Certifying for Operate Safely – Building Trust in Naval USVs

C M Baker, MSc, AMRINA, ARCNC[1], W Balfour, MEng, MSc, AMRINA, ARCNC[2]

[1]*Defence Equipment & Support, Bristol, United Kingdom*
[2]*NavyX, Portsmouth, United Kingdom*

Abstract

Autonomy and autonomous platforms are not just coming; they are here, and Defence isn't ready. Regulators are beginning to issue rules to certify autonomous vessels against; however, many rules do set empirically defined values that are to be achieved in support of a safe claim of compliance with goal-based certification. The Goal-based regulations for fully crewed platforms have centuries of empirical data demonstrating safe and unsafe practices and engineering. This has led to an inherent level of trust that Naval Vessels are Safe to Operate and can be Operated Safely. This breakdown of the Safe Claim, Argument, and Evidence is well understood within the defence's safety construct. Safe to Operate is the side of the argument provided and maintained by the Platform Authority (PA). Where Operating Safely is dependent on the Operating Organisation (Royal Navy (RN)), the separation between is unclear when certifying a vessel to operate at IMO Level 3 & 4 Autonomy. This is leading to platform teams being asked to demonstrate Operate Safely arguments on behalf of the operating bodies while the goal-based rules of Safe Operations remain undefined. Whilst there is some experience and commercial regulatory frameworks in place to provide an acceptable means of compliance, naval certification challenges remain. This is in part due to the need to assure a platform is Safe and Legal without limiting capability of the complex service vessels for broad scopes of operations. This paper aims to share the challenges, risks and opportunities identified through the lived experience of NavyX with the Autonomous Pacific 24 (APAC) Autonomy demonstrator.

Key Words: Autonomy; Naval Certification Safe to Operate; Trust in Autonomy

1. Introduction

Technological innovations are rapidly advancing, enabling Uncrewed Surface Vessels (USVs) to perform operations that were previously exclusive to crewed assets. These advancements are paving the way for Autonomous vessels, capable of making independent decisions and executing actions without human intervention, to complete their own Observe-Orient-Decide-Act (OODA) loop. The International Maritime Organisation (IMO) is in the process of developing a non-mandatory Goal-based Maritime Autonomous Surface System (MASS) code, set to be finalized by 2025, with a mandatory code to follow in 2028.

In the regulatory scoping exercise for this code, the IMO has defined 'degrees' of autonomy as:

- Degree one: ship with automated processes and decision support;
- Degree two: remotely controlled ship with seafarers on board;
- Degree three: remotely controlled ship with no seafarers on board;
- Degree four: Fully Autonomous ship.

[1] **Mr Chris Baker** is a Naval Architect at Defence Equipment & Support (DE&S) supporting the Queen Elizabeth Class Aircraft Carriers with experience in the Royal Navy's Autonomy and Lethality Accelerator, NavyX. He has previously presented on Alternative Fuels at the IMarEST Annual Conference and has been a Research Fellow at the University of Plymouth.
[2] **Mr William Balfour** is the NavyX's (the Royal Navy's Autonomy and Lethality Accelerator) Naval Architect & Marine Engineer, Supporting the RN's Autonomy Programme, XV Patrick Blackett, APAC & MADFOX. Having previously worked at DE&S within Ship Acquisition as well as wider Ministry of Defence innovation teams.

 https://doi.org/10.24868/11164

Crewed Naval vessels demonstrate their safety through a safety argument that outlines the vessel's scope of use and is certified against it. This argument is divided into Safe to Operate and Operate Safely. The Platform Authority (PA), the Technical Authority for ships operating within the Ministry of Defence (MOD), provides the Safe to Operate argument. The operating organisation, the Royal Navy (RN), provides the Operate Safely argument. This clear division of responsibility works well for crewed vessels, however as vessels transition to higher levels of autonomy this division becomes less apparent. Regulators are now having to assure both the vessel and the Autonomy Package, presenting new challenges in the certification process.

To better understand Autonomy the RN tasked NavyX to certify and experiment with APAC. This is a standard Pacific 24 Mk4 boat modified with an Autonomous package and has been used as an Unmanned Surface Vessel (USV) operational demonstrator for the Royal Navy. Early in October 2023 APAC was certified by the Naval Authority Technical Group (NATG) to operate to Degree 3 within limited areas for experimentation. This was one of the first boats NATG had certified to Degree 3 autonomy and consequentially identified that the means of acceptable compliance with their goal-based standards had not been fully explored. The PA found, at the time of the submission, that both sides of the safety argument were under subjective criticism having to comply with an as yet fully defined goal-based standard and limited trust in operators.

Figure 1- APAC24 in HMNB Portsmouth

This paper challenges the subjective setting of high and potentially inappropriate standards for autonomous systems. It explores the differences between certifying a degree one platform and an autonomous platform in a naval environment and how these differences impact the required evidence base and the balance between simulation and demonstration. The paper ultimately poses the question, 'Are standards for autonomous systems being set above and beyond what would be accepted for a crewed vessel because of a lack of trust?'. This paper critically examines the balancing act that this question poses.

2. The Autonomy Certification Delta

Certifying autonomous vessels, in theory, should have the same goals and aims as certifying a crewed vessel: that a vessel is safe to operate within a set of defined limits. Lived experience though shows a significant difference in attitude when determining the safe operational limits of autonomous vessels versus their crewed equivalent. This section will explore the differences required in evidence to

 https://doi.org/10.24868/11164

demonstrate an autonomous vessel is safe to operate, the impacts created by these deltas, and the positives and negatives of the current naval USV certification process.

2.1 Concept of Use

A Concept of Use (CONUSE) “describes the intended ways in which a specified capability is to be employed in a range of activities, operations, or scenarios” (Ministry of Defence, 2013). From a certification perspective, a good CONUSE will define what a vessel is intended to do and, equally, what it is intended not to do. A crewed vessel is assumed to not conduct an activity unless it is implicitly stated within the CONUSE. The emphasis on what the vessel will not do substantially increases with autonomous vessels.

Certification bodies are only prepared to certify USVs to perform specific tasks at pre-determined speeds and ranges. A crewed surface vessel could be permitted to carry a maximum number of souls up varying designated speeds, sea states and conditions. To perform additional tasks such as aviation operations or carriage of Weapons, Munitions and Explosives (WOME), additions to the certificate can be applied. For example, a Type 23 Frigate can carry 120+ people at 20+ knots in Sea States up to and including Sea State 9 to perform military operations.

However, a USV is not given the same level of flexibility. USVs will only be certified to perform designated tasks. For example, APAC24 can only operate at Level 3 autonomy for the purpose of experimenting with autonomous systems within a predefined area of operation.

2.2 *Safe Operating Procedures and Crewing Policy*

Safe Operating Procedures (SOPs) and Crewing Policies typically serve very different purposes. SOPs detail how the crew operate the ship safely whereas Crewing Policies dictate the number of crew required to perform SOPs. They contribute significantly to the Operate Safely argument and are often interlinked, but with USVs, they are almost the same document, particularly on smaller USVs.

Why is this? Fundamentally a USV’s operation changes dependant on the degree of autonomy in operation. In fact, most, if not all, Level 4 USVs will still maintain a level of oversight except for “fire and forget” items. An example of this can be seen with the US Navy’s Ghost Fleet Overlord programme where a LUSV completed a 4,421 nautical mile voyage at 98% Level 4 autonomy (Shelbourne & Lagorne, 2024). However, 6 crew members were still accommodated on board. Yes, there is an argument that this is because of the experimental nature of the vessels but even still, there are some areas where crew are still needed or preferred.

There is the argument of manned vs crewed. it is a topic worthy of discussion in its own right, but within this paper, crew are defined as Suitably Qualified and Experienced Personnel (SQEP). SQEP varies from platform to platform and very quickly the qualification and crewing burden becomes rather large to, in theory, to act as supervisor to an autonomous vessel. APAC24 is designed to be operated remotely with or without crew on board, the Degree 2 SOPs dictate there must be someone on board to physically hit the emergency stop and operate the boat. For safety reasons there then also needs to be a second person on board in case of a Man Overboard or injury expanding the crewing burden to just supervise.

As a result, whilst the technology might be operating at Level 3 or 4, in practice, SOPs prevent Level 4 operations in the true sense. However, this reliability on the human ability to intervene when increased levels of autonomy are in place is questionable. Several papers have shown that simply using crew to supervise instead of operate (Chan, et al., 2022a) (Chan, et al., 2022b) (Chan, et al., 2023), can increase risk due to behavioural changes, reduced practical experience, and complacency. All this raises the question, are crew still able to intervene as well as they could on a Level 1 or 2 platform? SOPs are designed to enforce best practices and safe operation, but is current regulatory hesitancy actually risking safe operation? Ultimately, USVs are designed to be cheaper alternatives to existing crewed platforms that reduce the risk to human life.

2.2 Hazard Log

A Hazard Log is a way of articulating a system’s risk to harm to As Low As Reasonably Practicable (ALARP). A log can be generated through a variety of means. It involves a SQEP panel reviewing a system and identifying all credible events that could cause harm to personnel, equipment, or the

 https://doi.org/10.24868/11164

environment. These events have an initiating Cause that could be safeguarded against occurring in the first instance and controlled before a situation or 'Hazard' becomes present. This 'Hazard' could lead to harm or accident. Once a Hazard is present mitigations can be implemented to reduce the severity or probability of the consequential accident occurring. This construct of events can appear as a Bow Tie with several Causes leading to a singular Hazard that could present a series of credible accidents, as shown in Figure 2.

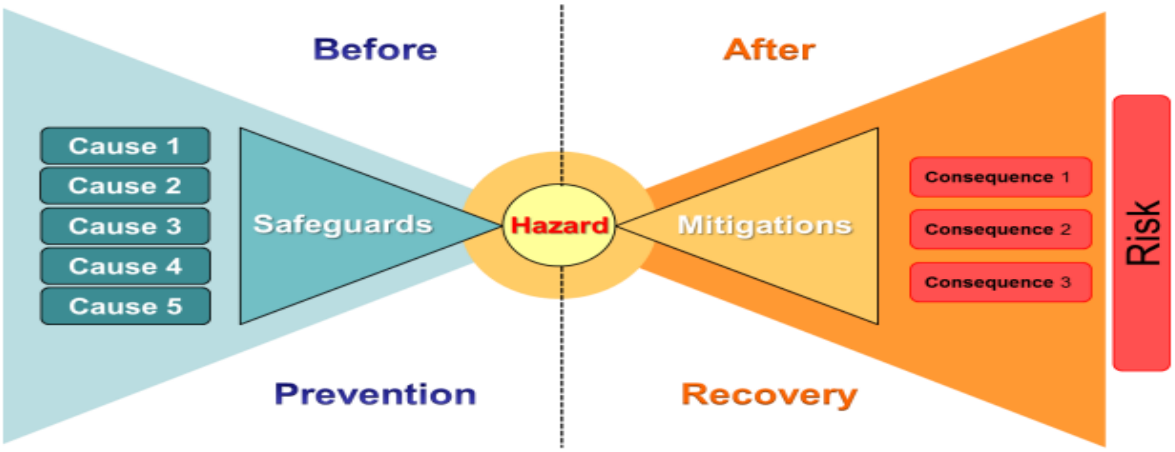

Figure 2 - Indicative Bow Tie Structure (DE&S, 2020)

The role of a SQEP panel is crucial in identifying and grading the risk of each accident against a standardised matrix. With its Subject Matter Experts (SME), this panel meticulously examines the elements that could lead to an accident. The highest risks to harm are then identified and prioritised for mitigation. This process involves identifying the design features, processes, or limits that must be present to satisfy the argument that a system's risk to harm is Tolerable and ALARP.

When considering autonomous vessels, many hazards and accidents must be reviewed for every degree of autonomous operation. Consequently, the harm to personnel focus moves from involved personnel to 3rd parties throughout the HAZID. For example, in the case of a collision, there will be harm to people, equipment, and the environment. When uncrewed, the personnel harm will only be to third-party personnel; this risk to 3rd parties was not scrutinised when crewed operations were examined. In the case when APAC was operating as a USV at Degrees 1 & 2, the responsibility fell within the operating organisation's safety construct; however, at Degree 3, the argument fell to the PA to satisfy.

This risk of harming 3rd parties can only be mitigated so far by systems and design; however, the accident's probability can only be mitigated by processes. i.e., operating within a safe space. This can be implemented by operating within controlled areas and using a geofence to set bounds for the autonomous system. The operating organisation must maintain this safe space outside the Safe to Operate scope. Within the Aviation space, this risk to 'uninvolved persons' from a Remotely Piloted Air System (RPAS) has many similar factors to USVs covering: Mass, dimensions, speed, quality of training and duration of exposure. Exposure is a function of the number of third parties at risk and the period that they are at risk. This risk is managed within the operating side of the safety checklist section of an RPAS's categorisation argument.

However, this argument was being asked to be made by the PA when certifying APAC within the Safe to Operate argument in which this enforced system would act as failsafe in case of loss of control of the system or connection failure. These are embodied within an emergency stop function that turns off the boat's automation if the connection is lost or under the operator's command. This is the extent of the PA's control of the operation, providing an assured system to a suitable Safety Integrity level. It remains for the operators to know to activate or manage connection links to not endanger uninvolved persons, equipment and the environment.

2.3 Class Certificates

 https://doi.org/10.24868/11164

Unlike certificates of class issued to crewed vessels, certificates of class for USVs are not as advanced. Using the example of a Lloyds Register (LR) Unmanned Marine Systems (UMS) Certificate it is clear that the main focus of LR is on the basic functionality of the systems that make the vessel autonomous, i.e., does the vessel start and stop when told to. It does not cover issues such as stopping distances or positive identification distances of other objects. This can be problematic as the level of assurance provided to naval certification bodies is reduced. The other basic functions of the boat and items, such as structures and stability, would still have to be covered by the crewed vessel equivalent. With APAC24 this requires a Work Boat Code certificate and an UMS certificate.

As a result, many naval certification bodies would not accept UMS certificates as significant evidence. The evidence would still prove useful but if the scope of certification covered items such as stopping distances, the UMS certificate would carry far more weight.

In defence of class societies, generating an UMS code that fits all USVs is challenging. Crewed vessels have been operating for hundreds of years, and codes have been derived from reviewing thousands of safety incidents to establish best practices. For USVs, this bank of experience does not exist. This means that the knowledge base is predominantly theoretical and experimental in comparison.

Further compounding the issue is that attitudes to risk have changed significantly. In the early 1700s safety was an afterthought to capability. Now, safety is an integral part of everything the Royal Navy designs and operates. Risk also now extends to reputational damage. Using the examples of historical incidents that have redefined what is believed safe have in the long-term bettered industry. The damage that an incident could cause however, has led to an increasingly risk averse approach. This is problematic because fundamentally to fully develop a UMS code fit for purpose, mistakes are going to occur and must be learnt from.

The key difference with USVs, and where there is scope to expand the pace and risk appetite of certification bodies, is that systems can be proven with a vastly reduced risk to humans, even if that comes at a cost to the asset. If stakeholders are willing to accept an increased risk to experimental assets in a controlled environment, then the pace of development could increase drastically. Combining this with the ability to run preliminary testing in simulated environments will massively speed up the rate of development of UMS codes.

However, UMS codes will also need to be broken down further. A UMS code for a work boat will not be appropriate for a 300m container ship. The knowledge and experience base is not available to be drawn from, but codes must clearly state safe separation, identification distances and other minimums of navigational safety.

2.4 Safety and Environmental Case Report

Safety and Environmental Case Reports (SECRs) are a fundamental part of certifying any vessel, no different from USVs. The basic structure of a SECR remains the same in that to prove a vessel is Safe, it must be Safe to Operate and can be Operated Safely. Broadly, the evidence and criteria are the same as those for crewed vessels, but with one key difference: cyber security.

For naval vessels, cyber security is becoming a priority. The ability to safeguard sensitive and operational information has never been more paramount when the capability of Signals Intelligence (SIGINT) has evolved rapidly. As a result, far more detail and evidence are required that USVs can a) prevent intrusion on systems, b) prevent a hostile takeover of control systems, and c) be rendered harmless in the event of an unrecoverable hostile takeover.

Whilst this is undeniably a critical safety aspect, there is one problem: what do naval architects, marine engineers, and safety managers – typically tasked with conducting certification activities know about cyber security? Inevitably, this will force the adoption of cyber security experts into the certification process, both within class societies and project teams. This presents opportunities, most obviously, to offer upskilling to existing engineers within the industry, which can only be a good thing in an increasingly digital and cyber-contested world. Operationally however, it represents the ability to place

 https://doi.org/10.24868/11164

maritime cyber security firmly in the mindset of vessel designers and drive the adoption of improved practices and growth of capabilities

2.5 Trust

The main theme throughout the autonomy deltas is trust. Naval certification bodies do not yet trust USVs to do their jobs and therefore require a far higher level of evidence to achieve the same operating capability compared to a crewed equivalent. The primary reason for this, as seen in the deltas of CONUSE, SOPs and Class Certificates is that the knowledge and experience base is severely lacking. Humans are risk averse and creatures of habit, two things which do not go well with rapidly deploying autonomous systems.

3. Building Trust in Autonomous Systems

In naval environments trust is built within the equipment and the combination of equipment and operators. Modern navies use institutes such as Flag Officer Sea Training (FOST) to assure that a ship and its crew can deliver capabilities safely and effectively. Remove the crew and replace them with a machine -how do you then conduct FOST? This section will explore the challenges presented by this and what can be done to fill this gap in assurance so that trust can be built in USVs.

3.1 Demonstration

As with most new technologies, demonstration is the best and most effective way of building trust. The demonstration allows authorities and parties to see and gauge risk for themselves. However, demonstrations are seen as more of an exam when it comes to certification, and unlike an exam, autonomy demonstrations rarely have fixed objectives. Using example of an APAC demonstration, the experiments team had a set trial to develop evidence to inform the NavyX Autonomy Programme. This Trial Plan had set evolutions to be conducted with success criteria, whereas the regulators witnessing the demonstration did not have specific criteria to satisfy them APAC was Safe to Operate or Operate Safely.

3.2 Simulations and LUSVs

Demonstrations for smaller vessels such as APAC can be frustrating but achievable. With comparatively low running costs and support requirements USVs can be trialled several times until the certifying body is content. For LUSVs this is not a feasible approach. The support requirements, operating costs, and operational pressure to deliver new capabilities at pace would make certification unrealistic. The Operate Safely argument then strays rapidly from crew to platform authorities.

There is a solution to this problem. Simulations allow for hundreds or even thousands of pre-determined trials as requested by a certification body. This gives the flexibility to establish a baseline relatively quickly for the vessel against which a potential demonstration could be held. For example, the certification strategy will request a speed and operating envelope. The simulations can be used to establish if a demonstration at the full extent of the operating envelope is useful or whether sufficient concerns exist that further enhancements are required.

However, simulations can present an insufficient representation of the realities of the maritime environment. They can only represent a simplified maritime environment, while more complex simulations can represent the chaotic physical of the natural environment. Depending on the level of chaotic variation permitted in the simulated environment, hundreds of the same events could be simulated yet lead to the production of varying resulting states. It may infer a predictive behaviour, although the probability of system hallucinations remains.

Due to this simulation can only suggest that an automated system will behave in a predictive manner to a degree of certainty within the simulation environment. It cannot be a final answer as only live trials will expose the OODA loop system to the chaotic reality of the maritime environment and therefore there remains the issue of trusting its decisions.

The benefits of this are sizeable. It allows certification bodies time to develop the tools, knowledge, and subsequent rules set to objectively certify USVs without significantly increasing the trial burden. It

 https://doi.org/10.24868/11164

also allows the pace of in-service dates for USVs to be accelerated. This would be akin to using the simulations to replace FOST before sea trials. Sea trials could then be expanded by a day or two to conduct the final phase of FOST, thus not creating potential days' worth of extra trials. This also would allow the utilisation of a key part of USVs. Every time a ship is docked the Ships Company is required to conduct an extensive workup period for the sake of the equipment post such intrinsic maintenance, and for the crew to regain confidence in their SQEP.

4. Finding the Balance between Objective and Subjective

So far, this paper has focused on evidence to persuade certification bodies of Safe to Operate and the theme of trust, but this goes beyond evidence. As lived experiences have demonstrated, multiple ways of producing evidence and building a safety argument exist. However, even if an objective rules-based approach does not exist the problem of certifying naval USVs will persist.

Before the discussion continues it is pertinent to make clear that this is not an advocation for a fully objective rules-based system. There still fundamentally exists a significant element of learning and operational experience to be fully content a USV is safe, and that can only be captured through a certain degree of engineering judgement. Indeed, this is common with crewed platforms where concessions are made to demonstrate that a platform is still safe but does not necessarily comply with every rule.

Nevertheless, there is an urgent need to establish a route to achieving a steady state similar to that of crewed platforms. Proposals of frameworks, which include introducing some minimum benchmarks based on vessel size and purpose, is a crucial step in this direction. This approach, like existing surface ship rules, will provide engineers with a clear objective to meet when designing these platforms. It will also reduce the excessive evidence pool required to get a certificate, freeing up valuable time, resources, and money. This will not only enable us to exploit the advantages of USVs best but to also ensure that the development of USVs can continue at pace with a greater degree of regulatory control.

5. Conclusion

Through the lived experience of attempting to certify APAC through the NATG, lessons have been identified by the NavyX PA and NATG to improve the requirements in order to certify an autonomous boat. However, hurdles remain to ensure consistent assurance for USVs and LUSVs.

The PA should continue to provide the Safe to Operate argument with the expanded responsibility of providing Claims, Arguments and Evidence to support assurance that an autonomous system will behave predictably when not under the direct control of an operator, and that operators have sufficient systems to ensure they can operate the system safely. I.e., a minimum standard of Situational Awareness proportional to the CONOPS of the USV or sufficient fail safes are present in case of a runaway system.

Operating organisations must remain responsible for the risk of harming equipment, environment and personnel (including third parties) when operating autonomous assets. As with crewed vessels, when within the bounds of their safety case the responsibility remains with the operators to know and understand the risks involved in using equipment. The assurance of this must be updated to reflect the evidence provided by the PA in the Operate Safely Argument.

Trust can only be built in these systems by enabling their function and enabling the iterative development of autonomous decision-making systems; however, in a controlled manner, accidents are expected to occur with all systems. Society will need to decide when it is comfortable operating around these systems, as with the autonomous automotive and aviation industry.

References

Chan, D. J., Norman, D. R. & Pazouki, D. K., 2022b. An Analytical Assessment of the Situational Awareness of Seafarers & Their Trust in Automated Systems. *International Naval Engineering Conference .*

 https://doi.org/10.24868/11164

Chan, D. J., Norman, D. R., Pazouki, D. K. & Golightly, D., 2022a. Autonomous Maritime Operations and the Influence of Situational Awareness Within Maritime Navigation. *WMU Journal of Maritime Affairs,* pp. 121-140.

Chan, D. J., Norman, D. R., Pazouki, D. K. & Golightly, D., 2023. Perception of Autonomy and the Role of Experience within the Maritime Industry. *Journal of Marine Science and Engineering.*

DE&S, 2020. *20201031-SHIPS_O_and_A-Leaflet_5-Issue_3-O.pdf,* Bristol: DE&S.

Military Aviation Authority, 2023. *Regulatory Article (RA) 1600: remotely piloted air systems (RPAS) Issue 9,* s.l.: Ministry of Defence.

Ministry of Defence, 2013. *Capability Management Practitioners' Guide.* 2.1 ed. s.l.:Ministry of Defence.

Shelbourne, M. & Lagorne, S., 2024. *US Naval Intelligence News.* [Online]
Available at: https://news.usni.org/2024/01/16/navy-wraps-first-unmanned-surface-deployment-to-westpac
[Accessed 13 May 2024].

 https://doi.org/10.24868/11164