.	What were your specific goals at the various decision points?
Cue identification.	What features were you looking for when you formulated your decision? How did you what you needed to in order to make the decision? How did you know when to make the decision?
Expectancy.	Were you expecting to make this sort of decision during the course of the event? Describe how this affected your decision-making process
Conceptual.	Are there any situations in which your decision would have turned out differently? Describe the nature of these situations and the characteristics that would have changed the outcome of your decision
Influence of uncertainty.	At any stage, were you uncertain about either the reliability or the relevance of the information that you had available? At any stage, were you uncertain about the appropriateness of the decision?
Information integration.	What was the most important piece of information that you used to formulate the decision?
Situation Awareness.	What information did you have available to you at the time of the decision?
Situation assessment.	Did you use all of the information available to you when formulating the decision? Was there any additional information that you might have used to assist in the formulation of the decision?
Options.	Were there any other alternatives available to you other than the decision you made?
Decision blocking – stress.	Was there any stage during the decision-making process in which you found it difficult to process and integrate the information available? Describe precisely the nature of the situation
Basis of choice.	Do you think that you could develop a rule, based on your experience, which could assist another person to make the same decision successfully? Why/why not?
Analogy/generalisation.	Were you at any time reminded of previous experiences in which a similar decision was made? Were you at any time reminded of previous experiences in which a different decision was made?

Stanton, N., Salmon, P. M., & Rafferty, L. A. (2013). Cognitive Task Analysis Methods.
Human factors methods: a practical guide for engineering and design. Ashgate Publishing, Ltd.

TRUSTED AUTONOMOUS SYSTEMS

AI Ethics Checklist

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

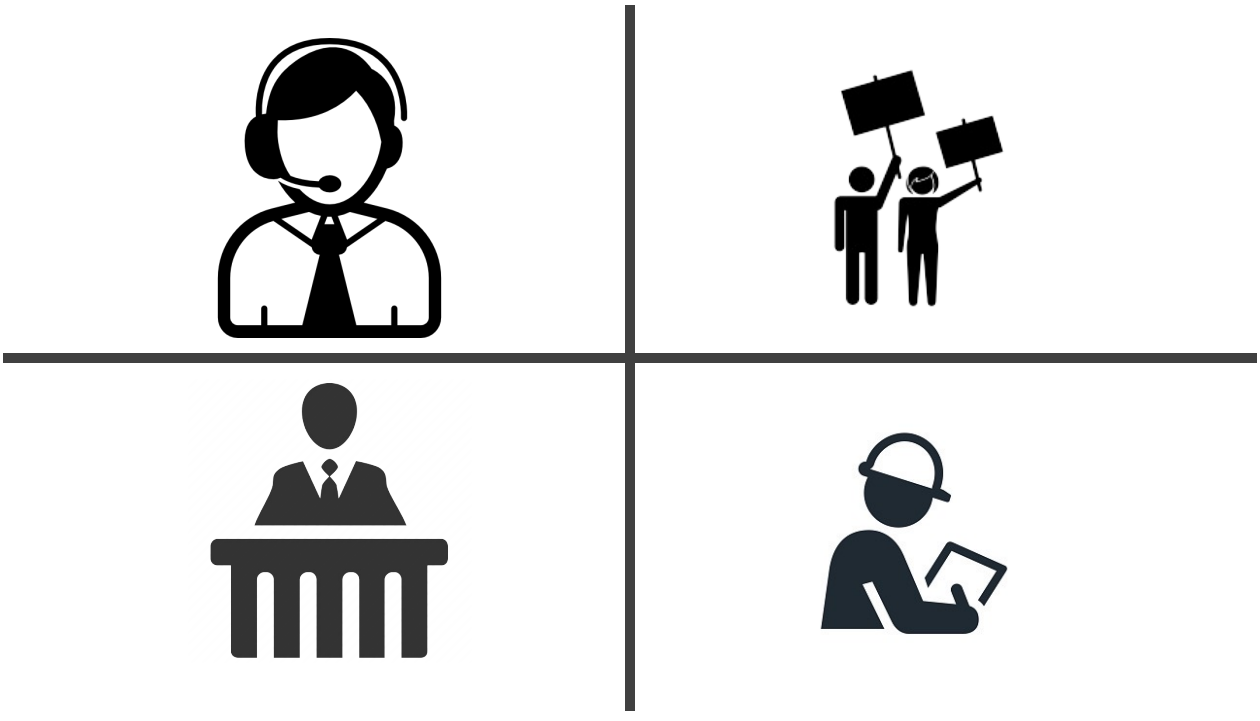
TRUSTED AUTONOMOUS SYSTEMS

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

TRUSTED AUTONOMOUS SYSTEMS

AI Ethics Checklist

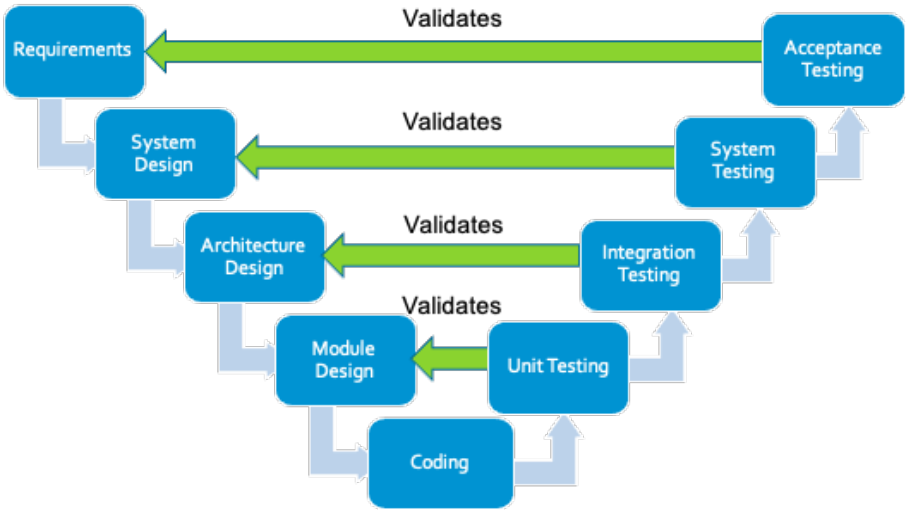
A	Describe the military context the AI is for	E.g. Force Application, Force Protection, Force Sustainment, Situational Understanding, Personnel, Enterprise Logistics, Business Process Improv.
B	Explain the sort of decisions AI helps with	E.g. Is it a single decision-maker, multi-decision maker; once-off 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



TRUSTED AUTONOMOUS SYSTEMS

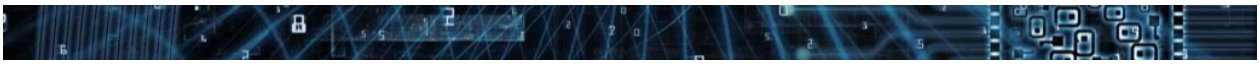
AI Ethics Checklist

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



AI Ethics Risk Matrix

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*



Legal & Ethical Assurance Program Plan (LEAPP)

DATA ITEM DESCRIPTION

1. DID NUMBER: DID-ENG-SW-LEAPP

2. TITLE: LEGAL AND ETHICAL ASSURANCE PROGRAM PLAN FOR ARTIFICIAL INTELLIGENCE SYSTEMS

3. DESCRIPTION AND INTENDED USE

3.1 The Legal and Ethical Assurance Program Plan (LEAPP) describes the Contractor's plan for assuring that Software acquired under the contract that is categorised as Artificial Intelligence (AI) meets the Commonwealth's Legal and Ethical Assurance (LEA) requirements.

3.2 For Contractors acquiring and/or supplying Software classified as AI under the Contract, the LEAPP is expected to describe the approach, plans and procedures to be applied to the management of the AI Software being acquired and/or supplied. This would typically include the monitoring and review of Subcontractors developing AI Software, the Configuration Management of acquired AI Software, and the integration and Verification of this AI Software with other elements being supplied under the Contract.

3.3 The Commonwealth uses the LEAPP:

a. to provide visibility into the Contractor's technical planning;
b. for progress and risk assessment purposes; and
c. to provide input into the Commonwealth's own planning.

4. INTER-RELATIONSHIPS

4.1 The LEAPP is subordinate to the following data items, where these data items are required under the Contract:

a. Software Management Plan (SMP);
b. Integrated Support Plan (ISP);
c. Configuration Management Plan (CMP); and
d. Verification and Validation Plan (V&VP).

See Appendix G. DID-ENG-SW-LEAPP https://www.dst.defence.gov.au/publication/ethical-ai

Summary: Reducing ethical risks of RAS-AI

A method for Ethical AI in Defence

...while not a formally adopted view of the Australian government, the Method is the clearest articulation of ethical AI for defence among the Indo-Pacific allies as well one of the most concrete practices that U.S. allies have thus far developed for AI ethics implementation in defence (Lockman, 2021, pp.21 & 23)

Key references:

Devitt, S. K., Gan, M., Scholz, J., & Bolia, R. S. (2021). A Method for Ethical AI in Defence. **Defence Science & Technology Group** (DSTG-TR-3786). https://www.dst.defence.gov.au/publication/ethical-ai

Roberson, T., Bornstein, S., Liivoja, R., Ng, S., Scholz, J. & Devitt, S.K. (2022). A Method for Ethical AI in Defence: A case study on developing trustworthy autonomous systems. [under peer review] *Journal of Responsible Technology*.

Gaetjens, D., Devitt, S.K. & Shanahan, C. (2021). Ethical AI in Defence Case Study: Allied Impact. DST Technical Report. **Defence Science & Technology Group**

Lockman, Z. (2021). Responsible and Ethical Military AI Allies and Allied Perspectives: CSET Issue Brief. **Centre for Security and Emerging Technology, Georgetown University's Walsh School of Foreign Service**. https://cset.georgetown.edu/wp-content/uploads/CSET-Responsible-and-Ethical-Military-AI.pdf

Copeland, D., & Sanders, L. (2021, 8 October). Engaging with the industry: integrating IHL into new technologies in urban warfare. **Humanitarian Law and Policy ICRC Blog**. https://blogs.icrc.org/law-and-policy/2021/10/07/industry-ihl-new-technologies/

TRUSTED
AUTONOMOUS
SYSTEMS

info@tasdcrc.com.au

www.tasdcrc.com.au

@TASDCRC

TASDCRC

Australian Government
Department of Defence

NEXT GENERATION
TECHNOLOGIES FUND

Queensland
Government