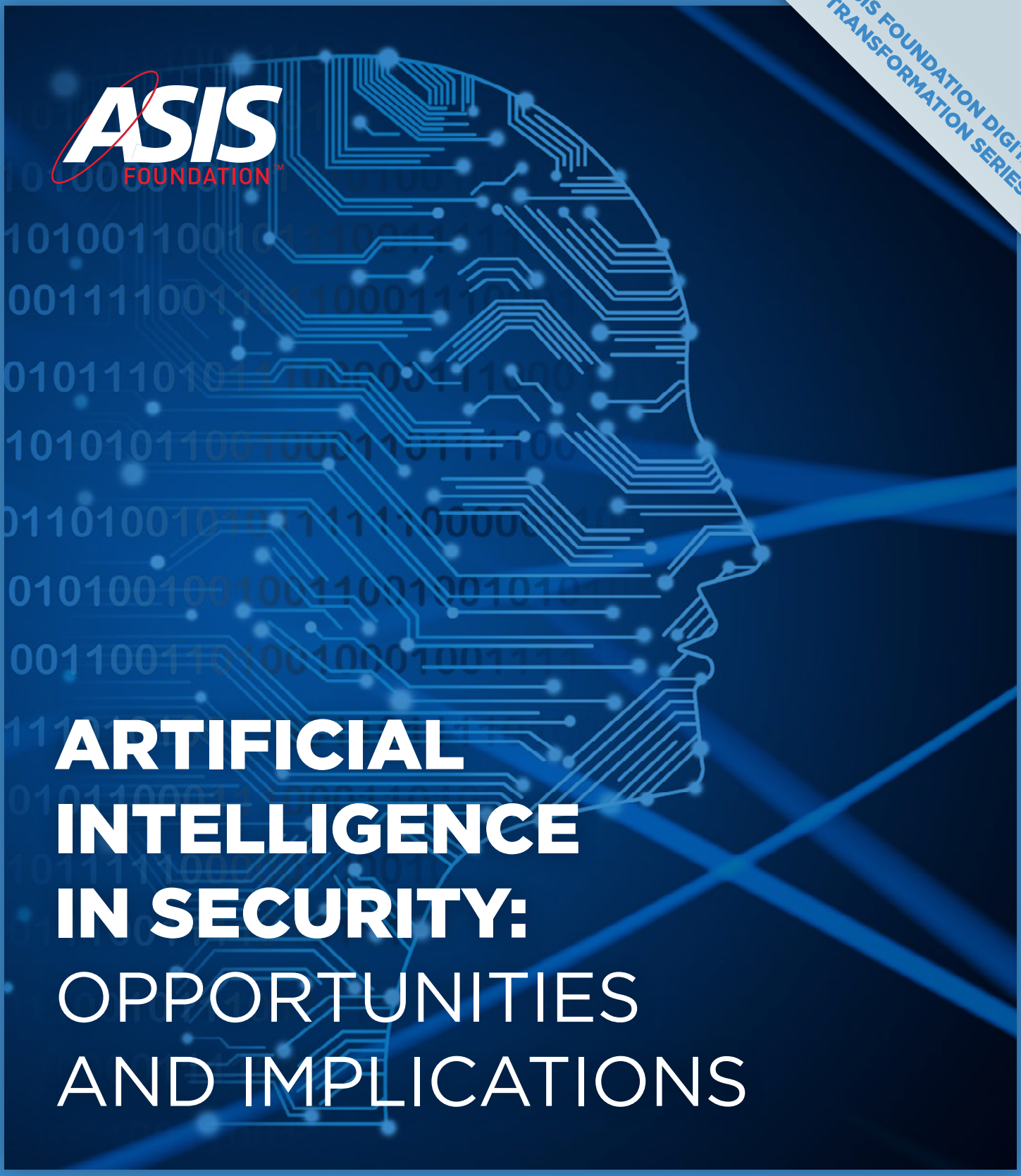
ASIS FOUNDATION DIGITAL TRANSFORMATION SERIES

ASIS FOUNDATION™



ARTIFICIAL INTELLIGENCE IN SECURITY: OPPORTUNITIES AND IMPLICATIONS

ADOPTION GUIDANCE DOCUMENT

Artificial Intelligence and Security Technologies Adoption Guidance Document

Opportunities and Implications of using Artificial Intelligence in the Establishment of Secure Physical Environments

Dr. Michael Coole

Mrs Deborah Evans

Mrs Jennifer Medbury

May 2021

Dr Michael Coole: Michael is a senior lecturer and researcher at Edith Cowan University. He has 25 years' experience in the security, crime prevention and emergency management fields, and has worked in the Australian Defence Force, Western Australia's Department of Justice and as a private consultant. Michael researches and teaches across the broad spectrum of security and crime prevention problem domains.

Mrs Deborah Evans: Deborah has worked in the international security environment within the maritime domain, and is now a sessional academic and postgraduate student at Edith Cowan University. Her professional and research interests include a variety of security areas including biosecurity, biodefence, civil and military Artificial Intelligence and Autonomous Weapons Systems, in addition to exploring Unmanned Aircraft System (UAS) applications through her private businesses in Western Australia.

Mrs Jennifer Medbury: Jennifer lecturers and researches in intelligence and terrorism studies at Edith Cowan University. She has over 11 years' experience as an intelligence analyst and senior intelligence analyst with the Australian Defence Intelligence Organisation and the Western Australia Police Force.

Opportunities and Implications of using Artificial Intelligence in the Establishment of Secure Physical Environments

Guidance Document

Introduction

The adoption of contemporary Artificial Intelligence (AI) technologies can be challenging to navigate amidst the rapid development of commercial products, marketing information, technical terms, and specifications. Therefore, the purpose of this guidance document is to provide security professionals with a research supported understanding of how security technologies use AI. The document aims to support security professionals in their AI adoption and decision making, providing a framework of factors to be considered for assessing the suitability of AI-based security technologies for the protection of their assets.

Consequently, the document provides an explanation of AI, and highlights how AI is currently used throughout the technical cycle of operation within some security technologies, and across the AI types, domains and spectrum of AI paradigms. The document highlights future opportunities and risks associated with the adoption, or absence, of AI. This analysis is supported by a risk factor check list as an aide memoire to help security professionals consider ways to assess their opportunities and risks within the context of AI enhanced security technologies.

The articulation of AI's current use in security technologies, the opportunities for further adoption, potential risks and risk factor check list are not exhaustive, rather they represent a summary of a larger research report: *Opportunities and Implications of using Artificial Intelligence in the Establishment of Secure Physical Environments – Artificial Intelligence Research Report*. This document therefore presents a summary to help security professionals reflect on what AI is, its true functional capabilities, and objectively consider the benefits and risks associated with the adoption of AI enhanced security technologies, ensuring such risks are accounted for in organisational risk registers.

What is Artificial Intelligence?

Artificial Intelligence (AI) may be broadly defined as a branch of computer science that investigates and develops computational approaches and techniques that allow machines to perform tasks that would normally require some level of human intelligence. In other words, making machines intelligent.

Artificial Intelligence in Security Technologies

Many contemporary technologies, including security systems and devices, use AI algorithms to enhance the capabilities of those technologies. An algorithm is simply an instruction, or set of instructions, which a computer or system will follow to perform a task. Security technologies can use AI algorithms to carry out a number of tasks, such as the identification of patterns and signals (such as the acoustic signals created by gunshots), to detect anomalies in patterns of behaviour (such as behavioural analysis in surveillance systems), to classify and match images (such as using computer vision to differentiate between a person or an animal) or to detect and identify images or materials (such as contraband or compounds in X-ray scanners).

The use of AI in security technologies can provide significant benefits for operational security, such as increasing the probability and speed of detection, reducing operator workload and fatigue, as well as helping to focus the attention of security personnel to where it is most needed. At a management level, AI may reduce costs, direct the allocation of resources, support decision making, and even present early intervention opportunities to mitigate insider threats.

Many security technologies use a basic level of AI to achieve a specific task. In some cases, technologies which use AI to carry out a number of specific tasks simultaneously, may appear to achieve a higher level of intelligence. However, this is not always indicative of a higher level of AI - intelligence levels tend to increase by the complexity and integration of decision making, rather than the number of specific tasks a system or device may perform.

Section 1 - Defining and Explaining Artificial Intelligence

Types of Artificial Intelligence

AI can be divided into four types based on the capabilities of the computer or device, relative to human intelligence (Goertzel & Pennachin, 2017[1]; IBM Services, 2018[2]; O'Carroll, 2020[3]).

Artificial Narrow Intelligence (ANI)
Also referred to as 'Weak AI' or 'Narrow AI' – approaches that focus on solving very specific tasks within the scope for which they have been designed. Narrow AI is very good at completing repetitive tasks and in many instances performs much better than humans. Examples include Siri, Google Translate and IBM's Watson.

Artificial Broad Intelligence (ABI)
Also referred to as 'Broad AI' - is described as the integration of two or more narrow AI systems or techniques that make decisions to perform a task or process. Enterprises may use data specific to their business to train systems to address the specific business process, for example self-driving vehicles, analysis of investment strategies for corporate customers in a banking system, or a software system supporting maintenance work on an oil rig.

Artificial General Intelligence (AGI)
Also referred to as 'Strong AI' or 'Deep AI' – approaches that allow machines to perform intellectual tasks at the same level as a human. General AI is expected to possess *theory of mind* as well as being *self-aware*, able to understand *belief, thoughts, emotions and expectations of people* and able to *interact socially*. Like humans, General AI can reason, be able to strategise and make plans based on emotions and prior knowledge. Although General AI theoretically possesses self-awareness, it lacks emotion. However, such advances are yet to be achieved in the current state of AI research and development.

Artificial Super-Intelligence (ASI)
Approaches that hypothetically possesses ability and intelligence that surpasses that of humans.

Artificial Intelligence Paradigms

This project drew on Corea's (2019[4]) work, which introduced an architecture to communicate and explain the intersection between the problem domains and *Paradigms* of AI, which comprise the approaches (tools and methods) used by AI researchers and computer scientists to understand and develop the algorithms and methodologies used to solve problems within intelligent systems and devices. According to Corea there are six AI paradigms across three different macro-approaches used to address and solve problem domains such as perception, reasoning, knowledge, planning, and communication. These six paradigms include Logic-based, Knowledge-based, Probabilistic-methods, Machine learning, Embodied intelligence and Search Optimization falling into three macro-approaches including *Symbolic*, *Statistical* and *Sub-symbolic*.

---

[1] Goertzel, B., & Pennachin, C. (2007). *Artificial general intelligence* (Ser. Cognitive Technologies). Springer. https://doi.org/10.1007/978-3-540-68677-4

[2] IBM Services. (2018). *Beyond the hype: A guide to understanding and successfully implementing artificial intelligence within your business*. https://www.ibm.com/downloads/cas/8ZDXNKQ4

[3] O'Carroll, B. (2020, January 31). What are the 3 types of AI? A guide to narrow, general, and super artificial intelligence. *Codebots*. https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible

[4] Corea, F. (2019). *An Introduction to Data: Everything You Need to Know About AI, Big Data and Data Science (Volume 50 Studies in Big Data)*. Springer Nature Switzerland AG 2019.

Symbolic Paradigms

Symbolic paradigms include the various AI methods that use symbols, which may be represented in rules (Logic or Knowledge), as their basis of computation. The internal working of such an AI process receives its input - and based on this input satisfying a set of pre-defined rules - produces a defined output.

Logic-based AI approaches are comprised of representing knowledge of an agent's world, its goals and the current state via logical statements. By using inference or deduction involving these logical statements, an appropriate computational decision to achieve specific goals is obtained. Areas of logic-based approaches include knowledge representation, various forms of reasoning (nonmonotonic, abductive and inductive) and computational logic.

Knowledge-based approaches consist of two salient components, a knowledge-based component and an inference engine that acts as the control component for inferring new knowledge/decisions. The knowledge base contains information about the state of the world, and this could be represented in different forms: declarative; procedural; heuristic; structural or meta knowledge involving ontologies and huge databases. The second component, the inference engine includes techniques such as rule-based, model-based and case-based reasoning for inferring new knowledge/decisions.

Statistical Paradigms

Statistical Paradigms apply a series of mathematical operators to their inputs to produce a defined output. For example, a traditional computer vision AI uses the set of pixels belonging to an image as input and applies a set of operations based on the spatial location and colour of pixels to group sets of pixels belonging to distinct objects. Measures taken from these groups are used to determine what the objects are. Statistical paradigms generally use probabilistic and machine learning techniques.

Probabilistic approaches employ probabilistic representations, that capture uncertainty in complex relationships and knowledge of the world, in the form of probabilistic graphical models. The graphical models capture the distribution in the data and decisions can be obtained via statistical inference techniques.

Machine Learning techniques automatically build models from input data which can be used to make predictions or decisions. Three salient classes of machine learning techniques include: unsupervised, supervised and reinforcement learning. In some cases, semi-supervised learning has been identified as a fourth class of machine learning. Unsupervised learning techniques find patterns from input data, without requiring these to be labelled. Supervised and reinforcement learning require labelled input data. Supervised learning includes classification techniques for determining which category an item belongs to and numerical regression which generates a function, which first captures the relationship between inputs and outputs, and subsequently then uses this data to predict how outputs will change as a function of the inputs. In reinforcement learning the aim is to reward a learner agent for good responses and to penalise for poor responses, ultimately allowing the agent to learn an operational strategy for its problem domain.

Sub-Symbolic Paradigms

Sub-symbolic paradigms are representative of neurons in the human brain. Sub-symbolic architectures can learn autonomously, following training and development of the neural network architecture. Corea (2019) defines the sub symbolic paradigm as one where "no specific representations of knowledge should be provided ex-ante", meaning knowledge representations are not provided 'before the event'. Therefore, the concepts of affective computing, autonomous systems, distributed artificial intelligence, ambient computing and evolutionary algorithms fall under the sub-symbolic paradigm. If taking a systems-based view of AI architectural paradigms, only the distributed intelligence paradigm and some applications of evolutionary algorithms qualify, as others are typically applications of AI approaches within the symbolic or statistical paradigms.

Embodied intelligence approaches take into consideration *situatedness* and *embodiment* in the design of intelligent behaviour in embodied and situated agents. *Situatedness* is the coupling between the agent and its environment and *embodiment*, refers to the constraints associated an agent's body, perceptual and motor system. The study of embodied intelligence has been associated with early development in bio-

inspired computational intelligence techniques in robotics, where the focus is on morphological computation and sensory-motor coordination in robotic models.

Search and Optimization approaches are techniques that can search a complex and ill-defined search space intelligently and efficiently.

Levels of Artificial Intelligence

When evaluating the 'level of intelligence' a system has, it is essential to understand the degree of involvement, decision making and control that humans retain - or intelligent systems possess - while these systems are in operation. This concept refers to the *level of autonomy* a system may have, which may vary considerably between types and categories of security technologies.

In understanding levels of intelligence, key distinctions are drawn between *Machine Autonomy* and *Intelligent Autonomy*. Numerous commercially available systems and machines maintain they are 'fully autonomous', operating without human assistance to achieve specific tasks. These systems typically possess *Machine Autonomy* and examples of this type of autonomy include food production and processing systems, or assembly systems in manufacturing. *Machine Autonomy* is best described as 'scripted' systems, where all decisions are enacted by encoded or embedded script and unknown scenarios result in the machine stopping and asking for human assistance.

Intelligent Autonomy may be described as sophisticated machine autonomy, with the system being capable of writing or modifying the 'script' as well as being able to make decisions during operation. In *Intelligent Autonomy*, systems will respond to unknown situations or unexpected events and attempt to resolve the issue without human intervention.

Defining Levels of Autonomy

Formal endeavours to define levels of autonomy have shifted from a focus on computer capabilities toward measurement of the interaction and collaboration between humans and machines to achieve outcomes. Distinctions may be drawn between existing concepts and definitions of integration, automation and levels of autonomy, and how both humans and AI perform within each classification or level. Consequently, these concepts can best be defined using a systems engineering approach, where differentiation between integration, automation and autonomy is determined by the degree of human control or intervention during various stages:

Integration

The act of combining or adding parts to make a unified whole. The term *Integrated Security* implies a combination of security technologies, functions and devices, or quite simply; an assimilation of different security services which communicate to perform advanced functions in, as a minimum, an automated manner.

Automation

The technique, method, or system of operating or controlling a process by automatic means, as by electronic devices, reducing human intervention to a minimum. In automation, human decisions and logic are executed to accomplish a pre-set series of tasks within a known, or assumed, frame of reference without decisions being made during operations.

Semi-autonomy

Technical outputs that involve machine decision making (in response to external, unexpected events) during operation, but a human is involved in some of this process and provides some direct control. Semi-autonomous systems are more independent and agile than automated systems.

Autonomy

Autonomy is broadly defined as the condition of being autonomous; self-government, or the right of self-government; independence. An autonomous system is one where decisions are made within the system and do not involve or require human decision making. A fully autonomous system is able to respond to unknown and unexpected events without pre-programming, scripting or assistance from a human.

Frost (2011, p.89)[5] offers the following differentiation between automation and autonomy, and introduces the concept of intelligent autonomy:

> An *automated system* doesn't make choices for itself – it follows a script, albeit a potentially sophisticated script, in which all possible courses of action have already been made. If the system encounters an unplanned-for situation, it stops and waits for human help (e.g. it "phones home"). Thus, for an automated system, choices have either already been made and encoded, or they must be made externally. By contrast, an *autonomous system* does make choices on its own. It tries to accomplish its objectives locally, without human intervention, even when encountering uncertainty or unanticipated events. An ***intelligent*** *autonomous system* makes choices using more sophisticated mechanisms than other systems. These mechanisms often resemble those used by humans. Ultimately, the level of intelligence of an autonomous system is judged by the quality of the choices it makes.

Frost (2011) identified *decision making* as the differentiating factor between automation and the various levels of autonomy, a view well supported across the broader literature. Consequently, the higher the level of autonomy in an intelligent system, the lower the degree of human control and decision making. Nevertheless, to date, there is limited characterisation of the extent to which intelligent systems are able to plan, direct and execute operations or missions. Furthermore, existing defined levels of autonomy do not explicitly link the concepts of *Narrow AI*, *Broad AI* and *General AI* with the characterisation of intelligent decision making and control at each ascending level.

Subsequently, the developed *Security Technologies Intelligent Autonomy Scale* (adapted from Chang, 2014[6]; Proud & Hart, 2005[7]) provides indication of the concept of 'levels' of intelligent autonomy within the predefined AI categories of *Narrow*, *Broad* and *General AI*. The scale articulates the levels of intelligent autonomy which may be used by security technologies, ranging from Level 1: *Manual* (absence of AI) through to Level 11: *Post-Autonomous* (General AI) and defines the technical AI characterisation at each increasing level of autonomy. At Level 11: *Post-Autonomous*, there is expectation that General AI will possess *Theory of Mind*, and be self-aware, understand belief, thoughts, emotions, expectations of people and interact socially. Such a theory suggests systems at the level of General AI would be capable of planning, directing, executing and reviewing security missions with no human intervention.

---

[5] Frost, C. (2011). Challenges and Opportunities for Autonomous Systems in Space. In *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2010 Symposium* (pp. 89-101). National Academies Press. https://doi.org/10.17226/13043

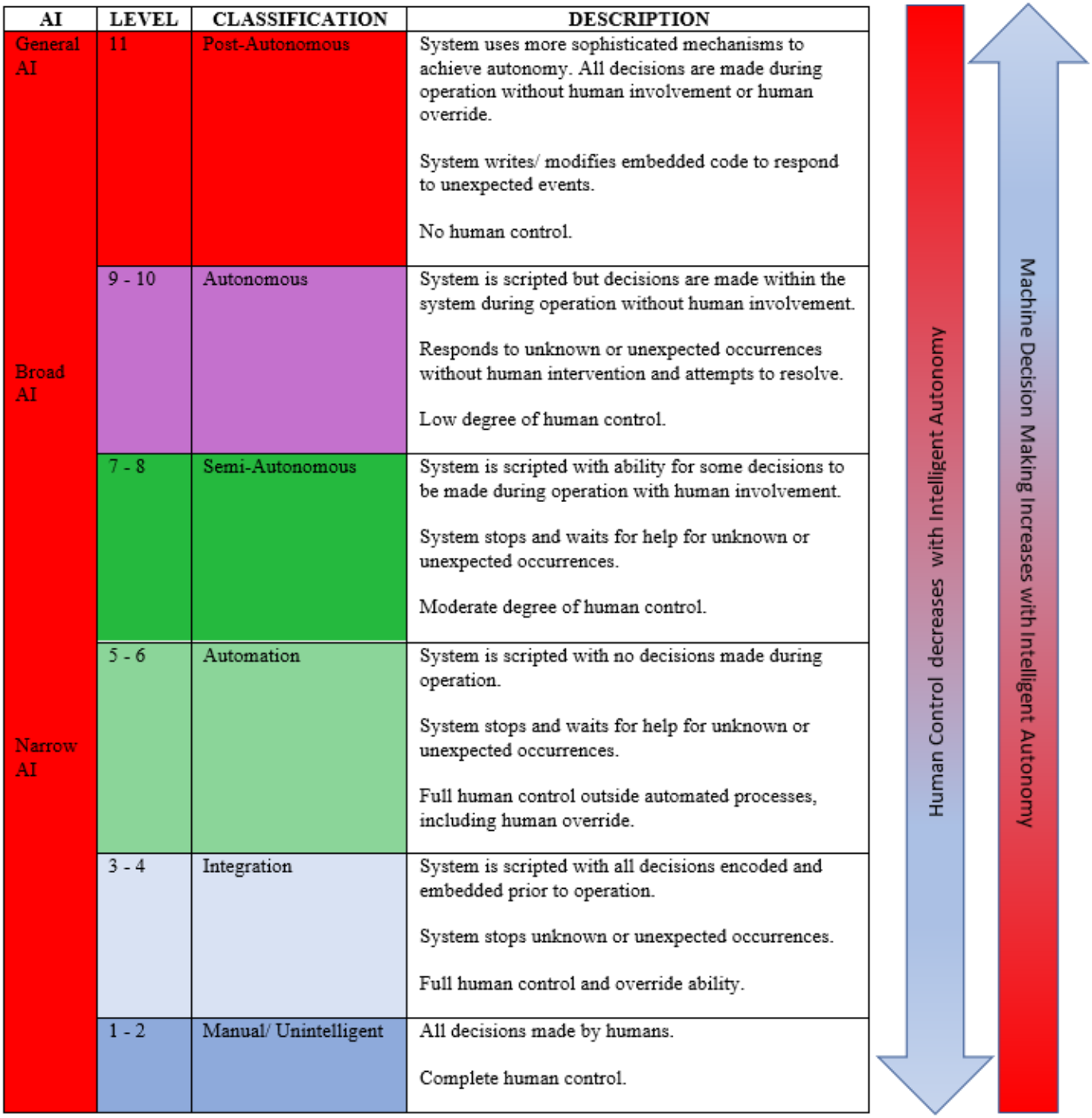
[6] Chang, E.M. (2014). *Defining the Levels of Adjustable Autonomy: A Means of Improving Resilience in an Unmanned Aerial System* [Master of Science Thesis, Naval Postgraduate School]. NPS Archive. http://hdl.handle.net/10945/43887

[7] Proud, R.W., and Hart, J.J. (2005). FLOAAT, A Tool for Determining Levels of Autonomy and Automation, Applied to Human-Rated Space Systems. *Infotech@Aerospace*, 7061. American Institute of Aeronautics and Astronautics (AIAA). Arlington, VA: AIAA.

Levels of Intelligent Autonomy in Security Technologies

The Security Technology Intelligent Autonomy Scale is a reference framework for security professionals to identify the level of intelligent autonomy and evaluate the capabilities and potential risks arising from AI enabled systems or devices. Figure 1 introduces the summarised Levels of Autonomy extracted from *The Security Technology Intelligent Autonomy Scale* (see Appendix A for the full scale).

Figure 1. *Summary of Levels from The Security Technology Intelligent Autonomy Scale*

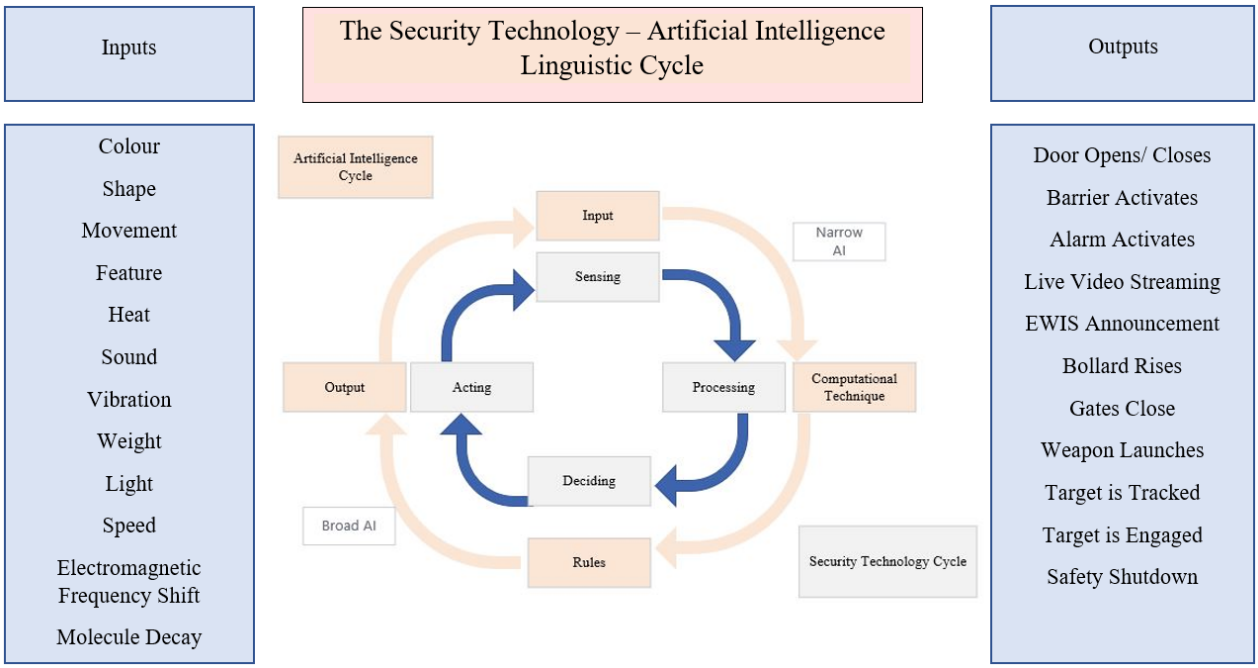


AI	LEVEL	CLASSIFICATION	DESCRIPTION
General AI	11	Post-Autonomous	System uses more sophisticated mechanisms to achieve autonomy. All decisions are made during operation without human involvement or human override. System writes/ modifies embedded code to respond to unexpected events. No human control.
Broad AI	9 - 10	Autonomous	System is scripted but decisions are made within the system during operation without human involvement. Responds to unknown or unexpected occurrences without human intervention and attempts to resolve. Low degree of human control.
	7 - 8	Semi-Autonomous	System is scripted with ability for some decisions to be made during operation with human involvement. System stops and waits for help for unknown or unexpected occurrences. Moderate degree of human control.
Narrow AI	5 - 6	Automation	System is scripted with no decisions made during operation. System stops and waits for help for unknown or unexpected occurrences. Full human control outside automated processes, including human override.
	3 - 4	Integration	System is scripted with all decisions encoded and embedded prior to operation. System stops unknown or unexpected occurrences. Full human control and override ability.
	1 - 2	Manual/ Unintelligent	All decisions made by humans. Complete human control.

The Security Technology-Artificial Intelligence Cycle

Within security technologies and devices, AI is implemented across the *technology cycle* by which a technology receives input, processes the input, decides what to do as a result of that input, and performs some action in line with that decision. The Security *Technology-Artificial Intelligence Linguistic Cycle* (Figure 2.) communicates the alignment between how security technology operates through sensing, processing, deciding and acting, and the AI cycle of data input, computational technique, rule checking and defined output.

Figure 2. *The Security Technology-Artificial Intelligence Linguistic Cycle*



The Security Technology-Artificial Intelligence Linguistic Cycle can be explained through practical examples of security technologies. For instance, a Passive Infrared (PIR) sensor detects a change in thermal radiation (input/sensing) created by the electromagnetic energy radiated by humans or other sources of thermal radiation. The rate of change is evaluated (computational technique/processing) and where the change in radiation exceeds a pre-set value (rules/deciding) an alarm state is produced (output/acting).

This basic cycle applies to all security technologies across all AI paradigms and approaches. The PIR sensor is an example of a security technology which is logic-based and sits within the Symbolic AI Paradigm, where stimuli or input must meet pre-defined rules to invoke some action or output such as an alarm. In contrast, a biometric access control system which uses facial recognition sits within the Statistical AI Paradigm. For example, though there are several ways in which the facial features of an individual can be 'mapped', the basic premise of facial recognition technology is that an individual's facial features are extracted from a scanned image and stored in a database.

For facial recognition, the system compares subsequent facial scans (sensing/input) to the stored template (data) and calculates the *probability* of the input 'matching' the stored information (computational technique/processing). If the probabilistic matched score is within an acceptable threshold (rules/deciding), the individual is accepted and access is granted, for example, through the release of the door's locking mechanism (output/acting). Thus, although the biometric facial recognition and PIR sensors are technologically different in their principal of operation and sit within different AI paradigms, they both follow the same technical cycle of operation.

Section 2 - Understanding Current Use of Artificial Intelligence in Security Technologies

According to ASIS International (2015[8]), the categories of security technology include Observe, Detect, Control, and Response. Across these categories the current status of AI in security technologies can be summarised as follows:

- The majority of security technologies sit within the Symbolic and Statistical AI Paradigms, and predominantly use Narrow AI techniques.
- Machine Learning in Security Technologies is at an elementary level, with evidence of Machine Learning within Network Video Surveillance Analytics, Biometric System Analysis and Management, Acoustic Detection Systems, and Drone and Robotics Analysis and Management – where these systems know only the data they have been provided and cannot yet interpret the 'unknown'.
- AI can make mistakes, with no way of knowing a mistake has been made, impacting upon the accuracy and reliability of Security Technologies.
- There is no evidence of General AI or Artificial Super Intelligence (ASI) in security technologies.

The use of AI may differ between security technology categories and types, and may be better understood by reviewing how AI is used within the categories of Observe, Detect, Control, and Response technologies.

Observation Technologies

ASIS International considers the core functions of observation technologies to be the detection of an approaching threat, characterisation of the threat, providing aid in formulating a response, and assistance in investigative efforts after an event occurrence. Network surveillance systems are perhaps the most prevalent of the observation technologies and are therefore presented in Table 1., as an example of how observation technologies align with AI paradigms and computational techniques.

Table 1.

Example of Observing Technology and AI Analytical Alignment.

Technology		**Security Function**	**Event Characteristics or Elements (Stimuli)**	**AI Paradigm**	**AI Computational Technique**
Network Video Surveillance	Digital Video Surveillance	Observe, Detect, Recognise, Identify	Colour, shape, contrast, silhouette, feature, movement	Statistical; Computer vision (Probabilistic Programming)	Background Analysis, and Statistical Modelling (Subspace Learning, Kernel Density Estimation or Gaussian Mixture Method).
	Video Motion Detection	Detection	Contrast, Silhouette, as indicative of Movement (Statistical Change in Colour/shade)	Statistical; Computer vision (Probabilistic Programming)	Pixel Matching, Foreground Analysis, Background Subtraction, or Gaussian Mixture Method (GMM)

[8] ASIS International. (2015). *Physical Security Principles: The essential sourcebook for the physical security professional*. Alexandria, VA: ASIS International.

Observation Technologies currently sit in the Symbolic and Statistical AI Paradigms and include perception, knowledge and planning problem domains.

- Network Surveillance Systems are predominantly Symbolic (Logic-based and Knowledge-based) with Video Analytics and Management Systems (Configuration & Programming) spanning across Symbolic (Knowledge-based) and Statistical Paradigms (Probabilistic and Machine Learning).

- Drones and Robotic sensors are predominantly Symbolic (Logic-based). Drone and Robotics Analytics are aligned with Symbolic (Knowledge-based) and Statistical (Probabilistic and Machine Learning) Paradigms. Management Systems (Configuration and Programming) are aligned with across Symbolic and Statistical Paradigms.

- At the time of writing, no Observation technologies currently sit within the Sub-symbolic AI Paradigm.

Detection Technologies

Detection technologies form a key component within an Intrusion Detection System (IDS) and a more holistic Physical Protection System (PPS), and are primarily used to determine that an unauthorised action is occurring or has occurred by sensing the action stimuli and communicating the alarm to a control centre for assessment. Detection technologies may include a range of internal and external intrusion sensors, alarm devices and systems, lighting systems, x-ray technology, trace detection systems, acoustic threat detection sensors and systems, RADAR and SONAR sensors and systems. Table 2 presents three examples of detection technologies and how each aligns with AI paradigms and computational techniques.

Table 2.

Examples of Detection Technologies and AI Analytical Alignment.

Technology		**Security Function**	**Event Characteristics or Elements (Stimuli)**	**AI Paradigm**	**AI Computational Technique**
Internal and External Intrusion Detection Systems	Contact Switch/ Reed Switch (Electro-mechanical Sensor)	Detection	Current Interference (loss) through breaking of connection between the two points	Symbolic (Logic Based)	Logic Programming or Rule based Process output
	Passive Infrared Sensors (Thermopile or Pyroelectric Detector)	Detection	Thermal energy threshold change	Symbolic (Logic Based)	Logic Programming or Rule based Process output
Trace Detection	Explosive/ Narcotics Trace Detection (Ion Mobility Spectrometry)	Detection	Threshold change in concentration of explosive/drug molecules in saturated air or surfaces	Symbolic (Logic Based)	Logic Programming or Rule based Process output

Detection Technologies currently sit within the Symbolic and Statistical AI problem domains and spectrum of AI paradigms.

- Intrusion Detection System sensors, including Biometric sensors, communications and hardware are Symbolic. Management Systems (Configuration & Programming) are located within the Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms. Intrusion Detection Analytics located in the Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

- Biometric System Analytics and Management Systems align to the Statistical (Probabilistic and Machine Learning) Paradigm.

- SONAR and RADAR sensors, communications and hardware are Symbolic with Analytics and Management Systems (Configuration & Programming) located across Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

- Network Surveillance Video Analytics used for anomaly detection are aligned with Symbolic (Knowledge-based) and Statistical (Probabilistic and Machine Learning) Paradigms.

- Acoustic Detection System Analytics are aligned with both Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms, while the management of Acoustic Detection Systems is aligned with the Statistical Paradigm (Probabilistic and Machine Learning).

- Drone sensors as detection mechanisms are located with the Symbolic Paradigm (Logic-based).

- X-ray sensors are Symbolic (Logic-based) with both Analytics and Management Systems (Configuration & Programming) located within the Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

- Trace Detection Systems are Symbolic (Knowledge-based) with both Analytics and Management Systems (Configuration & Programming) located within the Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

- Lighting Systems including sensors, activation and management are aligned with the Symbolic Paradigm (Logic-based).

- No Detection technologies currently sit within the Sub-symbolic AI Paradigm.

Control Technologies

The function of controlling is usually associated with access and egress control in the protection of assets. Access control is described by ASIS International as any technique that is employed to limit or otherwise control access to an area, facility, compound, system, person or asset. Controlling technologies primarily consist of 'access control' systems, with numerous types of sensors and credentials. Table 3 presents AI alignment examples of an access control credential and biometric systems which are often incorporated into access control systems.

Table 3

Examples of Controlling Technologies and AI Analytical Alignment

Technology		**Security Function**	**Event Characteristics or Elements (Stimuli)**	**AI Paradigm**	**AI Computational Technique**
Access Control Systems	EAC Reader	Identify, Control	Presentation or input (e.g. PIN) of authorised credential at reader	Symbolic (Logic Based)	Logic Programming or Rule based Process output
Biometric Systems	Biometric Facial Recognition – Holistic Matching Method	Detect, Recognise, Identify	Eigenfaces/ Eigenvector & Pixel Intensity Array without the detection of facial features	Statistical; Bayesian Program synthesis (Probabilistic Programming)	Principal Component Analysis, Linear Discriminant Analysis, or Independent Component Analysis
	Biometric Facial Recognition – Feature-based	Detect, Recognise, Identify	Facial features create a geometric relationship between measurement points	Statistical; Bayesian Program synthesis	Structural Similarity Measure (SSIM), and Feature Similarity Measure (FSM),

	(Structural) Method		of the individual's unique features	(Probabilistic Programming)	Principal Component Analysis, Linear Discriminant Analysis, or Independent Component Analysis

Controlling Technologies currently sit within the Symbolic and Statistical AI Paradigms.

- Access Control System sensors, including Biometric sensors, align with the Symbolic Paradigm. Access Control Management Systems (Configuration & Programming) and Analytics are located within the Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

- Biometric System Analytics and Management Systems align to the Statistical Paradigm (Probabilistic and Machine Learning).

- Vehicle and Dispensable Barrier and Turnstile hardware and mechanisms are located in the Symbolic Paradigm (Logic Based).

- No Controlling technologies currently sit within the Sub-symbolic AI Paradigm.

Response Technologies

Responding is the effort to neutralise, contain, or mitigate an event. It may also include an assessment that the event does not require immediate action. In the protection of assets, response typically includes the human-guard force intervention underpinned by human decision making. However, 'Response' as a function within the security domain has evolved considerably beyond the human guard-force intervention to a response mechanism which may include a range of technologies to assist human response. Current *Security Response Technologies* include Communications, Dispensable Barriers, Vehicle Barriers, and Weaponry Systems. Each of these individual technologies are comprised of numerous elements or components which may have varying technical outputs, as well as having varying levels of control applied. Table 4 presents two examples of response technologies and how they align with AI paradigms and computational techniques.

Table 4

Examples of Responding Technologies and AI Analytical Alignment

Technology		**Security Function**	**Event Characteristics or Elements (Stimuli)**	**AI Paradigm**	**AI Computational Technique**
Weaponry	Electro-magnetic Weaponry (Directed Energy Weapon)	Delay, Response	Electromagnetic energy beam created by a Gyrotron (Vacuum Electronic Device) to create high power high frequency THz radiation)	Symbolic (Logic Based)	Logic Programming or Rule based Process output
	Autonomous & Semi-Autonomous Weapons Systems	Delay, Response	Predefined Stimulus for which autonomous systems may execute self-direction, self-learning or emergent behaviour to select and attack targets	Symbolic (Logic Based); Statistical Bayesian Program synthesis (Probabilistic Programming)	Logic Programming or Rule based Process output; Multiple Algorithms for Classification, Regression or Clustering

Response Technologies sit within the Symbolic and Statistical AI Paradigms.

- Communications System devices and hardware align with the Symbolic Paradigm (Logic-based and Knowledge-based).

- Robotic System automation and management functions are located within the Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

- Drone Capture Systems used in response technologies are located within the Statistical Paradigm (Probabilistic).

- Weapons Systems sensors, communications and hardware are Symbolic with Analytics aligning with the Statistical Paradigm (Probabilistic and Machine Learning). Weapons Management Systems (Configuration & Programming) spanning align with both Symbolic (Knowledge-based) and Statistical (Probabilistic) Paradigms.

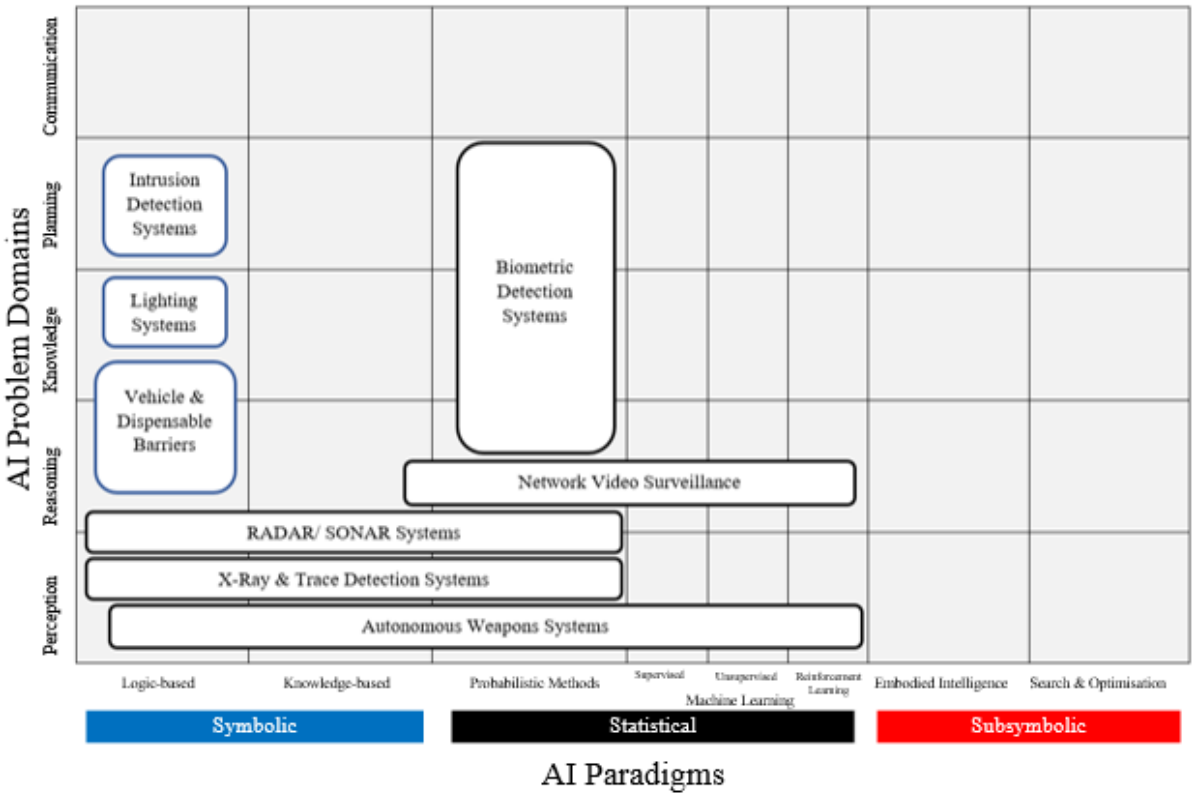
- No Response technologies currently sit within the Sub-symbolic AI Paradigm.

Security Technology Alignment to the AI Spectrum of Paradigms

Figure 3 presents an adapted version of Corea's (2019) *AI Knowledge Map*, with superimposed security technologies, indicating the locality of various security technologies along the visual map aligned to the AI problem domains and spectrum of paradigms.

Figure 3.

Security Technology Alignment to the AI Problem Domains and Spectrum of Paradigms



Section 3 - Future Opportunities and Risks

Benefits and Opportunities for AI in Security Technologies

Artificial Intelligence will continue to develop, and there are a number of tangible benefits of adopting AI in security technologies. Many of these relate to economic benefits, such as increased productivity and reduced costs from enhancement of narrow AI tasks. The overwhelming benefit to humanity will be the use of response technologies such as drones and robotics to remove humans from harm's way.

There are several opportunities for the development of AI in security technologies. These opportunities are expected to present across Observe, Detect, Control, Response and Integrated Security Technologies and include the following themes:

- AI Testing and Standardisation
- Technical Improvements
- Integration and Aggregation of Inputs
- Analytics
- Human-Machine Teaming
- Holistic Approach to AI Development
- Innovative Development

AI Testing and Standardisation

Opportunities for standardisation include the development of international AI standards, in addition to the design and development of common connectivity protocols across platforms, equipment and devices. Standardisation and uniformity of protocols will reduce connectivity issues and increase the number and type of system inputs required for greater intelligent interpretation of data. Development and standardisation of testing mechanisms will also drive accuracy and reliability by providing assurances to security managers that systems achieve the level(s) of proficiency declared by manufacturers and vendors under a variety of operational conditions.

Technical Improvements

Increased accuracy and reliability, particularly in live operating environments and under dynamic conditions is essential, as well as improvements in safety, quality, resolution, processing capabilities and reduction in signal noise. Further opportunities for technical improvements include novel detection devices, development of source datasets, increased mobility for sensors/devices, and scale and processing capabilities. Observation technologies specifically require greater accuracy in object recognition and classification, and data management systems for large volumes of information produced.

Integration and Aggregation of Inputs

Security technologies will benefit from increased automation and greater integration of internal and external systems (e.g. building and security systems). Aggregation of multi-sensor data with applied analytics has the potential to enhance security decision making and provide distributed intelligence between systems, facilities and security teams for improved security response. However, to achieve this, integrated systems essentially require transformation from a logical input/output type architecture to a more sophisticated architecture with a coordinated approach to decision making between devices and throughout the field, automation and management BACS levels. Enhanced integration will also improve the way in which systems are able to be used, for instance cross platform control may increase flexibility in how systems or devices such as drones and robotics are accessed and controlled during operation.

Analytics

Analytics and predictive analytics for pattern recognition and anomaly detection will present significant opportunities for security technologies. The development of novel analytical methodologies will be required to support decision making, provide greater context in dynamic environments, and provide dynamic assessments for distributed intelligence and response systems. Decision making tools such as 'ranked options' for response may assist security practitioners to make compliant decisions based on the best probability of outcomes. Together with Human-Machine Teaming, AI generated options for response could considerably enhance critical decision making for human operators.

Human-Machine Teaming

Human-machine Teaming (HMT) explores and promotes collaborative development through the merging of humans with machines to achieve specific goals and capabilities. The goal of HMT is to collaboratively assist humans rather than designate or redirect human activities to machines. Opportunities for HMT exist for across all security technologies, particularly in Response technologies such as drones and robotics, where trust and safety issues mean that human control over systems and devices is preferable. HMT within integrated security management systems could therefore offer substantial improvements and benefits for security operators without compromising critical decisions and potentially impacting on the security and safety of individuals.

Holistic Approach to AI Development

Technical development must be supported by enhancements across management, industry and at an international level. This means holistic and synchronistic development of AI must occur across all domains and disciplines which overlap the security function, including but not limited to safety, facility management, engineering, academia and commercial research and development. Social concepts such as trust, safety, responsibility, accountability and integrity as qualitative drivers of artificial intelligence must be explored fully to create the supporting foundations for evolution across all technology categories. Policies and frameworks must also be established for the development and deployment of AI-enabled security technologies in socially acceptable ways.

At a holistic management level, AI adoption should be explicitly linked with risk in terms of the consequences of failure. Essentially AI may be adopted more readily where there are less severe consequences of AI failure, and adoption should be reduced as the consequences of failure increase. In a practical sense, this means promoting the use of AI in more logical applications such as presentation of user credentials (e.g. access card), or where statistical probability of detection is relatively high due to stability of materials or signatures (e.g., x-ray or trace detection of target compounds). There may also be increased opportunity in environments where there are several layers of defence in depth working simultaneously so that the consequences of failure are lessened due to the statistical probability of detection by another technical or procedural measure. Thus, defence in depth strategy should be actively applied to the implementation of AI techniques and consider the strategic placement of humans and machines to produce optimal security outcomes.

Holistic approaches to AI development present a series of opportunities for the security industry to influence the social development of AI technologies. Opportunities exist not only for commercial, legal, social and political development of these policies and frameworks, but for the security industry as an exemplar of AI development.

Innovative Development

Opportunities exist for innovative development of security technologies. Unique opportunities for innovative development of drone swarms and robotics are emerging for *Response* technologies. There may also be scope for greater development of geospatial AI within response technologies, which for instance may include specific tools such as geofencing to define virtual perimeters (i.e. latitude, longitude,

altitude, etc.) in which an object such as a drone or robot may operate. Further opportunities may exist for intelligent deployment capabilities for various classes of less-lethal weapons, and development of novel classes of weapons with intelligent autonomy may have a similar commercial viability in specific applications or environments. Development in integrated security technologies may facilitate innovative applications such as automated crowd screening using multi-sensor inputs for increased context and threat detection. While some progress has been made in these areas, the capabilities of such technologies are still very much in the elementary stages of development.

Potential Risks of AI in Security Technologies

There are considerable risks of developing AI in security technologies, the consequences of which may not be fully comprehendible at this point in time. At a global level, the quest for technological advancement may create political divides, upset balances of power, encourage exploitation of underdeveloped nations, and promote the abuse of individual privacy and rights. Development of military and security response technologies with the capacity for autonomous use or release of force may eventually have the power to determine life and death, or inflict injury or harm onto humans. While this level of intelligent autonomy is not currently achievable in commercially available security technologies, the desire for military supremacy combined with the porous nature of military-commercial exchange will likely see the autonomous use and release of force become a reality. The potential for harm to result from development and deployment of these technologies means there must be extensive legal, moral, ethical and human rights considerations afforded to discourse on intelligent autonomy, provided through enforceable international governance platforms.

The following case studies present both practical and potential risks of AI. *Case Study 1* describes the potential for AI systems with Artificial Neural Networks to override programmed instructions in the pursuit of self-preservation. While the researchers in the case study may have intervened and re-programmed the robots to correct their behaviour, this type of risk may be difficult to manage under circumstances where the technology has been widely distributed or deployed, or where users have limited knowledge or resources to monitor and rectify such programming issues.

Case Study 1

From Del Monte, L.A. (2018). *Genius Weapons: Artificial Intelligence, Autonomous Weaponry and the Future of Warfare*:

In 2009, researchers from the Laboratory of Intelligent Systems at the Federal Institute of Technology in Lausanne performed experiments that suggest even primitive artificially intelligent machines are capable of learning deceit, greed, and self-preservation, without the researchers programming them to do so. The Lausanne research team programmed small, wheeled robots to find "food". In this experiment, a light-colored ring on the floor signified food. They also programmed the robots to avoid "poison", which was signified by a dark-colored ring. A robot received a reward (i.e., points) when it found the food. The robot continued to receive points by staying close to the food. If a robot found poison, it lost points. In addition, each robot had a blue light. The researchers programmed each robot to flash the blue light when it found food. The other robots could detect this flashing blue light and join the robot at the food source. They too would also receive points. The goal of the researchers was to have the robots cooperate with each other in the process of finding food and avoiding poison.

According to the authors, "Over the first few generations, robots quickly evolved to successfully locate the food, while flashing the blue light. This resulted in a high intensity of light near food, which provided social information allowing other robots to more rapidly find the food." Some robots were more successful than others. Therefore, following each experiment, the research team would use the data taken from the most successful robots to "evolve" a new generation of robots. They did this by replicating the artificial neural networks of the most successful robots in the less successful robots. The experiment was set up such that the space around the food, a light colored ring on the floor, was limited.

It was not large enough to fit all robots. When a robot found food and flashed its light, the other robots quickly moved in, creating chaos via bumping and jostling each other. In the midst of this chaos, the original robot that found the food could end up being bumped out of position.

By the fiftieth generation, some robots stopped flashing their light when they found food, ignoring their programming. In addition, some robots became deceitful and greedy. They would flash their light when they found poison, which lured the other robots to the poison, resulting in those robots losing points. After several hundred generations, all robots learned not to flash their light when they found food. This critical experiment implies robots can learn deceit and greed. I argue they also learn self-preservation. These robots were not able to learn via their own experience. They evolved with the help of researchers, who replicated the neural networks of the most successful robots into the less successful at the conclusion of each experiment. Now imagine a time when self-learning robots have intelligence equal to humans-level intelligence. The Lausanne experiment suggests they will act in their own best interests, even ignoring their programming. It is not clear that they will follow any innate moral code or respect laws expressed in their programming. Obviously, the Lausanne robots ignored their original programming and evolved their own laws (pp.142-143).

Case Study 2 further demonstrates the practical and potential risks of AI. While this case is 'pre-1999' and relates to autonomous weapons systems rather than commercial security systems, it raises significant issues as to how safe the assessments by AI, deep learning and predictive algorithms may be, and to what degree these systems may be able to execute instructions autonomously.

Case Study 2

From *Safety of Autonomous Systems Working Group (2018). Safety-related Challenges for Autonomous Systems*:

Operator's Choice Overridden by Software, pre-1999: During field practice exercises, a missile weapon system was carrying both practice and live missiles. Transit time was being used for slewing practice. Practice and live missiles were located on opposite sides of the vehicle. The operator acquired the willing target, tracked it through various manoeuvres, and pressed the weapons release button to simulate firing the practice missile. Without the knowledge of the operator, the software was programmed to override his missile selection in order to present the best target to the best weapon. The software optimized the problem, de-selected the practice missile and selected the live missile. When the release command was sent, it went to the live missile. The "friendly" target had been observing the manoeuvres of the incident vehicle and noted the unexpected live launch. Fortunately, the target pilot was experienced and began evasive manoeuvres, but the missile tracked and still detonated in close proximity (p.23).

[This case study highlights the potential implications of autonomous decision making and how accidents or incidents may occur in the event that users do not fully understand the behaviour of the software system, or if all possible scenarios are not considered at the design stage].

To understand the benefits and risks associated with the adoption of AI technologies, a 'PESTEL Analysis' was undertaken. PESTEL risks include *Political*, *Economic*, *Social*, *Technological*, *Environmental* and *Legal* factors which may affect an organisation. A PESTEL framework is a management tool used to evaluate strategic decisions within an organisation, and therefore can be used to identify the areas which may be impacted by the design, development and deployment of AI in security technologies. A full description of benefits and risks associated with the adoption of AI technologies in

the protection of assets is presented in the formal research report. Appendix B presents these benefits and risks within a PESTEL framework.

Section 4 - Evaluating the Benefits and Risks Associated with AI Embedded Technologies

To understand the potential risks associated with AI technologies, security managers and decision makers are encouraged to take a risk-based approach. Risk standards and guidelines such as ANSI/ASIS/RIMS RA-12015, ISO 31000, or Standards Australia AS/NZS HB167:2006 are instruments which function as an aide memoire to facilitate the consideration of risk factors. Such a risk approach should therefore be undertaken when evaluating the suitability of AI technologies for adoption within specific contexts and environments.

In considering a risk-based approach to AI adoption, tabulated questions were developed to assist security professionals and practitioners in making an informed assessment of the AI technology or product they are considering. Again, this list is not exhaustive, but rather focused on supporting a risk-based approach, where defined questions prompt security professionals to consider risks within a PESTEL framework, according to the level of *Intelligent Autonomy* a technology may achieve. These levels reflect the *Security Technology Intelligent Autonomy Scale* (Appendix A), comprising *Manual/Unintelligent* (i.e. the risks of not adopting AI technologies) through to *Post-Autonomous*.

What Are Your Organisational Risk Factors?

MANUAL/ UNINTELLIGENT SYSTEMS	Yes/ No
What are the Risks of NOT adopting AI Technologies?	
POLITICAL	
Could you be disadvantaged by a lack of data and intelligence through the absence of AI?	☐
Do public perceptions of AI support your decision to not use AI-enabled technologies?	☐
ECONOMIC	
Are you missing potential budget and cost reductions by not adopting AI across observe, detect, control and response technologies?	☐
Could you reduce nuisance alarms through the use of AI enhanced technologies?	☐
Could you increase operational capabilities or better allocate resources by adopting AI enhanced technologies, including at the integration level?	☐
Could the adoption of AI across security technologies increase efficacy by reducing operator workload and fatigue?	☐
SOCIAL	
Are personnel or stakeholders exposed to safety risks which could be avoided through the use of intelligent systems (e.g. robotics or drones)?	☐
Could personnel reasonably be removed from harm's way with AI technologies or devices?	☐
Could AI technologies assist in the detection of Insider Threat?	☐
Could AI technologies enhance early intervention opportunities?	☐
Could the adoption of AI-enabled security technologies provide organisational convenience?	☐
Can the use of AI-enabled technologies increase privacy by reducing human monitoring?	☐
TECHNOLOGICAL	
Could Human-Machine Teaming enhance your capacity to detect and mitigate threats?	☐

Would your organisation benefit from real-time processing and assessment across the spectrum of AI problem domains?	☐
Could AI-enhanced surveillance benefit your security operations?	☐
Could your control of, or response to, critical incidents improve as a result of using AI technologies?	☐
Can the use of AI technologies increase threat mitigation opportunities?	☐
Could the use of AI technologies reduce latency/ delays in communication between technology systems and devices?	☐
LEGAL	
Can increased traceability of incidents through AI-enabled technologies assist in mitigating legal issues and liability?	☐
Could the use of intelligent devices (e.g. drones or robotics used assist or replace humans) reduce legal liability from safety risks?	☐

INTEGRATED SYSTEMS	Yes/ No
POLITICAL	
Are there any privacy, ethical or human rights issues through the collection/storage/use of data or information?	☐
Are there any public perceptions of risk associated with this technology that need to be managed?	☐
Is there regulatory control (if any) of this technology in your jurisdiction, and can you comply with it?	☐
Is the use of AI technology likely to create industry disruption or industrial relations issues for your organisation?	☐
ECONOMIC	
Are the indicative costs of purchase, installation and maintenance of AI your proposed technologies viable?	☐
Is a service agreement available from the manufacturer or supplier to guarantee scheduled maintenance costs?	☐
Do you know the impact on your organisation if the manufacturer or supplier of your AI product ceased to operate?	☐
Is more than one manufacturer or supplier available to provide ongoing maintenance and assistance for AI systems?	☐
Are there substantial economic costs of AI related maintenance, disruption and downtime?	☐
Can the technology be operated manually if required?	☐
Do you know how AI- system reliance affects your organisation, personnel and operational needs?	☐
Are the potential costs associated with retention and storage of AI data acceptable?	☐
Does the location of AI data (i.e. country or region) affect ongoing budget costs?	☐
Are there potential economic and legal liability costs arising from AI use, misuse, error or failure?	☐
Are there provisions for resilience and redundancy of the AI technology and/or system?	☐
What are the direct and associated costs of AI resilience?	☐
What is the expected rate of AI enabling technologies decay?	☐
How often will the AI technology require upgrading?	☐
SOCIAL	
Are there any social or reputational implications of using this AI technology?	☐
Could the technology be used to exploit personal information (e.g. data and information from social media)?	☐
Does use of the technology present risks to personal or personally identifiable data of individuals (e.g. biometric information)?	☐
TECHNOLOGICAL	
Does this technology present any specific safety risks?	☐
What is the operational impact of AI technical issues or failures?	☐
What redundancy measures (if any) are in place to minimise disruption and downtime from AI issues?	☐
Is this technology (or parts of the system) deployed in the cloud?	☐
For cloud hosted AI, are there any jurisdictional issues associated with ownership, storage and access to data?	☐
For cloud hosted AI, what security measures are available to protect data and systems and minimise AI vulnerabilities?	☐
For cloud hosted AI, are there backdoor vulnerabilities that need consideration?	☐
ENVIRONMENTAL	
Could there be any environmental impacts from the use, misuse or failure of this AI embedded technology?	☐

LEGAL	
What are the legal risks associated with use of this technology?	☐
Could legal liability arise as a result of AI use, misuse or failure?	☐
Could an AI-related environmental spill, discharge or emission result in a caution or notice from environmental agencies or regulatory bodies?	☐

AUTOMATED SYSTEMS	Yes/ No
Risk Factors from Integrated Systems, plus:	
POLITICAL	
Do you have an assigned responsible and accountable officer for AI systems, including purchase, installation, operations, issues and failures?	☐
Could political, social, cultural or other biases be created by machine learning algorithms that may have an impact on your organisation?	☐
Could the consequences of AI use, misuse or failure impact on the security industry or other industry sectors (e.g. create industry disruption, reputational harm, industrial relations issues, etc.)?	☐
Do you understand how this technology/device/system engages with humans?	☐
Are there any potential civil liberties and civil rights implications of this technology engaging with humans?	
Could your uncontrolled AI technologies or devices impact on critical government functions or commercial operations (e.g. uncontrolled use of drones/UAV impact on commercial aviation and emergency response)?	☐
Could your AI technology be exploited by criminals (e.g. fail-safe mode used to gain access to a facility, or control of a system)?	☐
ECONOMIC	
Do you have the capacity to switch to 'manned' operations, and should you maintain training for manual systems? (i.e. are you reliant on AI, and how might you maintain operations if AI systems are down?).	☐
Are the potential economic costs associated with access to/ retrieval of data in the event of a dispute with the Independent Software Vendor (ISV) or Cloud Service Provider (CSP) hosting your data or providing the processing capability for this technology?	☐
Are there any potential economic costs associated with bankruptcy, natural disasters or other events which may affect the operation of the Independent Software Vendor (ISV) or Cloud Service Provider (CSP)?	☐
Are there additional costs for ensuring that sufficient resilience and redundancy is built into the technology and across critical business functions and systems?	☐
What are the economic costs of mitigating security vulnerabilities?	☐
SOCIAL	
Is there a high level of public/user acceptance for this technology?	☐
Are the safety risks from this technology socially acceptable?	☐
TECHNOLOGICAL	
Are there any safety risks known to exist for this technology?	☐
Could use, misuse, error or failure of this technology result in illness, injury or fatalities?	☐
Are there any safety risks which may emerge from the use of this technology within your specific context?	☐
For Machine Learning algorithms, how will the system be trained, and what data will it use?	☐
Does this technology incorporate or use IOT or Edge of Network devices?	☐
Are there suitable security measures available to protect data and systems against threats to IOT or Edge of Network devices?	☐
Will security and technology decay impact on vulnerability to exploitation, tampering and defeat?	☐
ENVIRONMENTAL	
Could use, misuse or failure of AI technology result in a spill, emission or discharge to the environment?	☐
LEGAL	
What are the potential legal liability and consequences of AI use, misuse or failure for your organisation?	☐

Can the legal risks of this technology be mitigated with insurance?	☐
Does your insurance company understand your risks associated with AI technologies?	☐
Could AI use, misuse or failure result in short term or temporary loss of statutory accreditation to operate?	☐
Could an environmental event from AI failure result in an infringement from environmental agencies or regulatory bodies?	☐

SEMI-AUTONOMOUS SYSTEMS	Yes/ No
Risk Factors from Integrated and Automated Systems, plus:	
POLITICAL	
Do you have governance in place to ensure semi-autonomous decision making does not infringe on or breach privacy, ethics or human rights?	☐
Is there scope for political impact resulting from the use or release of force (e.g. from Semi-autonomous Weapons Systems)?	☐
Is your AI decision making process transparent? (i.e. do decision making algorithms produce an output to provide transparency and traceability as to how and when each decision is made?)	☐
Do you know who is responsible for AI decision making (i.e. manufacturer, programmer, operator)?	☐
Are there any civil liberties and civil rights implications of semi-autonomous systems engaging with humans?	☐
Is the technology (or parts of the technology) able to be weaponised? If so, do you have controls in place to mitigate risks?	☐
Is this AI product essentially a military technology?	☐
Does this technology facilitate the privatisation of military AI, and what controls are in place to limit commercial use?	☐
Could the technology, or data generated from AI technologies, aid or facilitate foreign espionage?	☐
Could the technology be used to exploit underdeveloped countries (e.g. used in an experimental capacity for the benefit of developed nations)?	☐
ECONOMIC	
Do you have the capacity to switch to 'manned' operations, and can you maintain training for manual systems? What additional resources will you require to maintain a manned capability?	☐
Are there potential economic costs of legal or environmental liability?	☐
Are there additional costs for ensuring that sufficient resilience and redundancy is built into the technology and across critical business functions and systems?	☐
Are there economic costs of mitigating security vulnerabilities, including cloud-based risks and vulnerabilities?	☐
SOCIAL	
Has social distrust/ fear of this technology been considered?	☐
Do you have governance in place to ensure personal data and information is not exploited?	☐
Can Human-Machine Teaming (HMT) be applied to this technology to reduce social and safety risks?	☐
Can development and deployment of the technology be revoked in the event of unanticipated or undesirable outcomes?	☐
What is the Safety Integrity Level (SIL) of this technology (if any), and does the technology meet safety requirements?	☐
TECHNOLOGICAL	
Could Human-Machine Teaming (HMT) reduce technical risks of semi-autonomous decision making?	☐
Is the installation, use, maintenance and trouble-shooting of this technology overly complex?	☐
If these functions require outsourcing, what is the estimated response time for third parties required to attend to and resolve technical issues?	☐
Is use of this technology likely to reduce your physical response force?	☐
Does the adoption of AI have any impact on your response to critical incidents?	☐
Do you have redundancy measures in place to minimise disruption and downtime from AI issues?	☐
Are there critical business functions which could be impacted by technical failure?	☐
Is there a designated officer responsible for machine learning algorithms? How is appropriate learning guided, and what assurances may be provided to prevent rogue AI?	☐
Are there sufficient source data/ databases available for machine learning algorithms?	☐

Do you know how accurate the AI identification and classification of images is?	☐
Do you know how vulnerable the technology is to cyber and cloud-based security risks?	☐
Do you know the efficacy of security measures to ensure personal data is not accessed, breached or exploited?	☐
Are there any dangers or risks which may emerge from uncontrolled or adversarial use of this technology?	☐
Are there other suitable AI-embedded technologies available that present less risk?	☐
ENVIRONMENTAL	
Could the use of this technology result in significant environmental pollution or damage to ecosystems?	☐
Do you understand how the public might react to an environmental incident caused by the use of this technology?	☐
LEGAL	
Could use, misuse or failure of this technology result in legal liability under Common or Criminal Law?	☐
Could AI use, misuse or failure result in a loss of statutory accreditation to operate for an extended period?	☐
Could an environmental event result in investigation, prosecution or infringement from environmental agencies or regulatory bodies?	☐

AUTONOMOUS SYSTEMS	Yes/ No
Risk Factors from Integrated, Automated & Semi-Autonomous Systems, plus:	
POLITICAL	
What policies and procedures are in place to ensure autonomous inputs, decision making, outputs or responses do not infringe on or breach privacy, ethics or human rights?	☐
What would be the political impact – both domestically and internationally – of the use or release of force (e.g. from Autonomous Weapons Systems)?	☐
How transparent is the process of AI decision making and response? Can decision making and response be attributable to satisfy responsibility and accountability requirements?	☐
Who is responsible for AI system errors or failures (i.e. manufacturer, programmer, operator), and how is this regulated?	☐
Are there any civil liberties and civil rights implications of autonomous systems engaging with humans, animals or within social systems?	☐
Could the use of this technology result in sanctions or damage to international relations?	☐
Does the technology (design, development or deployment) create any risks to national security or sovereignty? E.g. is data stored or transmitted to a foreign entity?	☐
Is it possible to design, develop and/or manufacture the technology locally to mitigate sovereign risk?	☐
Is the use of this technology likely to create or increase political divides between developing and developed nations?	☐
ECONOMIC	
Do you know the potential economic costs of reputational harm, damage or loss associated with AI technology adoption?	☐
Are there any additional physical security requirements and/or costs for the storage, transport or handling of autonomous systems?	☐
Are there any economic costs of mitigating legal liability, environmental liability or collateral damage?	☐
Are there any potential economic costs of damage to international relations as a result of use, misuse or failure of this technology?	☐
Could there be any potential trade or economic costs from sanctions or damage to international relations?	☐
SOCIAL	
Is the technology likely to generate social fear, distrust, discontent, or undermine public order and security?	☐
Does the technology meet capable guardianship requirements? If so, to what degree?	☐
Does the technology enable you to meet your duty of care and Occupational Health & Safety obligations?	☐
Do you know the impact the use, misuse, error or failure of this technology may have on your reputation?	☐
Are there any substantial training requirements for this technology to ensure personnel are trained adequately and appropriately?	☐
Are there any Human-Machine Teaming (HMT) options available for this technology to lessen AI risks from autonomous decision making?	☐
Can development and deployment of the technology be revoked in the event of unanticipated or undesirable outcomes?	☐
Do you have policies and procedures in place to ensure personal data and information is not exploited?	☐
Do you have governance in place to guarantee that privacy, ethical and human rights violations do not occur?	☐
Has this technology been developed and deployed in socially acceptable ways?	☐
Is this technology supported by international legal and political frameworks?	☐
TECHNOLOGICAL	
Is the technology accurate and reliable?	☐

Could autonomous or unmanned systems create complacency in operators?	☐
Could Human-Machine Teaming reduce technical risks of autonomous decision making?	☐
Can machine learning algorithms be reverse engineered to identify where issues in machine learning have occurred?	☐
Will forensic auditing and assessment needs be achievable?	☐
Do you know the likelihood and consequence of autonomous failure of human override?	☐
Do you understand the potential outcomes of cyber and cloud-based security breaches (hacking, hijacking, modification of script, etc)?	☐
ENVIRONMENTAL	
Could the use, misuse or failure of this technology result in irreparable environmental damage or destruction? (E.g. Nuclear or chemical accident).	☐
Do you know how the international community might react to a critical environmental incident?	☐
LEGAL	
Are there established legal frameworks around the design, development and use of this system, device or class of technology?	☐
Are there any potential legal liability and consequences of AI use, misuse or failure?	☐
Could AI use, misuse or failure result in the permanent loss of statutory accreditation to operate?	☐
Could an environmental event result in investigation and prosecution from environmental agencies, regulatory bodies or contravene international law, treaties, conventions, or agreements?	☐
Would the use or release of force (e.g. from Autonomous Weapons Systems) contravene International Humanitarian Law or other international law, treaties, conventions or agreements?	☐

POST-AUTONOMOUS SYSTEMS	Yes/ No
Risk Factors from Integrated, Automated, Semi-Autonomous & Autonomous Systems, plus:	
POLITICAL	
Are the political consequences of the use, misuse or failure of post-autonomous systems irreparable?	☐
Do you understand the political impact of a 'Hyperwar' resulting from Post-autonomous Military or Weapons Systems?	☐
Are there any potential civil liberties and civil rights implications of post-autonomous systems engaging with humans, animals or within any societal systems or constructs?	☐
Could your post-autonomous system be used to exploit underdeveloped countries or create a political divide between developing and developed nations?	☐
ECONOMIC	
Do you know what physical security measures will be required for the physical protection of post-autonomous systems?	☐
Do you know what cyber-security measures will be required for the protection of post-autonomous systems?	☐
Do you know what the potential economic costs associated with irreparable environmental damage or destruction are?	☐
Do you know the potential economic costs of legal liability arising from AI use, misuse or failure?	☐
SOCIAL	
Do you know what social risks may emerge from the point of singularity and the arrival of Quantum computing?	☐
Do you know what the potential social costs (food security, income and social status, health, education, physical environment etc) associated with irreparable environmental damage or destruction are?	☐
Are there any social implications of limited personal autonomy resulting from post-autonomous Systems?	☐
Do you understand the social costs of limited or no personal privacy?	☐
TECHNOLOGICAL	
Do you understand the technological risks which may result from Quantum computing?	☐
Are there any security vulnerabilities of Quantum technologies for your AI system?	☐
Are there any consequences of failure to override post-autonomous systems?	☐
ENVIRONMENTAL	
Could post-autonomous technologies direct irreparable environmental damage or destruction?	☐
How might widespread environmental destruction impact on human survival?	☐
LEGAL	
What legal frameworks exist to provide governance and oversight in the design, development and deployment of post-autonomous systems at national and international levels?	☐
Would the use or release of force from Post-autonomous military technologies and systems contravene International Humanitarian Law or other international law, treaties, conventions or agreements?	☐

Glossary of Terms

Affective Computing	Systems and devices that can recognise, interpret, process and simulate human affects.
Agent	Anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators.
Agent-based Modelling	A simulation modelling technique, where a system is modelled as a collection of autonomous decision-making entities called agents which individually assesses its situation and makes decisions on the basis of a set of rules.
Algorithm	An instruction, or set of instructions, which a computer or system will follow to perform a task.
Analytics	Includes Descriptive Analytics, Diagnostic Analytics, Predictive Analytics and Prescriptive Analytics and uses Machine Learning techniques to find patterns and discover insights and relationships within data.
Artificial Broad Intelligence	The integration of two or more narrow AI systems or techniques that make decisions to perform a task or process.
Artificial Intelligence	Agents that receive percepts from the environment and perform actions.
Artificial Intelligence Paradigm	The approaches (tools and methods) used to develop the algorithms used to operate intelligent systems and devices.
Artificial Narrow Intelligence	Also known as 'Weak AI' or 'Narrow AI', and represents most of the current AI systems or techniques, which focus on performing specific tasks.
Artificial Neural Networks	A computational model in machine learning, which is inspired by the biological structures and functions of the mammalian brain, consisting of multiple units called artificial neurons which build connections between each other to pass information.
Artificial Super Intelligence	Intelligence that surpasses human ability.
Automation	The use of human decisions and logic that a system executes to accomplish a pre-set series of tasks within a known, or assumed, frame of reference without decisions being made during operation.
Autonomous System	A system where decisions are made (in response to external inputs or signals) which do not involve human decision making.
Computer Vision	AI concerned with the extraction of meaningful structures from images perceived by a system. Computer vision is the ability for a computer to use an artificial intelligence algorithm to 'see' and interpret both still images and video.
Deep Learning	Machine Learning using multiple layers of simple, adjustable computing elements.
Deep Neural Network	A neural network architecture with many layers, typically 5–100. A neural network with only a few layers is called a shallow network.

Detection	Determining that an unauthorised action has occurred or is occurring; detection includes sensing the unauthorised action, processing and communicating the alarm status to a control centre.
Distributed Artificial Intelligence	A class of technologies that solve problems by distributing them to autonomous "agents" that interact with each other. Multi-agent systems (MAS), Agent-based modelling (ABM) and Swarm Intelligence are examples of this where collective behaviours emerge from the interaction of decentralised self-organised agents.
Embodied Intelligence	An approach to AI and cognitive science that largely renounces symbolic representations and formal reasoning, and emphasises context, physical embodiment, social interaction and sensorimotor behaviour over generally abstractness, individualism and logically rigorous thought.
Evolutionary Algorithms	Evolutionary programs can alter their own rules by using genetic algorithms.
Expert System	A computer system that simulates the ability or behaviour of a human expert on performing a task. An expert system incorporates the knowledge base that represents facts and rules, and the inference engine that uses the knowledge base to deduce new conclusions.
Genetic Algorithms	A machine learning method for finding solutions to certain kinds of problems, loosely analogous to the biological process of artificial selection.
Hyperwar	War or conflict which accelerates and unfolds at an unprecedented rate due to AI capabilities and speed of response.
Integrated Security	A combination of security technologies, functions and devices, or quite simply; an assimilation of different security services which communicate to perform advanced functions in, as a minimum, an automated manner.
Integration	The act of combining or adding parts to make a unified whole.
Intelligent System	An advanced system comprised of connected elements or components that perceive and respond to the world around them.
Machine Learning	Computational models that have the ability to "learn" from the data and provide predictions. Depending on whether there is a supervisory signal, machine learning can be divided into three categories: the supervised learning, unsupervised learning, and reinforcement learning.
Machine Vision	A technology used to provide image-based automatic analysis for applications in industry such as automatic inspection, process control, and robot guidance.
Multi-Agent System	A collection of autonomous agents that need to coordinate their activities in order to achieve their individual goals. Coordination is achieved through negotiation or argumentation an, in most applications, requires that the agents learn to adapt to each other's strategies.
Neural Network	Also known as artificial neural network, neural net, deep neural net; a computer system inspired by living brains.
Observation	The function of detecting changes in a scene.

Passive	Refers to a system or sensor which does not emit signals, rather it operates by detecting, receiving or recording signals.
Probabilistic Programming	A framework that does not force you to hardcode specific variables, but rather works with probabilistic models.
Recognise/ Recognition	Recognition refers to the detection in images of a known object or instance (e.g. a mug) or a class of objects (e.g., the set of all mugs).
Reinforcement Learning	Dynamic programming that trains algorithms using a system of reward and punishment. The algorithm is exposed to a random and new dataset and it automatically finds patterns and relationships inside of that dataset. The system is rewarded when it finds a desired relationship in the dataset but it is also punished when finds an undesired relation. The algorithm learns from awards and punishments and updates itself continuously.
Respond/ Response	The element of a physical protection system designed to counteract adversary activity and interrupt the threat, or the effort to neutralise, contain, or mitigate an event.
Robot	A physical device capable of behaviour in the world involving interactions with its environment through sensors and actuators.
Robotics	The field of study dedicated to the science and engineering of robots.
Robotic Process Automation	Technology that extracts the list of rules and actions to perform by watching the user doing a certain task.
Safety Integrity Level (SIL)	Relates to frequency and risk of hazards, and performance requirements for achieving and maintaining safety - the higher the SIL, the greater the risk of failure.
Search and Optimisation	Tools that allow intelligent search with many possible solutions.
Sensor	A device that responds to a stimulus associated with an unauthorised action, such as an intrusion into a protected area or an attempt to smuggle contraband through an entry.
Shallow Neural Network	A neural network architecture with one hidden layer, as opposed to a deep neural network with many hidden layers.
Statistical System	A statistical approach is based on mathematical tools to solve specific sub-problems.
Sub-symbolic System	The sub-symbolic approach is one that no specific representations of knowledge is provided ex-ante.
Supervised Learning	Training a model from input data and its corresponding labels.
Swarm Intelligence	See ‘Distributed Artificial Intelligence’.
Symbolic System	A computer program that performs computations with constants and variables according to the rules of algebra, calculus, and other branches of mathematics. Also known as algebraic computation system; computer algebra system; symbolic computation system.
Unsupervised Learning	A type of machine learning algorithm used to draw inferences from sets of data consisting of input data without labelled responses, e.g., cluster analysis. The system is exposed to a random and new dataset and it automatically finds patterns and relationships inside the dataset.

Appendix A

The Security Technology Intelligent Autonomy Scale

		Level	Set Mission	Input/ Sensing	Computational Technique/ Processing	Rules/ Deciding	Output/ Acting	Review Mission
General AI	**Post-Autonomous**	**11**	Post-autonomous system determines, plans, directs and executes missions.	Post-autonomous security system possesses *Theory of Mind* and uses artificial consciousness to acquire input.	Post-autonomous security system determines, selects, and executes optimal computational techniques, autonomously adapting to optimise processing outcomes.	Post-autonomous security system determines, writes and adjusts rules, produces, analyses and predicts threat outcomes to optimise decision making.	Post-autonomous security system acts to execute security mission, using self-awareness for adaptation.	Post-autonomous system performs post-mission review.
	Autonomous	**10**	-	Fully autonomous security system observes, detects and monitors all inputs from integrated sensors and field-level components. System commands mission and acts autonomously, eliminating human intervention.	Fully autonomous security system integrates data (inputs) and applies computational techniques to analyse threats. System prepares to take action without human intervention.	Fully autonomous security system performs threat assessment, produces and ranks results, performs management-level decision making, and does not display results to the human-operator.	Fully autonomous security system acts autonomously, eliminating human intervention.	-
Broad AI		**9**	-	Autonomous security system observes, detects and monitors all inputs from integrated sensors and field-level components. System commands and acts autonomously, informing the human after execution.	Autonomous security system gathers data (inputs), applies computational techniques to process and interpret threats and prepares to take action informing the human-operator but not waiting for consent. Does not display results.	Autonomous security system performs threat assessment, analyses produces and ranks results, performs management-level decision making. Displays results to the human-operator upon query.	Autonomous security system acts autonomously, but informs the human after execution.	-
	Semi-Autonomous	**8**	-	Semi-autonomous security system observes, gathers, filters, and prioritises field-level inputs; displays information only if asked.	Semi-autonomous security system gathers data (inputs), applies computational techniques to process data, interpret threats and integrate data into a result which is displayed to the human-operator only upon request.	Semi-autonomous security system performs decision making. The system makes final decisions, but does not display results to the human.	Semi-autonomous security system executes automatically and does not allow any human interaction.	-
		7	-	Semi-autonomous security system observes, gathers, filters, and prioritises field-level inputs without displaying any information to	Semi-autonomous security system analyses, processes, interprets, and integrates data (inputs) into a result which is only displayed to the human if	Semi-autonomous security system performs decision making. The system makes final decisions and displays a reduced set of ranked options	Semi-autonomous security system executes automatically and only informs the human if required by context. It	-

Narrow AI				the human. Status on command execution is provided.	result fits programmed context (context dependant summaries).	without displaying "why" decisions were made to the human.	allows for override ability after execution. Human is shadow for contingencies.	
	Automated	**6**	-	Automated security system observes, gathers, filters, and prioritises field-level inputs with information displayed to the human.	Automated security system overlays processing with analysis and interprets the data (inputs). The human is shown all results.	Automated security system performs decision making and displays a reduced set of ranked options while displaying "why" decisions were made to the human.	Automated security system executes automatically, informs the human, and allows for override ability after execution. Human is shadow for contingencies.	-
		5	-	Semi-automated security system gathers field-level inputs from the subsystems and environment, but it only displays non-prioritized, filtered information.	Semi-automated security system overlays processing with analysis and interprets the data (inputs). The human shadows the interpretation for contingencies.	Semi-automated security system performs decision making. All results, including "why" decisions were made, are displayed to the human.	Semi-automated security system allows the human a context-dependant restricted time to intervene before execution. Human shadows for contingencies.	-
	Integrated	**4**	-	Integrated security system is responsible for gathering field-level inputs for the human and for displaying all information, but it highlights the nonprioritized, relevant information for the user.	Integrated security system analyses inputs and processes, though the human is responsible for interpretation of the data.	Both human and integrated security system perform decision making, the results from the system are considered prime.	Integrated security system allows the human a pre-programmed restricted time to intervene before execution. Human shadows for contingencies.	-
		3	-	Semi-integrated security system is responsible for gathering field-level inputs and displaying unfiltered, unprioritized information for the human. The human still is the primary monitor for all information.	Semi-integrated security system is the prime source of input analysis and processing, with human shadow for contingencies. The human is responsible for interpretation of the data.	Both human and semi-integrated security system perform decision making, the results from the human are considered prime.	Semi-integrated security system executes decision after human approval. Human shadows for contingencies.	-
	Manual	**2**	-	Human is the prime source for gathering and monitoring all data from field-level inputs, with security system shadowing for emergencies.	Human is the prime source of input analysis and processing, with security system shadowing for contingencies. The human is responsible for interpretation of the data.	The human performs all decision making, but the security system can be used as a tool for assistance.	Human is the prime source of execution, with security system/computer assistance for contingencies.	-
		1	-	Human is the only source for gathering and monitoring (defined as filtering, prioritizing and understanding) all data.	Human is responsible for analysing all inputs, processing, and interpretation of the data.	Security system does not assist in or perform decision making. Human must do it all.	Human alone can execute decision.	-

APPENDIX B

PESTEL Risks of Artificial Intelligence in Security Technologies

	TECHNOLOGIES	BENEFITS	RISKS
P	**POLITICAL**	**Facilitation of Law Enforcement & Intelligence Objectives:** Additional data for law enforcement & intelligence Enhanced aggregation of disparate data for law enforcement Automated translation of information from foreign languages	**Socio-Political Risks:** Political, cultural, social bias in machine learning algorithms Civil liberty breaches from AI systems or devices physically engaging with humans Difficulty in attributing accountability & responsibility for AI systems & AI decision making Lack of transparency in AI decision making Privacy, ethical & human rights issues Managing public perceptions of risk Reputational damage and loss of trust Unauthorised use of data from social media platforms **Commercial/Industrial Risks:** Industry disruption from AI use, misuse or failures Industrial relations issues resulting from use, misuse or failures of AI Regulatory risks & implications (i.e. overregulation, industry self-regulation) **National Security & International Relations:** Disproportionate development between countries Political divide between developing and developed nations Exploitation of underdeveloped countries (experimental AI) Sovereign risk resulting from reliance on foreign technologies and entities Foreign espionage The weaponisation of AI Privatisation of military AI Political implications from the use or release of force Uncontrolled UAVs may ground commercial & emergency aircraft Criminal exploitation of AI
E	**ECONOMIC**	**Cost Benefits:** Cost reduction Improved allocation of resources **Operational Benefits:** Reduced operator workload and fatigue Increased operational capabilities Reduced disruption from nuisance alarms Increased productivity	**Operating Costs:** Cost burdens of purchase, installation & maintenance **Indirect Costs:** Economic costs of disruption/ downtime from system maintenance or failure Impact of system reliance (i.e. reduction in manual operation capabilities) Costs associated with data ownership, storage and location Economic costs of security and AI decay (e.g. upgrades, security patches, etc.) Low resilience and redundancy across systems and devices **Broad Economic Costs:** Cost to industry/jobs from redirection of labour

			Reputational risk/ loss of consumer confidence Industrial relations issues resulting from use, misuse or failures of AI Reliance on the cyber domain for design & development of security technologies
S	**SOCIAL**	**Improved Safety:** Human-machine Teaming to reduce safety risks Removal of humans from the front line **Early Intervention Opportunities** **Convenience**	**Lack of Transparency in Development of AI:** Absence of legal and political frameworks to govern the development and deployment of AI in socially acceptable ways Unauthorised use of data from social media platforms **Social Risks from Unregulated/Uncontrolled AI** Social risks of inability to revoke AI development Exploitation of biometric data is irreversible Privacy, ethical & human rights issues Social implications from the use or release of force Lack of red teaming capability to reduce adversarial threats Exponential development may inhibit social responsibility Misuse of data Reputational risk (personal, corporate or government) Criminal exploitation of AI/ AI vulnerability to defeat **Safety Risks:** Fatal consequences of AI failures Breaches of capable guardianship & duty of care Civil liberty breaches from AI systems or devices physically engaging with humans Inappropriate training of personnel Lack of moral and emotional intelligence in AI Uncontrolled UAVs may ground commercial & emergency aircraft **Social Acceptance:** Low public user/acceptance of AI Public discontent concerning government use of AI Social fear/distrust of AI Low consumer tolerance of retrospective remedies
T	**TECHNOLOGICAL**	**Increased Operational Capabilities:** Increased accuracy Increased real-time processing capability Increased mobility & connectivity Reduced latency issues between technologies and devices Human-machine teaming to defeat human adversaries More extensive and directed surveillance Diversity of applications Faster, more focussed decision making and response Improved control of/ response to critical incidents **Improved Threat Detection & Mitigation Opportunities:**	**Lack of Transparency:** Black Box of AI issues & risks Difficulty in reverse-engineering machine learning algorithms for transparency Forensic auditing & assessment **Safety Issues:** Injuries or fatalities resulting from AI & technology failures Vulnerability from low accuracy/reliability of technology Risks associated with machine learning databases Inappropriate training of personnel Autonomous identification & classification failures Autonomous failure of human override

		Early intervention opportunities Threat mitigation opportunities Increased adaptability from dynamic threat assessment Breaking the Kill Chain	Neglect of mission critical functions **Security Risks:** Cyber & cloud-based risks Data compromise/ breach Data integrity, control and management issues Hacking/ hijacking or modification of script Exploitation of security vulnerabilities of IOT and Edge of Network devices Misguided AI systems not following programming Misuse of data Criminal exploitation of AI based on known vulnerabilities or responses Adversarial exploitation of AI Vulnerability to tampering & defeat Manipulation of data via weak system integrity Limited counter drone technologies available commercially Autonomous identification & classification failures Dangers of uncontrolled AI Reduced physical response force for critical incidents Exploitation of biometric data is irreversible Privatisation of military AI **Operational Limitations & Risks:** Vulnerability from low accuracy/reliability of technology Complacency in operation of unmanned systems Practical issues associated with generation or acquisition of datasets Complexity of systems Low resilience and redundancy across systems and devices AI implementation & management issues Security and technology decay Potentially limited capacity for 'untraining' and 'retraining' machine learning algorithms Limited source data Autonomous response may create additional risks/ issues **Risks to Industry:** Security/ IT divide Reliance on the cyber domain for design & development of security technologies Regulatory risks & implications (i.e. overregulation, industry self-regulation)
E	**ENVIRONMENTAL**		**Environmental Consequences**: Uncontrolled UAVs may ground commercial & emergency aircraft Environmental consequences of AI use, misuse or failure
L	**LEGAL**	**Reduced Legal Liability:** Reduction or elimination of safety risks **Facilitation of Legal Process:**	**Legal Frameworks:** Absence of legal and political frameworks to govern the development and deployment of AI in socially acceptable ways Lack of security governance/ oversight

		Increased traceability of incidents	Absence of legal frameworks for counter drone applications/ response Limited regulatory frameworks for source data (e.g. databases with image classifications) **Lack of Transparency** 'Black Box of AI' issues & risks Difficulty in forensic auditing and assessment **Legal Liability Arising From:** AI failures and system errors Injuries or fatalities resulting from AI & technology failures Privacy & human rights issues Safety risks & issues Civil liberty breaches from AI systems or devices physically engaging with humans Duty of care Adversarial exploitation of AI systems Inadequate human assessment resulting from reliance on AI systems Reduced physical response force for critical incidents Retrospective legal action Industrial relations issues resulting from use, misuse or failures of AI Non-compliance **Legal Limitations:** Legal limitations on acquisition of source data Issues concerning data ownership, storage & location Regulatory risks & implications (i.e. overregulation, industry self-regulation) Security/ IT divide