Autonomy is the answer, what was the question again?

WO2 PJ Spayne[a&b*] MSc, BEng(Hons), CEng, MIET, RN; Dr LJ Lacey[b] PhD, MSc, BEng(Hons), MRAeS, FHEA; Dr M Cahillane[b] PhD, MSc, BSc(Hons), CPsychol; Prof AJ Saddington[b] EngD, BEng(Hons), CEng, FRAeS, FHEA

a Royal Navy (United Kingdom); b Cranfield University (United Kingdom)
* Corresponding Author. Email: Peter.Spayne@Cranfield.ac.uk

Synopsis

In recent years aspirations regarding the implementation of autonomous systems have rapidly matured. Consequently, establishing the assurance and certification processes necessary for ensuring their safe deployment, across various industries, is critical. In the United Kingdom Ministry of Defence distinctive duty holder structures - formed since the publication of the Haddon Cave report in 2009 - are central to risk management. The objective of this research is to evaluate the duty holder constructs suitability to cater for the unique merits of artificial intelligence-based technology that is the beating heart of highly autonomous systems.

A comprehensive literature review examined the duty holder structure and underpinning processes that form two established concepts: i) confirming the safety of individual equipment and platforms (safe to operate); and ii) the safe operation of equipment by humans to complete the human-machine team (operate safely). Both traditional and emerging autonomous assurance methods from various domains were compared, including within wider fields, such as space, medical technology, automotive, software, and controls engineering. These methods were analysed, adapted, and amalgamated to formulate recommendations for a single military application.

A knowledge gap was identified where autonomous systems were proposed but could not be adequately assured. Exploration of this knowledge gap revealed a notable intersection between the two operating concepts when autonomous systems were considered. This overlap formed the development of a third concept, *safe to operate itself safely*, envisioned as a novel means to certify the safe usage of autonomous systems within the UK's military operations.

A hypothetical through-life assurance model is proposed to underpin the concept of *safe to operate itself safely*. At the time of writing the proposed model is undergoing validation through a series of qualitative interviews with key stakeholders; duty holders, commanding officers, industry leaders, technology accelerator organisation leaders, requirements managers, system designers, Artificial Intelligence developers and other specialist technical experts from within the Ministry of Defence, academia, and industry.

Preliminary analysis queries whether a capability necessitates the use of autonomy at all. Recognising that some autonomous systems will *never* be certified as safe to operate themselves safely, voiding ambitious development aspirations. This highlights that autonomy is simply one of many tools available to a developer, to be used sparingly alongside traditional technology, and not a panacea to replace human resource as originally thought.

This paper provides a comprehensive account of the convergence between *safe to operate* and *operate safely*, enabling the creation of the *safe to operate itself safely* concept for autonomous systems. Furthermore, it outlines the methodology employed to establish this concept and makes recommendations for its integration within the duty holder construct.

Keywords: LAWS; AI; Autonomy; MoD; RN; Duty Holder; Safe to Operate; Operate Safely

1. Introduction: How Do You Trust a Robot?

Preliminary research was conducted to analyse Lethal Autonomous Weapon Systems (LAWS) and aimed to develop a model for building trust among stakeholders integrating new autonomous technologies. However, research quickly revealed a significant knowledge gap requiring investigation that forced a change of direction towards human-machine teaming and artificial intelligence (AI). The initial literature surveys concluded early that established assurance mechanisms designed to ensure that human operators *operate safely* and that conventional machines are *safe to operate* are not fit for purpose when applied to the assurance of AI based autonomous systems. Therefore, a new method to account for technology able to *operate itself* was devised.

Authors' Biographies

WO2 Peter Spayne is a Weapon Engineer in the RN, majoring on mine disposal systems. He studied Autonomous systems at BEng and explored weaponising high powered ultrasound to agitate submerged explosives at MSc with Staffs Uni. He is a PhD student researching the assurance of lethal autonomous weapon systems at Cranfield University.

Dr Laura Lacey is a Senior Lecturer in Military Aviation Safety and Airworthiness at Cranfield University

Dr Marie Cahillane is Head of the Applied Psychology Group and a Senior Lecturer in Applied Cognitive Psychology within the Centre for Electronic Warfare, Information and Cyber at Cranfield University. She has over 15 years' experience in leading and collaborating on defence and Security research.

Prof Alistair Saddington is Professor of Defence Aeronautics. He has over 25 years of experience in aerospace engineering across both industry and academia including research leadership, management, and postgraduate teaching and supervision.

 https://doi.org/10.24868/11162

This study focuses on risk management and assurance methods used by the British Royal Navy (RN), covering LAWS within a broader spectrum of autonomous systems. Given the RN's operations across various domains, LAWS is understood here as a self-sufficient weapon, capable of learning to independently performing target acquisition, discrimination, and engagement in compliance with International Humanitarian Law (IHL). The emphasis on LAWS, as opposed to general Autonomous Systems, arises from their potential to make life-or-death decisions, necessitating a deeper exploration of an AI's role in making judgments. LAWS can be static, part of a larger system, or attached to conventional vehicles, including advanced targeting and guidance systems capable of selecting targets for engagement post-human intervention. Henceforth, 'autonomous system' will encompass all forms of autonomy, including LAWS.

2. Literature Review: Automatic, Automated and Autonomous

Kenneth Payne's 'I-Warbot' (Payne, 2021) highlighted that the WWII V2 rocket was considered as the first autonomous weapon by earlier standards. Its analogue mechanics - altimeters, barometers, fuel float switches, and gyroscope - allowed it to adjust its flight based on environmental feedback, without AI or complex computing. Today, such systems are classified as automatic or automated, not autonomous.

To investigate further the research protocol was given a favourable opinion by the Ministry of Defence Research Ethics Committee (MODREC) ref: 2265/MODREC/23. Twenty interviews with MOD and industry personnel identified a significant gap in the understanding of autonomy. Despite detailed knowledge of systems such as the Outfit DLH decoy launcher, Phalanx, and Seawolf, the majority of participants confused *automated* capabilities with *autonomy*. This confusion is partly due to a large number of misleading 'autonomy scales' in circulation, such as the IMO's 'Degrees of Autonomy' (IMO, 2019), which focuses on variations of remote control rather than any true autonomy.

Today the concept of autonomy in technology remains vaguely defined. While the Oxford English Dictionary (2010) defines it as 'self-governance', applying this to technology implies the need for intelligence. However, true autonomy, suggesting free will, or unpredictable actions due to unforeseeable inputs, poses safety concerns in technology applications. This distinction is crucial, as understanding the difference between manual, automatic, automated, and autonomous systems - illustrated in Table 1 using fire detection systems – is the foundation of the challenge in defining and safely integrating true AI based autonomy into technology.

Table 1: Distinctions of Autonomy

Statement	Example	Description
Manual	A person witnesses a fire, raises the alarm by shouting and fights the fire with a handheld extinguisher	A Manual Process of detecting and responding to a fire
Automatic	Smoke from a fire is sensed by a detector triggering an audible alarm	An environmental change automatically triggers an electro-mechanical process
Automated	Indications of a potential fire (Smoke and Heat) is detected to different degrees by multiple sensors causing software within an alarm system to *logically* conclude there is a fire. This triggers a predefined output from a Boolean table, such as an audible alarm, text message, or the activation of a spray system	An environmental change is noted by multiple sensors. An appropriate output is concluded by the system based on the logical sum of defined inputs. For example; Heat NOT Smoke = Investigation Warning, Heat AND Smoke = Alarm and Spray, Smoke NOT Heat = Alarm Only
Autonomous	The AI model central to a ships autonomous platform management system becomes aware of a fire in a machinery space, concluded from data received through inputs from an extensive array of sensors distributed throughout the ship and connected externally. This system gathers data on the location of the human crew, the state of machinery and fuel tanks, etc. Additional information about the ship's current tasking, the task group's situation, enemy activities, and broader mission	An Artificial Intelligence trained from the data of millions of human examples is central to a platform management system. It analyses complex data from a vast array of input sources and makes a complex decision that cannot be mapped to a single input, nor plotted in a Boolean table. The resultant output may need to be justified in a court of law. In this example the system replaced the cognitive reasoning, critical thinking and judgement required of a chief engineer, who may have acted in a similar fashion or made different decisions on a

 https://doi.org/10.24868/11162

Statement	Example	Description
	objectives provides essential situational context for decision-making. Upon assessing the situation, the AI probabilistically assesses that the human engineers in the compartment would not be able to effectively combat the fire or evacuate the area before it spreads to a nearby fuel sampling valve, which was already compromised due to a minor leak - being temporarily managed with rags and buckets as recorded in the digital maintenance logs. The resultant output is that the AI opts to activate mechanical ventilation valves to hermetically seal the compartment, trapping the engineers inside. Following this, it triggers a gas suppression system to extinguish the fire by removing oxygen from the compartment. Although this action results in the death of the engineers, it prevents a larger catastrophe, prioritising the ship's overall safety. Should any aspect of input data or situational context been different, the AI might have selected from an infinite array of other potential actions.	case-by-case basis, were a human in command of the control room during this incident.

Degrees of Autonomy

Table 1 suggested that a fire protection system's autonomy status - be it manual, automatic, automated, or autonomous - is fixed by design. Yet, autonomy is better understood as a state rather than a fixed capability. Interviews revealed a common misconception among technology developers from various sectors, including those working on autonomous ships and cars, who initially attempted to develop systems under the assumption that they would not require any form of manual control. This approach often led to the elimination of human-machine interfaces, comfort, and life-support systems, resulting in the need for retrofitting or the abandonment of trials (PSN9876, 4567, 0123, *Personal Communication (Anonymised Research Interview)*, 2024) (Lee and Wu, 2023).

Interviewees agreed that autonomy should be adjustable, analogous to a volume control, allowing for modulation as necessary (All Participants, *Personal Communication (Anonymised Research Interview)*, 2023/24). This concept aligns with the Yerkes-Dodson Law (Cohen, 2011) (Figure 1), which correlates performance with operator attention. Linking a need for an operator to remain alert to exercise the ability to dial up and down autonomy in line with the complexity of a situation. Building on this, Table 2 proposes an adjustable autonomy scale, and Table 3 introduces examples of degrees of autonomy within the hypothetical scenario of a minor warship, building on established crewing states (Navy Lookout, 2024). This demonstrates that a vessel can operate under various automation levels depending on the required task, suggesting that autonomy can and should be fine-tuned at both the system and subsystem levels for optimal performance and safety.

Table 2: Degrees of Automation

Degree	Definition	Description
M	Manual	Human operated manual control
R	Remote Control	System is manually operated from a remote location
1	Assisted	Very low-level automation for operator assistance. For example, cruise control aiding a driver may be considered automatic

 https://doi.org/10.24868/11162

2	Partial	Increased automation with more complex input, for example adaptive cruise control where a RADAR aided cruise control device monitors vehicle speed and distance from the vehicle ahead, causing the vehicle to accelerate or decelerate in response to sensed inputs. May be considered automated
3	Conditional	An autonomous system operating independently within a defined environment. For example, a driverless car on a racetrack. Complex examples may include geofencing via GPS when referring to ships or aircraft. Human override is available remotely or by an onboard operator who is monitoring the system *hands off*
4	High	As degree 3 but operating in an unrestricted (or less prescribed) environment. The ability for human override remains available but would be very rarely used
5	Full	As degree 4 but with no human override capability indicating that the system would never need human intervention. *Degree 5 autonomy, sometimes known as General AI is not yet technologically possible

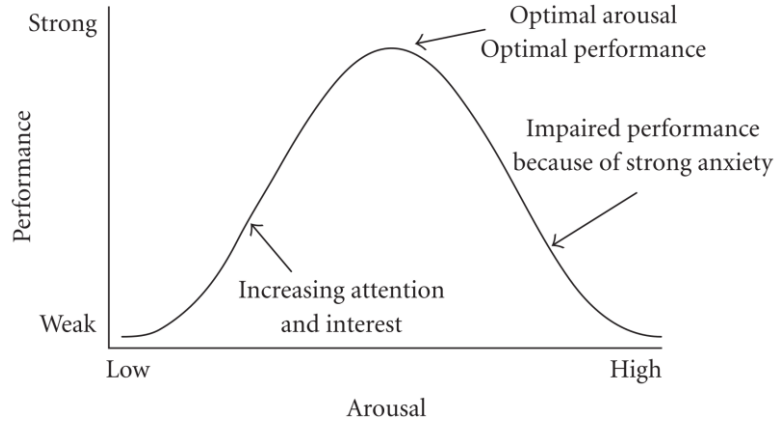

Figure 1: Yerkes-Dodson Law, Cohen (2011)

Table 3: Crew States

Crew State	Definition	Description
0	Action (Uncrewed)	In operations deemed extremely high risk, verging on suicidal, where crew safety is significantly compromised, the ship is transitioned to a state of full autonomy. Following crew evacuation, an integrated network of platform and combat management systems takes over, autonomously controlling all internal machinery, navigation, and combat operations. To minimise risks such as fire, life-support systems are deactivated, and some compartments may be depleted of oxygen. Operations proceed unattended, though the option to monitor and exercise remote control from an off-ship command centre remains, allowing for strategic oversight while prioritising human safety
1	Action (Crewed)	The ship is at action stations with all positions locally crewed ready to respond to enemy action. This posture is only maintained for short periods when an engagement is imminent. All weapons are readied
2	Defence (Crewed)	The ship's crew is keeping defence watches operating in a high-threat area where enemy engagement is possible, but not imminent. 50% of the crew are *on watch* operating weapons and sensors and

 https://doi.org/10.24868/11162

		contributing to the day-to-day running of the ship's routine. The other 50% are resting
3	Cruising (Crewed)	The ship is running a normal daytime routine akin to a merchant vessel, there is no risk of enemy engagement, and the crew are undertaking routine business in normal working hours
4	Cruising (Uncrewed)	For low-risk, routine tasks not requiring crew presence, the ship autonomously navigates to a set destination with life support and combat systems off, showcasing the efficiency of autonomous technology for benign operations.

* For tasks that are monotonous yet hazardous (mine hunting), a mix of State's 4 and 0 might be necessary. The ship methodically scans areas at slow speeds with SONAR. It is crucial to understand that autonomy is a state, not a fixed capability, and thus, the autonomy level should be adjustable to regain control as operational risks or environmental conditions change.

3. Safe to Operate and Operating Safely (The Duty Holder Construct)

The RN manages risks associated with the operation of technology by aligning with legislation from the UK's Health and Safety Executive (HSE) and wider bodies. Compliance is optional but mandated where possible by the Secretary of State. Procedures for adhering to this guidance (including deviations for military gain) when operating warships, weapons and equipment are detailed in Defence Safety Authority Book 2 - Defence Maritime Regulations (DSA02-DMR) (Defence Maritime Regulator, 2023). Specific guidance is then percolated through subordinate internal publications. The approval to depart from legislation and take risks for military gain, once thoroughly identified, is owned and authorised by Duty Holders, whose seniority dictates the severity of risk to life that may be tolerated.

Risk management involves a network of HSE, IMO and specific International Standards Organisation (ISO) regulations, tailored to equipment, scenarios, and environments. Military operations often require deviations from standard guidelines; therefore, deviations are managed internally by duty holders who strive to keep risks 'as low as reasonably practicable' (ALARP) (Defence Maritime Regulator, 2016).

Duty holder facing organisations throughout defence (Defence Maritime Regulator, 2018) administer safety management and ensure equipment and services meet safety standards, reporting non-compliance as required. They also curate policy for the safe use of equipment that does not fall under civil regulation, such as weapons and bespoke submarine operations. Specifically, platform and equipment authorities ensure equipment is *safe to operate* by defending safety compliance arguments in safety cases, achieving certification from the Naval Authority, which sets maritime compliance rules and standards by translating civil regulations to policy and accrediting curated policy where civil regulations do not exist.

Training authorities ensure personnel are well-trained, equipped and vetted to undertake the safe operation (*Operating Safely*) of equipment, with unit or platform Commanding Officers (COs) ensuring adherence to training and legislation, making their platform a human-machine team under a single command.

This model, pairing humans and machines in two distinct silos parallels civilian systems, in the United Kingdom (UK) vehicles in use on public roads are regulated by design standards and Ministry of Transport (MOT) testing, resulting in an MOT Certificate, and drivers are trained, examined and accredited by the Driver and Vehicle Licencing Agency (DVLA), resulting in a driving licence, to form human-machine teams as road users, safety assured by their combined certification.

For military gain, exceptionally, platforms may operate non-compliant with COs managing minor risks. Major risks escalate up the duty holder chain of more senior risk holders for approval. Immediate, unconsented, deliberate but justifiable breaches may be taken, but fall under the CO's accountability no matter the risk to life.

While effective for conventional technology and human resource management, this system struggles with autonomous systems due to unpredictable AI model behaviours that are akin to human unpredictability. An Autonomous systems' model training and outputs, comparable to human actions, require analysis from data scientists which is beyond the capabilities of training authorities optimised to train humans and not validate training data. A third concept must therefore be created.

4. Conclusions: Operating Itself Safely – The Third State

As the degrees of autonomy that were detailed in Table 2 increase the two statements of *safe to operate* and *operate safely* begin to converge, as depicted in Figures 2, 3 and 4. Figure 4 depicts a system that has been tuned up to fourth degree autonomy. It therefore requires specialist assurance to ascertain its ability to *operate itself safely*.

 https://doi.org/10.24868/11162

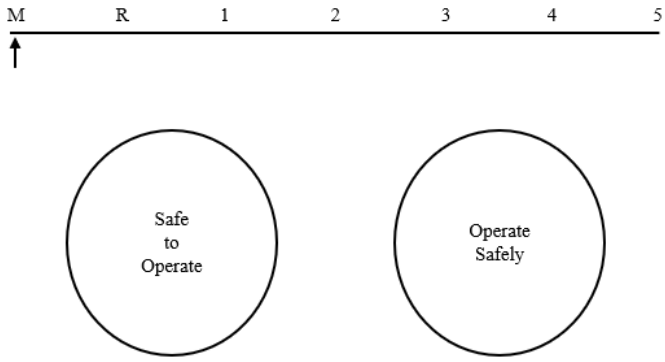

Figure 2: A manual system with separate agencies defining the assurance of machine aspects and human aspects. (x-Axis scale as per Table 2)

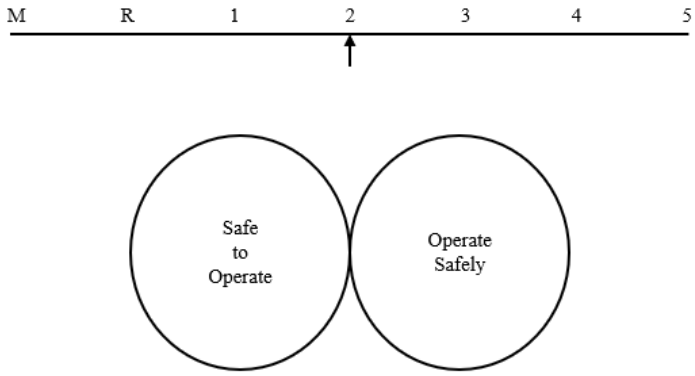

Figure 3: A second degree system (partial autonomy – automated) utilises separate agencies defining the assurance of mechanical aspects and human aspects but working much closer together.

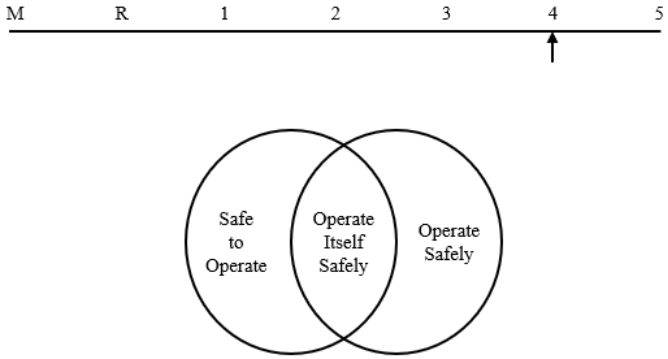

Figure 4: A fourth degree system (high autonomy), the two statements have converged and created a third in the overlap. This third space deals with the unique data science aspects of autonomy and AI model training data, outside of the scope of traditional equipment authorities or training agencies.

Autonomy is the sum of AI, Machine Learning (ML), and Data Science (DS) (Ministry Of Defence, 2022) This research has concluded that AI and ML demand specific assurance beyond the standard practices of training people and risk assessing to certify machinery. ML's reliance on training data necessitates extensive validation via DS, a process far removed from traditional human operator training scopes. Given the unpredictable nature of

 https://doi.org/10.24868/11162

decision-making in ML and AI, traditional equipment authorities cannot assess these systems using the usual Software Integrity Level (SIL) certification (International Standards Organisation, 2023).

However, the new third concept is not a replacement, a fundamental need for assurance of traditional hardware and software remains as an AI model sitting within an autonomous system consisting of normal hardware and software complements, rather than replaces, these components.

Additionally, even if a system achieved full autonomy, eliminating the need for constant human operation, personnel involved in the systems infrastructure or potential emergency intervention would still require training through established human training methods.

5. Output: Assuring a System is Safe to Operate Itself Safely

Assuring an autonomous system is a thorough and expensive process, which has previously deterred development progress beyond the conceptual stage (Thomas, 2022). Typically, unless a critical safety function necessitates autonomy – something unachievable by human or automated means – the system is unlikely to reach production or prove its viability past testing (Thomas, 2022). Autonomous systems, as decision-making entities mimicking human cognition, often become far more expensive than the human alternative. Continuous investment in autonomy must acknowledge the inherent safety limitations; achieving an acceptable ALARP safety level is challenging, given the absence of human-like accountability mechanisms.

Autonomous systems span three levels, as outlined in Table 4. This tiered structure introduces further complexity, as each level might function independently as an autonomous entity, necessitating individual assurance at each level, further multiplying the effort required to complete the assurance process (Safety of Autonomous Systems Working Group, 2022).

Table 4: Autonomous System Levels

Level	Definition	Description
3	System	The ‘platform’ representing the final autonomous system as a product. Directly interacts with the environment
2	Architecture	System software that hosts the AI model(s), operating software, and hardware at component level
1	Computation	Individual processes that translate inputs into outputs. Complex systems typically contain billions of computations across multiple architectures

The National Institute of Standards and Technology (NIST) (2023) study informed the AI Risk Management Framework (RMF) and underscored the need to consider AI-specific risks alongside conventional ones, due to their potential to affect a wider audience with more significant implications. In 2020, The Artificial Intelligence High Level Expert Group created the Assessment List of Trustworthy AI (ALTA) framework; allowing developers to self-assess their products during development, focusing on wider social and political aspects. ALTA addresses seven areas to assure a system is *operating itself safely*; these include human agency, analysing AI's impact on human behaviour; oversight, defining human interaction and training requirements; technical robustness, enhancing system resilience against unforeseen events and threats; privacy, ensuring data collection complies with human rights laws; explainability, clarifying the reasoning behind unusual decisions; and also addresses diversity, societal & environmental wellbeing, and accountability to ensure broad, responsible AI application (High Level Expert Group on Artificial Intelligence, 2022).

An autonomous system's integration within an organisation involves scrutinising every aspect of the TEPIDOIL[1] framework across the three levels detailed in Table 4. This comprehensive assessment encompasses the implications of AI-specific risk factors and the traditional assurance processes for human and non-AI system elements (Ministry of Defence, 2009). This ensures a holistic evaluation of the system's operation, highlighting the need for a coordinated approach to manage both conventional and AI-related risks effectively.

Assurance for deployment in specific environments initially permits only interim accreditation. Full *operating itself safely* status follows extensive operational use, with reliability proven through consistent documentation and the verification of expected outcomes.

Whilst deployed, autonomous systems can be comparatively assured to humans. Both process inputs; training, experience and judgement for humans is similar to programming and environmental interaction for an autonomous system. Inputs inform outputs that are subject to scrutiny. Therefore, the unpredictability of unknown responses

[1] 'Training, Equipment, Personnel, Information, Doctrine, Organisation, Infrastructure and Logistics

 https://doi.org/10.24868/11162

to novel situations means *full* safety assurances are unattainable. However, through repeated satisfactory performance, a basis for *reliability trust* may be established.

Assurance efforts should therefore focus on verifying training data and monitoring real-time responses to ensure the expected results. This approach can be delivered by way of a multilayer assurance model (Table 5) that encapsulates the system within its components, the organisation, and the wider operational context.

Table 5: Assurance Layers of Autonomous Systems

Layer	Definition	Description
-2	Computation	Identify and assure all computations, where reasonably practicable.
-1	Architecture	Define the system architecture, assure traditional hardware and software using established methods.
0	Platform (System)	Establish system boundaries, set requirements for operators and interacting agents to ensure they operate and interact with the system safely, conduct risk analysis for traditional system aspects to ensure traditional hardware and software components are safe to operate.
1	System of Systems	Identify the intended and unintended nodes in the wider system of systems, including other autonomous systems, traditional systems, and humans that may interact with this platform within the operational environment.
2	Operating Environment	Define the operating environment, identify sources of live data, potential sources of shared live data, and actively feed this back to training data creators for curation and reissue, to optimise the system for operations in a specific environment.
2a	Data	Assure and validate training data, secondary sources and third-party (shared) validation data, ensure data validation and curation tools are reliable. Assure the integrity of the simulated training environment.
3	Deployment	Implement live monitoring of operations, analyse returned data, and update assurance activities based on continuous feedback.
4	Real World Impact	Evaluate the system's impact on real-world communities, considering social, political, and economic influences in the operating environment.

This broad assurance model illustrates the extensive scope of examination required to assert system safety. Extending across the system's lifecycle (Table 6), this new approach diverges from the traditional CADMID[2] cycle and reflects a requirement to assure wider than typical system boundaries.

Table 6: Through Life Assurance of Autonomous Systems

Step	Definition	Description
1	Prepare	Prepare the organisation to receive an Autonomous System prior to procurement by setting the highest-level capability requirements, terminology, and definitions. Engage stakeholders to outline required governance before embarking on a typical engineering procurement cycle
2	Cartography	Map requirements and broader considerations for the entire system of systems, considering the intended operating environment and context
3	Global Risk Assessment	Assess known and predicted risks in 3-dimensions, factoring controllability and orientation of an operator during a state change. An appropriately instructed Generative Pre-Trained Transformer (GPT) may be used to risk assess a vast combination of scenarios and environmental variables

[2] Concept, Acceptance, Design, Manufacture, In-Service, Disposal

 https://doi.org/10.24868/11162

4	Go/No Go	Answer a series of Go/No-Go questions to determine the costs and feasibility of continuation of the project. What is also considered is whether the system (and which aspects) needs to be autonomous given the cost
5	In Service Monitoring	Active monitoring during systems development, model training, and deployment
6	Disposal	Outline a disposal and emergency withdrawal plan that adheres to ethical guidelines and considers data security aspects

The impact and influence of each layer must be considered within every step. When the system initially deploys with *interim* accreditation as *safe to operate itself safely*, Step 5 shall enact a feedback loop to Step 2 to enable the system to work towards full accreditation. This accreditation is based on the consistent safe performance of the system and may be withdrawn at any time.

Disposal at Step 6 is more complex than simply shutting the system down and dismantling it for disposal. The data must be sanitised for security and recycled into training data stock to inform the next generation of systems that will operate in the environment.

6. Summary

Autonomous systems fall outside the traditional UK MOD categorisations of being either 'safe to operate' or 'operating safely' due to AI's distinctive characteristics. They necessitate a third categorisation, *operating itself safely*, to meet their unique requirements. Despite the buzz around autonomy as a solution to global industry challenges, the complexity, and costs of integrating AI into LAWS or safety-critical systems have been underestimated. The expense of rigorous and widespread risk management and the additional costs for curating and validating training data greatly surpass those of existing manual or automated systems. Care should be taken to avoid the creation of an autonomous assurance cottage industry where it is not required.

Autonomy – misunderstood and often confused with automation – has been ambitiously presented as a complete solution, which has misled stakeholders regarding the capabilities of new systems. It is crucial to recognise autonomy as a *state* rather than a *capability*, ideally employed as a safety mechanism, battle override or emergency sub-system, activated only in critical situations, such as the incapacitation of an operator. These states require assurance of their safe self-operation outside of the scope of traditional agencies, demanding further development of the methodologies proposed in this paper for their readiness.

Whilst further research and development into autonomous systems is vital for technological progress, procurement of systems intended to be autonomous as the rule and not the exception, for widespread application and implementation today, is deemed unviable at this time.

7. Acknowledgement

The recommendations in this paper are preliminary conclusions from PhD research funded by the Royal Navy under the supervision of Cranfield University. Research is qualitative and participants have had their identities protected in compliance with rules set out by MODREC, in accordance with JSP 536 guidance. The authors would like to thank all those who agreed to be interviewed and gave their time to take part. Redacted transcripts are available upon request.

8. References

Cohen, R.A., 2011. Yerkes–Dodson Law, in: Encyclopaedia of Clinical Neuropsychology. Springer New York, New York, NY, pp. 2737–2738. https://doi.org/10.1007/978-0-387-79948-3_1340

Defence Maritime Regulator, 2023. DSA02 Defence Maritime Regulations for Health, Safety and Environmental Protection.

Defence Maritime Regulator, 2018. DSA01.2.

Defence Maritime Regulator, 2016. DSA01.1 Defence Policy for Health, Safety and Environmental Protection.

High Level Expert Group on Artificial Intelligence, 2022. Assessment list for trustworthy Artificial Intelligence (ALTAI) for self-assessment. https://doi.org/10.2759/791819

IMO, 2019. Autonomous shipping [WWW Document]. URL https://www.imo.org/en/MediaCentre/HotTopics/Pages/Autonomous-shipping.aspx (accessed 3.8.24).

International Standards Organisation, 2023. ISO/IEC/IEEE 15026-3:2023, Systems and software engineering: System integrity levels.

Lee, C.K.H., Wu, K.Y.K., 2023. Making autonomous vehicle systems human-like: lessons learned from accident experiences in traffic. Enterp Inf Syst 17. https://doi.org/10.1080/17517575.2021.1998641
Ministry of Defence, 2009. Defence lines of development analysis with MODAF.
Navy Lookout, 2024. After action report – Royal Navy's busiest air defence activity since 1982.
NIST, 2023. AI RMF PLAYBOOK.
Oxford English Dictionary, 2010. Autonomy.
Payne K, 2021. I Warbot.
Safety of Autonomous Systems Working Group, 2022. Safety Assurance Objectives for Autonomous Systems.
Thomas, 2022. Swatting of Mosquito latest UK uncrewed dead end [WWW Document]. Airforce Technology.com. URL https://www.airforce-technology.com/features/mosquito-swat-becomes-the-latest-in-a-long-line-of-uk-uncrewed-dead-ends/ (accessed 3.9.24).